THE ASTROPHYSICAL JOURNAL, 379:52–79, 1991 September 20

GALAXY FORMATION THROUGH HIERARCHICAL CLUSTERING

SIMON D. M. WHITE[1]
Steward Observatory, University of Arizona

AND

CARLOS S. FRENK
Physics Department, University of Durham, Durham DH1 3LE, UK
Received 1990 December 10; accepted 1991 March 28

ABSTRACT

We develop analytic methods for studying the formation of galaxies by gas condensation within massive dark halos. Our scheme applies to cosmogonies where structure grows through hierarchical clustering of a mixture of gas and dissipationless dark matter. It is an elaboration of the ideas of White & Rees. We adopt the simplest models consistent with our current understanding of N-body work on dissipationless clustering, and of numerical and analytic work on gas evolution and cooling. We also employ standard models for the evolution of stellar populations, and construct new models for the way star formation heats and enriches the surrounding gas. Although our approach is phenomenological, we avoid assumptions which have no clear physical basis. Our methods allow us to predict star formation as a function of location and time, and so the following properties of the galaxy population: current star formation rates and halo X-ray luminosities; current luminosity functions both for galaxies and for virialized systems; relations between present luminosity, circular velocity, metallicity, and stellar or total M/L ratio; the history of the OB star contribution to the metagalactic ionizing flux; and the distribution of faint blue (star-forming) galaxies in both apparent magnitude and redshift. In this paper we give detailed results only for a cold dark matter universe with $\Omega = 1$ and $H_0 = 50$ km s^{-1} Mpc^{-1}, although our methods are easily applied to other models. Even for this case, predictions depend strongly on the mean baryon density, on the fluctuation amplitude, on the models for heating and metal enrichment by massive stars, and on the initial mass function with which stars form. Our most successful models require a large baryon fraction ($\Omega_b/\Omega \gtrsim 0.1$) and efficient heating and enrichment of halo gas. They then approximately reproduce the characteristic luminosities of galaxies and of galaxy clusters, the observed relations between galaxy properties, and the kind of bias needed to reconcile $\Omega = 1$ with the observed kinematics of galaxy clustering. However, the amplitude of this bias is too small, and additional sources of bias must be invoked. Our luminosity functions contain significantly more faint galaxies than are observed. This is a serious discrepancy which may be alleviated by starbursts in dwarf galaxies, by selective merging of such systems, and by observational selection against low surface brightness dwarfs. Successful models form their stars late, typically more than half of them since $z = 1$, making the epoch of galaxy formation easily accessible to observation.

Subject headings: galaxies: clustering — galaxies: formation — galaxies: stellar content — galaxies: structure

1. INTRODUCTION

Over the last few years the study of galaxy formation has become an observational science. For a long time quasars were the only known objects beyond redshift 1, and their relationship to galaxies and galaxy formation was very uncertain. There are now at least 20 galaxies with measured redshifts exceeding 1.5, some almost as distant as the most distant known quasars. Although almost all of these are strong radio sources, their spectra appear to be dominated by starlight, and in many cases they show the presence of large and distributed star-forming regions (Spinrad 1989). The interpretation of these spectra in terms of stellar populations is controversial (see Spinrad 1989; McCarthy et al. 1987; Lilly 1988; Chambers, Miley, & Joyce 1988; Chambers & Charlot 1990), but it seems clear that they are undergoing star formation episodes which are much more vigorous than those seen in nearer radio galaxies. Some of the properties of these systems are quite similar to those expected for a protogalaxy in the later stages of formation (e.g., Baron & White 1987).

Evidence for substantial evolution of the galaxy population in the recent past has also been accumulating. It is more than 10 years since Butcher & Oemler (1978) discovered that some galaxy clusters at relatively modest redshift ($z \sim 0.3$) contain a population of blue galaxies not present in their nearby counterparts. Recent work (e.g., Gunn 1989) has shown that this blue population is heterogeneous and reflects higher levels both of star formation and of nuclear activity in the recent past. The counterpart of this phenomenon in the general field seems to be the excess of blue galaxies found in galaxy counts at faint magnitude levels (Peterson et al. 1979; Tyson & Jarvis 1979; Kron 1980; Ellis 1987; Tyson 1988). In particular, the total flux observed from faint blue galaxies in the very deep CCD counts of Tyson (1988) is so large that the associated star formation may account for the production of a substantial fraction of the entire heavy-element content of galaxies, and perhaps also for a similar fraction of their stellar content (Cowie 1988). This population may therefore be the long-sought population of "primeval" galaxies, although deep spec-

[1] Present address: Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK.

troscopic surveys have so far found most blue objects at brighter magnitudes ($m_J \leq 22.5$) to be relatively normal systems at modest redshift and with elevated star formation rates (Colless, Ellis, & Taylor 1989; Colless et al. 1990).

Absorption lines in quasar spectra are a window on properties of the intergalactic medium, and may also teach us something about galaxy formation. The damped Lyα systems at $z \sim 2$–3 studied by Wolfe (1989) are particularly interesting in this context. The column densities of these clouds appear quite similar to those of present H I disks in galaxies (although their metal and dust content may be lower; Pettini, Boksenberg, & Hunstead 1989; Fall & Pei 1989). However, the total amount of neutral gas they contain can be calculated directly from Wolfe's observations, and exceeds the abundance of neutral gas in nearby galaxies by an order of magnitude. Indeed, the abundance Wolfe finds is comparable to the present mass density in stars. An intriguing possibility is thus that the observed clouds contain the raw material for most of the stars in galaxies. Although this material has been assembled into large cool units by the time it is observed, it has apparently still to be converted into stars. Further indirect evidence that the universe may be very active on galactic scales at $z \sim 2$ comes, of course, from the strong evolution of the quasar and radio galaxy populations, which reach their peak volume densities at this epoch (see, e.g., Green 1989).

This rapidly changing observational situation demands a critical reevaluation of theories of galaxy formation. There has been substantial progress in constructing such theories over the last decade, but comparisons with observation have focused mainly on current properties of the galaxy population and of large-scale structure. There is now a need for evolutionary predictions to compare with observed objects at high redshift. In addition, there is a need for more detailed models of galaxy formation to exploit the rapidly growing body of information on the multicomponent nature of stellar populations in our own Galaxy, and so, by extension, in other spirals (see Gilmore, Wyse, & Kuijken 1989). Both these problems involve severe difficulties. In principle, they require understanding the dynamics of turbulent, multiphase, self-gravitating, and star-forming clouds in which cooling, radiative heating, and shock heating are all of major importance. In addition, they require quite detailed modeling of the stellar populations which form, and of the radiative transfer effects which determine their observable emission. *Ab initio* treatment of these processes is not feasible, and real progress is likely only if heuristic models are built which maintain close contact with observation. The present paper combines recent numerical and analytic work on the evolution of structure and on the evolution of stellar populations to derive "plausible" predictions for the observable properties of young galaxies. We focus specifically on the major uncertainties inherent in any such prediction. Although we work within the cold dark matter model, our methods can be applied to other currently popular models; many of the sources of uncertainty are common to all models.

As a theory for galaxy formation, the cold dark matter (CDM) model continues a major line of research from the 1970s. Almost single-handedly, Peebles (1970, 1973, 1974, 1980; Peebles & Dicke 1968) established the hierarchical clustering or "isothermal" scenario in which structure builds up through the continual aggregation of nonlinear objects into larger and larger units. A theory for the development of the mass distribution in this model was presented by Press & Schechter (1974) and tested using the first published cosmological N-body simulations. The large program of Aarseth, Gott, & Turner (1979), following earlier studies of individual clusters (Peebles 1970; White 1976), showed how such simulations could be used both to follow hierarchical clustering from quasi-linear initial conditions and to present the results in a way that allowed direct comparison with observation. An independent advance was the realization that the characteristic stellar mass of galaxies could be viewed as arising from the cooling properties of protogalactic gas clouds (Binney 1977; Rees & Ostriker 1977; Silk 1977). A final ingredient was the hypothesis that dark matter was present in substantial quantities around individual galaxies as well as in clusters and groups of galaxies (Ostriker, Peebles, & Yahil 1974; Einasto, Kaasik, & Saar 1974). White & Rees (1978, hereafter WR) combined these elements in a unified scheme; a dominant dissipationless background of dark matter of unspecified type was assumed to cluster hierarchically, carrying with it gas which cooled and condensed to form galaxies in the cores of heavy halos. Further work within this picture showed it to account both for the angular momentum structure of spiral and elliptical galaxies (Fall 1979; Fall & Efstathiou 1980; Barnes 1990) and for the overall scaling properties of the galaxy population (Faber 1982; Gunn 1982). WR calculated a galaxy luminosity function for their model and found it to predict the right characteristic luminosity but too many faint galaxies, a problem which carries over to the present extension of their theory.

This body of work led to the conclusion that random phase fluctuations with a power-law index $n = -1$ (in $|\delta_k|^2 \propto k^n$) were needed to explain the scaling properties of galaxy clusters (e.g., Fall 1979), and that $n = -2$ could explain the scaling properties of galaxies (Faber 1982; Gunn 1982). Thus, when the CDM power spectrum was derived (Peebles 1982; Blumenthal & Primack 1983) and found to have just these properties, it provided a beautiful a priori justification for the initial conditions which had previously been postulated ad hoc to fit the observations. A synthesis of these earlier galaxy formation ideas was included by Blumenthal et al. (1984) in their presentation of the CDM model. An explicit demonstration that the CDM power spectrum leads to clustering very like that observed was provided by the N-body simulations of Davis et al. (1985) and White et al. (1987a, b). The latter studies also incorporated the notion of biased galaxy formation. This grew out of Kaiser's (1984) explanation of the strong clustering of Abell clusters through the statistical properties of high peaks in Gaussian noise. It was developed collaboratively during a 1984 workshop at the Institute for Theoretical Physics in Santa Barbara, and published almost simultaneously by a number of workers (Davis et al. 1985; Bardeen et al. 1986; Bardeen 1986; Kaiser 1986; White 1986). The main attraction of the idea is that it may reconcile observation with the theoretical "imperative" that $\Omega = 1$. Throughout this paper we shall assume that $\Omega = 1$, and our models therefore require bias, perhaps arising through the "natural" mechanism described by White et al. (1987a) and further investigated by Frenk et al. (1988) and Carlberg & Couchman (1989). It is still uncertain whether this mechanism can work and whether its predictions are consistent with observation (Cole & Kaiser 1989; White, Tully, & Davis 1988; Eder et al. 1989). Our methods allow us to estimate how such bias affects the M/L ratios of rich clusters. The results suggest that it may need to be enhanced by dynamical segregation effects during cluster formation (Barnes 1985; Evrard 1987; West & Richstone 1988; Carlberg 1991).

Studies of galaxy formation within the CDM model have so

far been rather limited. One approach has been to identify the locations and characteristic internal velocities of forming galaxies with those of dark halos in N-body simulations. This allows one to verify that the CDM model can produce the correct abundance of objects as a function of circular velocity, and to measure the bias in their spatial distribution (Frenk et al. 1985, 1988; White et al. 1987a). It also gives considerable insight into the dynamical environment in which the galaxies must form. However, it does not permit the calculation of galaxy luminosities, nor does it allow an investigation of the overmerging problem first discussed by WR. This problem arises because both simulation results and analytic arguments suggest that the dark halos of galaxy clusters should have little substructure; galaxies must therefore be able to avoid merging as clusters form and their halos amalgamate. Inclusion of a dissipative component in the simulations permits a crude treatment of both these problems, and results from a first investigation by Carlberg & Couchman (1989) are very encouraging. However, cooling processes can only be treated schematically in this kind of work, and no attempt has so far been made to treat the stellar populations of galaxies (and so their observable properties) in a realistic way.

More detailed cooling arguments were applied to the CDM model by Blumenthal et al. (1984), whose assumptions about galaxy formation parallel those of WR. In particular, they assumed that perturbations of gas and CDM virialize as *homogeneous* spheres. A single cooling time then applies to each system; if it is sufficiently short, the gas turns into stars, otherwise no galaxy forms. Blumenthal et al. did not attempt to describe the time development of galaxy formation or to calculate a galaxy luminosity function. However, a first analysis of these problems was given by WR for their hierarchical clustering models. They found that cooling gas had to make stars with an efficiency well below unity for their models to work; following Larson (1974a), they assumed this efficiency to be proportional to the depth of the galactic potential well. Even simpler models for galaxy formation in a CDM universe have been considered by Schaeffer & Silk (1985) and by Evrard (1989); these avoid any explicit treatment of cooling and instead investigate the consequences of simple ad hoc rules for galaxy formation. None of this work addresses the influence of halo structure on cooling and star formation rates; in fact, the cooling time will be strongly position-dependent in a real halo, and at least some cooling is expected in the densest parts of every system. In addition, these early models did not consider the stellar populations in galaxies, and so made no evolutionary predictions that can be compared with recent faint galaxy data. A first approach to these problems came with Baron & White's (1987) study of galaxy formation within "generic" hierarchical clustering. This work investigated the general appearance of forming galaxies, but did not attempt specific predictions for a CDM universe.

The present paper extends the methods of WR to study galaxy formation in a flat CDM universe. We begin by presenting a simple model for the nonlinear evolution of the population of massive halos, which we justify by comparison with earlier N-body work. This model predicts the abundance and internal structure of dark halos as functions of mass and time, as well as the way in which the halos merge with each other as clustering proceeds. We then use the recent numerical work of Evrard (1989) and the similarity solutions of Bertschinger (1989) to motivate a model for the accumulation and cooling of gas within these halos. We find, as did WR, that feedback processes, perhaps associated with energy injection by supernovae, must be invoked to reduce the efficiency of cooling and star formation. Otherwise, much of the gas present turns into stars in small objects and at early times. We explore the consequences of assuming that these feedback processes lead, in addition, to chemical enrichment of the circumgalactic gas. This leads to predictions for the star formation rate in a halo as a function of its mass and of redshift. At this stage the theory can already be compared with observed star formation rates in galaxies and with X-ray data for their gaseous halos. Other aspects which we investigate at this stage are the metallicity-mass relations for galaxies, the metallicity of the intergalactic medium, and the effect of heavy-element cooling on the overall galaxy formation process. The final step is to use the stellar population models of Bruzual (1983) to convert star formation histories into luminosities and thus to follow the luminosity and color of the stars which form. This allows us to make direct contact both with the luminosity function of local galaxies and with observations of faint, high-redshift objects. Preliminary accounts of some of this work are given in White (1989, 1990). Related analyses have recently been presented by Cole (1991) and Lacey & Silk (1991).

A number of the processes we model are highly uncertain, and our treatment of them at best schematic. This applies particularly to the hydrodynamics of the strongly radiative and possibly multiphase gas, to the effect of feedback processes on its thermodynamic state and heavy-element content, to the relationships between local gas conditions, star formation rates and stellar initial mass functions, to the effects of dust on the observable emission from young stars, and to the merging history of stellar complexes subsequent to their formation. The combined uncertainties in the "predictions" of the CDM model from all these sources are enormous. We have therefore elected to study a variety of very simple models in order to get a first indication of the range and nature of the possibilities, and to try to isolate aspects of the data which are insensitive to some of the more uncertain elements. We do, in fact, find a number of relatively robust predictions for galaxy formation in a flat CDM universe. Most stars form quite late; feedback processes must be surprisingly efficient; an uncomfortably large baryon density is needed to get sufficiently bright galaxies to cool; substantial merging of intrinsically faint systems seems necessary to get an acceptable galaxy luminosity function; bias effects are too weak (without dynamical segregation) to give the low M/L ratios; and thus the high luminosities, of the most massive clusters, and dust must be relatively unimportant if we are to understand the observed counts of faint galaxies.

The layout of this paper is as follows. In § 2 we discuss our model for the dark matter distribution and the extent to which it can be justified by reference to the numerical experiments of White et al. (1987a), Frenk et al. (1988), and Efstathiou et al. (1988b). Section 3 discusses the accumulation and cooling of gas and the energy input from supernovae. Simple models for chemical enrichment are presented in § 4, and stellar population models are introduced in § 5. At this point we set values of the parameters in our models so that they give the correct luminosity density for the present universe. In § 6 we derive mass-luminosity-metallicity relations for galaxies. Employing a variety of assumptions, we also calculate luminosity functions for virialized systems, for galaxies, and for star-forming regions. The results are compared with data on nearby systems as well as with counts and redshift surveys of faint galaxies. Finally, § 7 tries to assess the status of the CDM model in the

light of these results, and to pinpoint the most promising avenues for further investigation.

2. THE EVOLUTION OF DARK HALOS

By hypothesis, the dominant component of a flat CDM universe is the dissipationless dark matter. The baryonic material influences the evolution of the mass distribution significantly only within the observed regions of galaxies, and N-body methods can be used to simulate the evolution of structure directly on all larger scales. Such simulations yield a wealth of information about the structure and evolution of dark halos (e.g., Frenk et al. 1985, 1988). For our purposes, we need to distill from this information the simplest analytic models which seem adequate to describe the structure, evolution, and statistics of the halos within which galaxies form. We begin by assuming these halos to be spherical even though the simulations show them to be typically quite elongated. We model their density profile by that of a singular isothermal sphere, an approximation supported by the data of Frenk et al. (1988) over the range 30–300 kpc for all but the most massive halos. The most appropriate halo "mass" parameter is thus $V_c = [GM(r)/r]^{1/2}$, since it is approximately independent of the radius at which it is measured. We obtain the abundance of such isothermal halos as a function of V_c and redshift z, by using the Press-Schechter (1974) argument in the form presented by Narayan & White (1988).

At some very early epoch, consider spherical regions of comoving radius r_0, in current units. The matter in such a region is assumed to be part of a single collapsed structure by redshift z, if its linear overdensity, $\delta = \delta\rho/\langle\rho\rangle$, extrapolated forward to that epoch, exceeds the critical value, δ_c. The probability of a random region satisfying this condition is

$$F(r_0, z) = \int_{\delta_c}^{\infty} \frac{1+z}{(2\pi)^{1/2}\Delta(r_0)} \exp\left[\frac{-\delta^2(1+z)^2}{2\Delta^2(r_0)}\right] d\delta , \quad (1)$$

where $\Delta(r_0)$ is the rms linear overdensity in a sphere of radius r_0, extrapolated to the present day. Press & Schechter suggest that the fraction of matter in clumps with masses corresponding to initial radii in the range $(r_0, r_0 + dr_0)$ can be obtained from this as

$$f(r_0, z)\, dr_0 = -2 \frac{\partial F}{\partial r_0} dr_0 . \quad (2)$$

The validity of this *Ansatz* is quite controversial, and, in particular, it is hard to justify the factor of 2 which was inserted purely as a fudge to ensure that all the mass is included in a clump of some size or another. However, the identical formula has recently been rederived by a different route which avoids many of the conceptual and mathematical problems of the argument presented here (Bond et al. 1991); as a result its theoretical underpinning now seems more secure.

For a spherical perturbation, each shell recollapses to the origin at a time when the mean linear overdensity interior to it extrapolates to the value 1.68. Narayan & White (1988) therefore chose $\delta_c = 1.68$, and assumed that the mean density of the nonlinear clumps to which the theory applies is $178\rho_0(1+z)^3$, where ρ_0 is the present critical density. This last expression is derived by assuming that the outermost mass shell virializes at the time of collapse at one-half of its maximum expansion radius (see, e.g., Peebles 1980). The mass and circular velocity of a clump are then related to its initial size and to redshift by

$$M = \frac{4\pi}{3}\rho_0 r_0^3 , \quad V_c = 1.67(1+z)^{1/2} H_0 r_0 , \quad (3)$$

where H_0 is the present Hubble constant. Multiplying equation (2) by the appropriate Jacobian and carrying out the differentiation gives the fraction of matter which is in halos of circular velocity V_c at redshift z:

$$f(V_c, z)\, dV_c = -\left(\frac{2}{\pi}\right)^{1/2} \frac{\delta_c(1+z)}{\Delta^2} \frac{d\Delta}{dV_c} \exp\left[\frac{-\delta_c^2(1+z)^2}{2\Delta^2}\right] dV_c , \quad (4)$$

where equation (3) is used to express Δ as a function of V_c. Dividing by the mass of a halo with circular velocity V_c, and multiplying by the mean mass density of the universe, gives the number density of halos as a function of V_c and z,

$$n(V_c, z) dV_c = \frac{-3(1.67^3)\delta_c H_0^3 (1+z)^{5/2}}{(2\pi)^{3/2} V_c^4 \Delta} \frac{d\ln\Delta}{d\ln V_c} \times \exp\left[\frac{-\delta_c^2(1+z)^2}{2\Delta^2}\right] dV_c . \quad (5)$$

This is the abundance per comoving volume, in current units.

The variance, Δ^2, which appears in equations (1), (4), and (5) is related to the CDM linear power spectrum through the convolution

$$\Delta^2(r_0) = \int_0^{\infty} 4\pi k^2\, dk\, |\delta_k|^2 W^2(kr_0) , \quad (6)$$

where

$$W(x) = 3(\sin x - x\cos x)/x^3 . \quad (7)$$

We take the power spectrum to have the form given by Davis et al. (1985). Extrapolated to the present day, and for $H_0 = 50$ km s^{-1} Mpc^{-1}, the value we shall adopt throughout this paper, this gives

$$|\delta_k|^2 = 1.94 \times 10^4 b^{-2} k(1 + 6.8k + 72k^{3/2} + 16k^2)^{-2} \text{ Mpc}^3 , \quad (8)$$

where the square of the "biasing parameter," b, is defined as the ratio of the variances of the galaxy and mass fluctuations within randomly placed spheres of radius $r_0 = 16$ Mpc (see, e.g., Bardeen et al. 1986). (Note that this definition of b differs from that of Narayan & White 1988, but is consistent with what is now the common usage in this field.) Evaluating the integral in equation (6) using this expression gives a result which can be approximated to within 10% over the range 0.05 Mpc $< r_0 <$ 40 Mpc by

$$\Delta(r_0) = 16.3 b^{-1}(1 - 0.3909 r_0^{0.1} + 0.4814 r_0^{0.2})^{-10} . \quad (9)$$

This gives $\Delta(16\text{ Mpc}) = 1$ for $b = 1$, the traditional "unbiased" normalization of the power spectrum based on the fact that the observed galaxy density has unit variance in a sphere of this radius. The N-body models of Davis et al. (1985), Frenk et al. (1985, 1988), and White et al. (1987a, b) have a power spectrum amplitude corresponding to $b = 2.6$. More weakly biased models have been advocated by a number of workers. The most convincing arguments for smaller b come from attempts to fit the CDM model to the large-scale flows discussed by Lynden-Bell et al. (1988) and by Faber & Burstein (1988).

Kaiser & Lahav (1989) argue that these data can be accommodated for $b \sim 1.5$, although others—for example Groth, Juszkiewicz, & Ostriker (1989)—believe that no CDM model can be reconciled with them. This question is largely decoupled from that of evolution on galactic scales. Throughout this paper we will present results for $b = 2.5$ and for $b = 1.5$ in order to illustrate the effects of altering the normalization of the CDM power spectrum.

Equations (5) and (9) provide a simple model for the abundance of isothermal CDM halos at each redshift. Notice that all the free parameters have been set a priori. It is clearly important to check that the result is an acceptable fit to simulations of nonlinear clustering. For the 375,000 Mpc^3 surveyed by the N-body models of White et al., our equations predict 5057, 682, 46, and 0.5 halos at $z = 0$, with circular velocities exceeding 100, 200, 400, and 800 km s^{-1}, respectively. The corresponding numbers of halos found in the simulations are 3806, 564, 50, and 1. Similarly, for the 25,000 Mpc^3 surveyed at higher resolution by Frenk et al. (1988), our Press-Schechter theory predicts 1116, 337, 45, and 3 halos with circular velocities exceeding 64, 100, 200, and 400 km s^{-1}; the numbers found were 943, 329, 46, and 4. Unfortunately, this remarkably good agreement appears not to extend to earlier epochs.

The halo abundances shown in Figure 11 of Frenk et al. (1988) are considerably smaller than we would predict for $z \geq 1$. On the other hand, Efstathiou & Rees (1988) used these same simulations (among others) to compare the abundance of halos as a function of *mass* to a slightly different version of Press-Schechter theory, and found good agreement at all redshifts. This apparent contradiction arises because the masses assigned by the Davis et al. (1985) "friends-of-friends" group finder (which was also used by Efstathiou & Rees) are not related to the circular velocities assigned by the Frenk et al. algorithm through equations (3). In fact, with a constant comoving linking length, the group finder bounds halos at systematically lower overdensity (and so higher mass for a given circular velocity) at earlier times. This appears to reflect the changing slope of the CDM power spectrum. In addition, Frenk et al. measured V_c at a fixed absolute density, corresponding to a much lower relative overdensity at high redshift. Their halo rotation curves are gently declining in the relevant radial range, so at early times their circular velocites underestimate the values at an overdensity of 178, the overdensity appropriate for our Press-Schechter formalism. Finally, some of the small halos found by the group finder are not counted in the circular velocity diagrams because they lie within the outer boundary of a larger system. These effects produce the substantial apparent difference between the abundances of Efstathiou & Rees and of Frenk et al. However, the difference is largely illusory. If V_c is taken as the *maximum* circular velocity of a halo (out to an overdensity of 50, say), the N-body results at high redshift agree much better with our Press-Schechter theory.

Because of the rather sketchy justification for the assumptions made in deriving it, the Press-Schechter formalism has been repeatedly questioned (e.g., Peacock & Heavens 1989). As noted above, their final formula has gained considerably in plausibility as a result of the alternative derivation provided by Bond et al. (1991). In addition to the original tests of Press & Schechter (1974) and the CDM tests described above, this formula has been checked against N-body simulations from a variety of initial conditions by Efstathiou et al. (1988b). These experiments show surprisingly good agreement over the part of the mass distribution containing the bulk of the total mass. Nevertheless, small systematic deviations from the predictions are clearly visible. The most vulnerable part of the theory may be its predictions of the low-mass behavior of the mass distribution; the power-law extension of this distribution to small objects is not well established by the N-body work. However, it is important to realise that the slope of this power law does *not* determine the corresponding slope of the faint end of the galaxy luminosity function. According to the theory of WR, the latter is set by the binding energy of the typical halos present at early times. This determines how effectively supernova feedback can suppress star formation in these objects which are the main formation sites for low-luminosity galaxies. Formation of faint objects in the *present* low-mass tail of the halo mass distribution is predicted to be relatively unimportant. Thus WR give a "cluster" luminosity function, $n(L)dL \propto L^{-a}dL$ for small L, where the exponent, a, is that of the Press-Schechter mass function and equals $(9 - n)/6$ for $|\delta_k|^2 \propto k^n$, whereas their "galaxy" luminosity function is $n(L)dL \propto L^{-b}dL$ with $b = (13 - n)/(7 - n)$ at faint luminosities. Over the relevant range, $1 \geq n > -3$, the two exponents vary with n in opposite senses.

The evolution of halo abundance predicted by equations (5) and (9) is shown in Figure 1. The two panels of this diagram refer to the power spectrum normalizations, $b = 2.5$ and $b = 1.5$, for which we will present results throughout his paper. (Note that these plots correct similar plots in White 1989 which were in error. None of the later results of that paper were affected by this plotting mistake.) The abundance of halos of each circular velocity initially rises steeply as such objects begin to form. It reaches a peak over a rather broad redshift range, and thereafter declines as halos merge to form more massive systems. For $b = 2.5$ the abundance of halos with $V_c = 400$ km s^{-1} is peaking today, that of 200 km s^{-1} halos peaked at $z \sim 1$, and that of 100 km s^{-1} halos peaked at $z \sim 3$. For the larger initial fluctuation amplitude, $b = 1.5$, all these vents are shifted to higher redshift, the abundance of 200 km s^{-1} halos peaking, for example, at $z \sim 3$. [The value of $(1 + z)$ at which structures are predicted to form on a given scale varies as b^{-1}.] In the theory of WR it is effectively the peak of these curves which determines the present abundance of galaxies with the corresponding circular velocity. A similar assumption was adopted as an *Ansatz* in recent work by Cole & Kaiser (1989).

The final aspect of dissipationless clustering which we will need to model is the merging of halos with time. This is necessary in order to identify the systems which currently contain stars formed at high redshift. There are two aspects to this problem. The first is to identify the current halos which contain old stars. This requires a model for *halo* merging, which we base on an extension of the Press-Schechter theory due to Bond et al. (1991) and Bower (1991). This allows us to construct a luminosity function for nonlinear galactic systems, by which we mean isolated galaxies, galaxy groups, and galaxy clusters; it also allows us to estimate their mass-to-light ratios, gas-to-star ratios, and mean metallicities. The second and more difficult problem is to identify the individual *galaxies* which contain stars formed at earlier epochs. This requires an understanding of when and how galaxies merge as their halos coalesce. The simplest extreme assumption, following WR, is that galaxy merging is negligible; as we shall see, this appears to lead to an unacceptable luminosity function.

The time evolution of the merging process in hierarchical

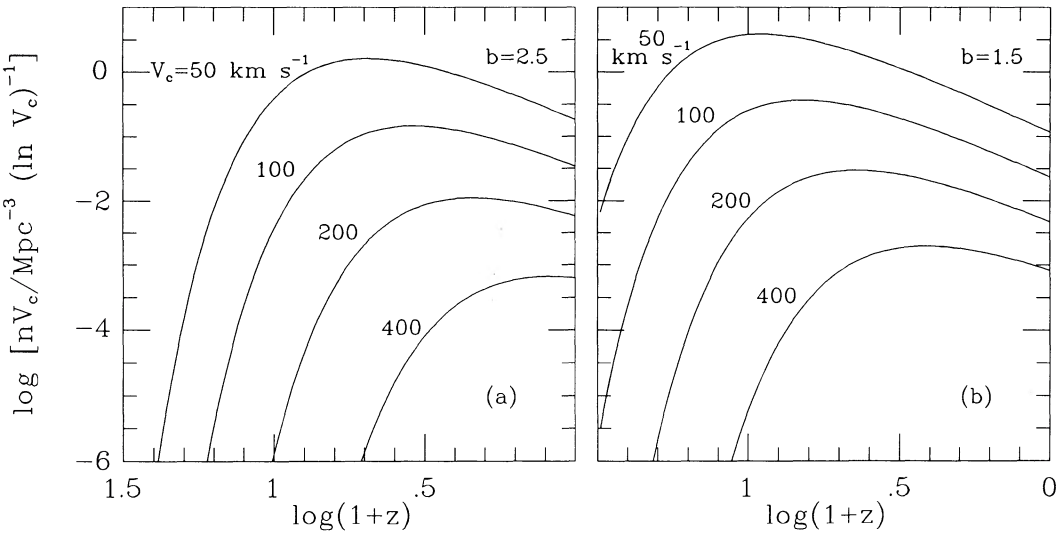

FIG. 1.—Abundance of dark matter halos as a function of redshift and circular velocity. These abundances were calculated using the Press-Schechter theory (eq. [5]) for cold dark matter models (eq. [9]). Each curve is labeled by the circular velocity (in km s^{-1}) of the population. (*a*) Biasing parameter $b = 2.5$ (*b*) Biasing parameter $b = 1.5$.

clustering has been studied in some detail by Frenk et al. (1988) and Efstathiou et al. (1988b). The statistics of halo merging are complex, but many aspects of it are well modeled by the following simple extension of the Press-Schechter theory. Consider a region of initial comoving radius r_1, entirely contained within a larger region of radius r_0. Bower (1991) shows that the probability that these two regions will have mean linear overdensities δ_1 and δ_0 is given by

$$f(\delta_1, \delta_0)d\delta_1\, d\delta_0 = \frac{(1+z)^2\, d\delta_1\, d\delta_0}{2\pi\Delta_0(\Delta_1^2 - \Delta_0^2)^{1/2}} \times \exp\left\{-\frac{(1+z)^2}{2}\left[\frac{(\delta_1-\delta_0)^2}{\Delta_1^2-\Delta_0^2} + \frac{\delta_0^2}{\Delta_0^2}\right]\right\}. \quad (10)$$

where Δ_1 and Δ_0 are given by equation (9) evaluated at r_1 and r_0, respectively. This formula assumes that all points in the interior of the larger volume are *equally* likely to be contained in the smaller volume; this is a plausible but highly nontrivial assumption which is responsible for the simplicity of the result. An exactly equivalent formula was derived by Bond et al. (1991) through an entirely different chain of reasoning. Given equation (10), an argument directly analogous to that which gave equation (4) produces an expression for the fraction of matter which is in clumps of circular velocity V_1 at redshift z_1, and later in clumps of circular velocity V_0 at $z_0 < z_1$:

$$f(V_1, V_0, z_1, z_0)\, dV_1\, dV_0 = \frac{2\delta_c^2\, \Delta_1(z_1 - z_0)(1+z_0)}{\pi(\Delta_1^2 - \Delta_0^2)^{3/2}\Delta_0^2}\frac{d\Delta_1}{dV_1}\frac{d\Delta_0}{dV_0} \times \exp\left\{-\frac{\delta_c^2}{2}\left[\frac{(z_1-z_0)^2}{\Delta_1^2-\Delta_0^2} + \frac{(1+z_0)^2}{\Delta_0^2}\right]\right\} dV_1\, dV_0\,, \quad (11)$$

provided that $V_1/(1+z_1)^{1/2} < V_0/(1+z_0)^{1/2}$, corresponding to the requirement $r_1 < r_0$. When integrated over V_1 or V_0, this reduces to equation (4) at z_1 or z_0, respectively. One can combine this expression with equation (4) to get the conditional probability densities,

$$f(V_0, z_0 \mid V_1, z_1) = f(V_1, V_0, z_1, z_0)/f(V_1, z_1)\,, \qquad f(V_1, z_1 \mid V_0, z_0) = f(V_1, V_0, z_1, z_0)/f(V_0, z_0)\,. \quad (12)$$

The first of these gives the probability that material which was in a halo of circular velocity V_1 at z_1 will end up in a halo with V_0 at z_0, while the second is the probabiity that material which is in a halo with V_0 at z_0 had previously been in a halo with circular velocity V_1 at z_1. One can show that these expressions obey the integral relations

$$f(V_0, z_0 \mid V_1, z_1) = \int_0^\infty f(V_0, z_0 \mid V_2, z_2) f(V_2, z_2 \mid V_1, z_1) dV_2\,, \qquad f(V_1, z_1 \mid V_0, z_0) = \int_0^\infty f(V_1, z_1 \mid V_2, z_2) f(V_2, z_2 \mid V_0, z_0) dV_2\,, \quad (13)$$

when $z_1 > z_2 > z_0$. Thus this model has all the analytic properties required of a consistent picture for the merging of halos. Furthermore, Bower (1991) and Bond et al. (1991) show that it also gives a good fit to N-body data on halo merging. We shall use this model to assign the stars and the metals formed at each time to the halos present at later times. This allows us to follow the depletion and enrichment of gas as star formation proceeds. We will, however, require further assumptions to model the merging of galaxies.

3. COOLING AND STAR FORMATION

Beyond a redshift of 100 or so, we expect a fraction, Ω_b, of the universe to be in the form of cool, near-uniform gas. Once structures with $V_c > 10$ km s^{-1} begin to appear, the associated gas will be collisionally ionized by the shocks associated with collapse and virialization. There are then two efficient mechanisms by which it can cool and collapse further relative to its dark matter halo. The first is inverse Compton scattering of microwave background photons. At redshifts exceeding 10, this process can remove the thermal energy of the gas on time

scales shorter than the age of the universe (e.g., Peebles 1971). The second process is radiative cooling by bound-bound and bound-free transitions; we find below that this is also efficient enough at these epochs to produce very short cooling times. As a result we will make no error by neglecting inverse Compton cooling in the analysis that follows. At recent times halos form with high enough equivalent virial temperature ($kT > 2$ keV) for free-free transitions (thermal bremsstrahlung) to dominate the cooling. This is the relevant regime for the X-ray–emitting gas in rich galaxy clusters. Unfortunately, the most critical temperature regime for galaxy formation is one to two orders of magnitude cooler, where the cooling rate is quite sensitive to the assumed heavy-element content. As a result we will find that our models depend significantly on the way we treat protogalactic chemical enrichment. The relatively high metal abundance of the X-ray gas in clusters is a clear indication that such enrichment is unlikely to be negligible. (See Sarazin 1986 for a comprehensive review of the intracluster medium.)

The properties of rich clusters also provide us with a direct argument for the appropriate value of Ω_b to insert in our models. Gas cooling times are considerably longer than the age of the universe throughout the main body of these systems. Thus, the mass ratios of gas, dark matter, and stars not only are directly measurable but are plausibly unchanged since the epoch of cluster formation. The Coma Cluster is perhaps the best-observed rich cluster. Within $1h^{-1}$ Mpc the optical and X-ray data are reliable, and can be used to derive masses without extrapolation and with little model dependence. The gas constitutes about $4.6h^{-1.5}$% and the galaxies about 6% of the total mass. The former number is taken from the analyses of The & White (1986, 1988), Merritt (1987), and Hughes (1989) and is unlikely to be in error by more than 30%; the latter uses, in addition, a mean stellar M/L taken from Table 4.2 of Binney & Tremaine (1987) and is accurate to better than about 50%. In our models Ω_b is identified with the ratio of baryons to dark matter within the regions of halos from which cooling takes place. Since we adopt $\Omega = 1$ and $H_0 = 50$ km s^{-1} Mpc^{-1}, the Coma Cluster data (which are in no way atypical of clusters in general) suggest that we should take $\Omega_b \sim 0.19$.

While this value of Ω_b is only marginally inconsistent with older bounds from the standard theory of cosmic nucleosynthesis (e.g., Yang et al. 1984), more recent analyses suggest that the maximum allowed baryon density could be as small as 6% for our chosen value of H_0 (Olive et al. 1990). The cluster data certainly do not allow such a small ratio of baryons to dark matter. This apparent discrepancy could be eliminated in at least three ways. A baryon density $\Omega_b \simeq 0.19$ is allowed if primordial nucleosynthesis occurred in a weakly inhomogeneous medium (Applegate & Hogan 1985; Kurki-Suonio et al. 1990). Alternatively, one could abandon the assumption that $\Omega = 1$ and choose, for example, $\Omega = 0.2$ and $\Omega_b = 0.04$, so that $\Omega_b/\Omega = 0.2$, the ratio suggested by the Coma data. The flat geometry predicted by the inflationary model could then be, rather inelegantly, preserved by introducing a cosmological constant. Finally, one could imagine that baryons are, for some reason, substantially overabundant in clusters. For example, cooling might concentrate gas toward the centers of dark halos at early stages of the clustering hierarchy, and this separation might be partially preserved when the halos later merge to form the observed clusters. This idea is supported by the inferred presence of cooling flows in the centers of many clusters, but appears contradicted by the fact that the cluster gas is, if anything, inferred to be less concentrated than the dark matter (Sarazin 1986; Hughes 1989). In the context of our models it also sits somewhat uncomfortably with the claim we make below that strong feedback effects are required to prevent excessive cooling and star formation at early epochs. On the other hand, we shall see that our models fail to produce bright galaxies for values of Ω_b as small as 0.05, so some concentration effect must be invoked if we are to avoid the conclusion $\Omega < 1$.

In the local universe the visible regions of galaxies contribute less than 1% of the closure density. If we adopt $\Omega_b \sim 0.19$ and assume that all star formation produces visible stars, then gas must be converted into stars with an overall efficiency well below 10%. This turns out to be a stringent constraint on our models. For such values of Ω_b (or, more generally, for such ratios of gas to dark matter) cooling is so effective in the dark halos present at moderate redshift that all their contained gas can cool. In this situation the supply of gas is limited by the overall accretion onto halos, rather than by cooling, and overproduction of stars (or at least of cold gas) can be avoided only by assuming that star formation pumps energy into the remaining halo gas with high efficiency.

Once significant cooling starts within a halo, it seems likely that it will continue and even accelerate as the gas shrinks to higher and higher density. This process can only plausibly be arrested by the formation of a gaseous disk or the formation of stars. Star-forming regions dump energy into the surrounding gas through photoionization, stellar winds, and supernova explosions, and may halt or even reverse the collapse. At late times it is very difficult to specify the state of the gaseous component; it might well have the multiphase structure engendered in the local interstellar medium by similar processes (e.g., McKee & Ostriker 1977). As each new structure collapses, the associated gas will be shocked, and in some circumstances it may be heated to the new virial temperature and effectively rehomogenized within its halo. However, cooling rates may be short enough for the multiphase structure to survive the shocks, and it is then unclear how the dynamics of the gas component should be modeled. Studies of the interstellar media of interacting galaxies, where similar physical conditions prevail, may provide some observational insight into this problem. In this paper we treat it in a way which clearly oversimplifies the dynamics, but which should be appropriate to obtain a plausible estimate of the star formation rate as a function of halo mass and of epoch.

We assume that, as a halo forms, the gas initially relaxes to an isothermal distribution which exactly parallels that of the dark matter. The hydrostatic equilibrium equation then relates the gas temperature to the circular velocity of the halo,

$$kT = \frac{1}{2}\mu m_p V_c^2 , \quad \text{or } T = 35.9\left(\frac{V_c}{\text{km s}^{-1}}\right)^2 \text{K} , \qquad (14)$$

where μm_p is the mean molecular weight of the gas, which we fix by assuming a fully ionized gas which is 25% helium by mass. At each radius in this distribution we can then define a cooling time as the ratio of the specific energy content to the cooling rate,

$$t_{\text{cool}}(r) = \frac{3}{2}\frac{\rho_g(r)}{\mu m_p} kT \Big/ n_e^2(r)\Lambda(T) , \qquad (15)$$

where $\rho_g(r)$ is the gas density profile and $n_e(r)$ is the electron density. Substituting in our assumed density profile and gas

composition gives

$$t_{\rm cool}(r) = \frac{192\pi G m_p^2 r^2}{49 f_g \Omega_b \Lambda(\mu m_p V_c^2/2k)}, \quad (16)$$

where f_g is the fraction of the initial baryon density which remains in gaseous form. A cooling radius can now be defined as the point where the cooling time is equal to the age of the universe,

$$t_{\rm cool}(r_{\rm cool}) = \tfrac{2}{3} H_0^{-1}(1+z)^{-3/2}. \quad (17)$$

We distinguish two different cases. When $r_{\rm cool}$ is larger than the virialized region of a halo, cooling is so rapid that infalling gas never comes to hydrostatic equilibrium. The supply of cold gas for star formation is then limited by the infall rate rather than by cooling. When $r_{\rm cool}$ lies deep within the halo, the accretion shock radiates only weakly, a quasi-static atmosphere forms, and the supply of cold gas for star formation is regulated by radiative losses near $r_{\rm cool}$.

We consider the virialized part of a halo to be the region within which the mean overdensity is 200. Its radius and mass are therefore defined by

$$r_{\rm vir} = 0.1 H_0^{-1}(1+z)^{-3/2} V_c, \qquad M_{\rm vir} = 0.1 G^{-1} H_0^{-1}(1+z)^{-3/2} V_c^3. \quad (18)$$

This choice is motivated by the simple spherical collapse model discussed above (before eq. [3]). However, additional support comes from simulations of the collapse of rich clusters carried out with a hybrid N-body/smoothed particle hydrodynamics code (Evrard & Davis 1988; Evrard 1989). These models show the accretion shock to be close to, but perhaps a little outside, the radius given by equation (18). Thus, when $r_{\rm cool} \gg r_{\rm vir}$, we are in the accretion-limited case in which infalling gas never reaches hydrostatic equilibrium. We obtain an expression for the accretion rate by differentiating the second of equations (18) with respect to time, and multiplying by the fraction of the mass of the universe which remains in gaseous form:

$$\dot{M}_{\rm inf}(V_c, z) = 0.15 f_g \Omega_b V_c^{-3} G^{-1}. \quad (19)$$

Note that, except for the weak time dependence of f_g, this infall rate does not depend on redshift.

In the opposite limit, $r_{\rm cool} \ll r_{\rm vir}$, the halo will contain a quasi-static atmosphere of hot gas. Evrard's simulations provide some justification for our assumption that the distribution of this gas will parallel that of the dark matter. His models (which include *no* radiative energy losses) produce clusters in which the gas is nearly isothermal and is distributed very similarly to the dark matter. Cooling will cause a flow of gas into the central regions of such a halo in a manner exactly analogous to that envisaged for cooling flows in galaxy clusters (e.g., Fabian, Nulsen, & Canizares 1984). A simple expression for the inflow rate is

$$\dot{M}_{\rm cool}(V_c, z) = 4\pi \rho_g(r_{\rm cool}) r_{\rm cool}^2 \frac{dr_{\rm cool}}{dt} = \frac{3}{4} f_g \Omega_b H_0 (1+z)^{3/2} \frac{V_c^2 r_{\rm cool}(V_c, z)}{G}. \quad (20)$$

Despite the rather naive basis for this equation, exact similarity solutions for cooling flows by Bertschinger (1989) show that it is approximately correct for a rather wide range of halo and gas structures. For the particular isothermal model that we have adopted, the coefficient should, perhaps, be reduced by 28%. However, we will see that the true situation is probably more akin to a galactic fountain than to Bertschinger's single phase flow. The proper coefficient is therefore quite uncertain, and we have preferred to retain the simple assumption of equation (20). From equations (16) and (17) we see that

$$r_{\rm cool} \propto (f_g \Omega_b)^{1/2}(1+z)^{-3/4}.$$

Substituting this in equation (20), we find

$$\dot{M}_{\rm cool} \propto (f_g \Omega_b)^{3/2}(1+z)^{3/4};$$

the inflow rate due to cooling decreases with time, and is quite sensitive to the gas fraction.

The gas supply rates predicted by equations (19) and (20) are illustrated in Figure 2. To construct these diagrams, we used cooling functions, $\Lambda(T)$, appropriate to a pure H/He mixture (*a*) and to gas enriched with heavy elements (*b*). Our cooling functions are interpolated from Figure 9.9 of Binney & Tremaine (1987), but the actual metallicity and gas content of halos are evolved according to the detailed models we discuss in § 4. The particular models used for Figure 2 both start with $\Omega_b = 0.1$ and have fluctuation amplitude $b = 2.5$. The admixture of heavier elements increases the values of $\dot{M}_{\rm cool}$ substantially (up to a factor of 5) in halos with circular velocity in the range 50 km s^{-1} $< V_c <$ 500 km s^{-1}, where we expect the bulk of galaxy formation to occur. In any particular halo, the rate at which cold gas becomes available for star formation is not simply $\dot{M}_{\rm cool}$, but rather the minimum of $\dot{M}_{\rm cool}$ and $\dot{M}_{\rm inf}$. If the cooling time is shorter than the infall time ($\dot{M}_{\rm inf} < \dot{M}_{\rm cool}$), all the accreted gas can cool and no more than this can be available for star formation. Conversely, if the infall time is shorter than the cooling time ($\dot{M}_{\rm cool} < \dot{M}_{\rm inf}$), only the gas that can cool is accreted. For small halos and at early times, the gas supply is infall-limited. In the halos of present-day bright galaxies, on the other hand, the gas supply is regulated by cooling (unless $f_g \Omega_b$ is large and contamination by heavy elements is very efficient). In the following we assume that the rate at which gas is made available for star formation is min ($\dot{M}_{\rm cool}$, $\dot{M}_{\rm inf}$).

The predicted gas content of galaxy halos suggests an immediate test of the framework we have constructed so far. According to the models of Thomas et al. (1986), the bolometric X-ray luminosity of the region within the cooling radius of a galactic cooling flow is

$$L_X \sim 2.5 \dot{M}_{\rm cool} V_c^2. \quad (21)$$

(The coefficient here is only approximate, since it depends on the assumed structure of the cooling flow and of the galactic potential well.) The predictions of this formula are superposed on Figure 2. For large circular velocities the mass cooling rates correspond to quite substantial X-ray luminosities. This radiation will be emitted with a near-isothermal spectrum at a characteristic temperature given by equation (14). This is given on the upper horizontal axis in Figure 2. A present-day bright spiral with $V_c = 250$ km s^{-1} is predicted to have a halo temperature of only 0.19 keV and a cooling radius in the range 100–300 kpc, depending on the enrichment model. Such emission is too cool and too diffuse to have been detected easily by the *Einstein* satellite, but should be measurable with careful observations by *ROSAT*. [Note that the luminosity scales as $(f_g \Omega_b)^{3/2}$ and the surface brightness as $(f_g \Omega_b)^{1/2}$.] Bright ellipticals are thought to have massive halos with even larger values of V_c. Although some, like M87, do show the kind of X-ray

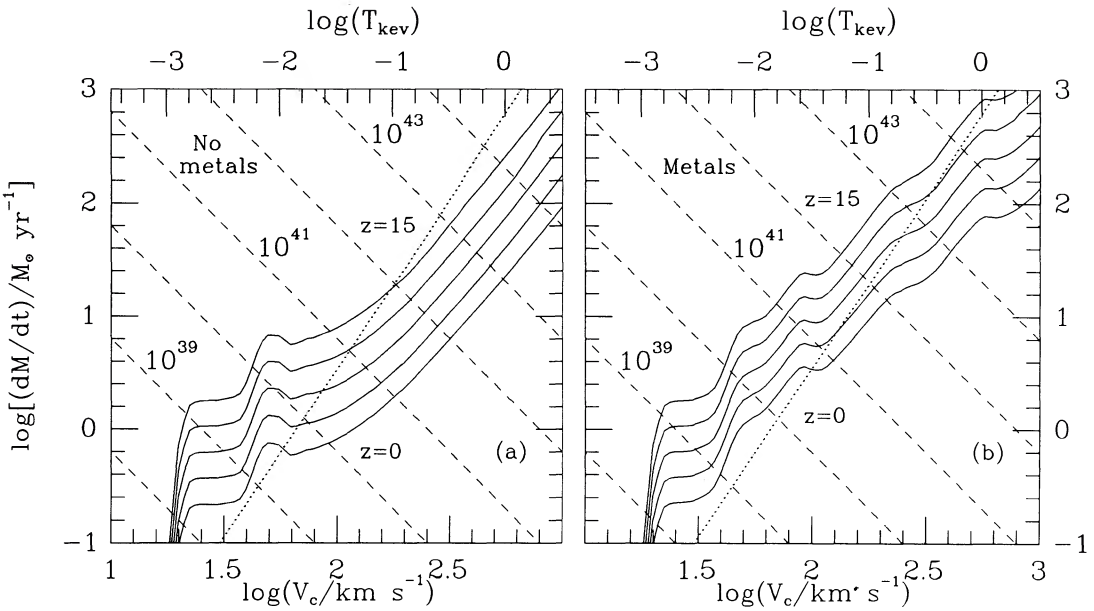

FIG. 2.—Gas infall rate and cooling rates in dark matter halos as a function of circular velocity and redshift. The infall rate (*dotted line*) is essentially independent of redshift; the cooling rates (*solid lines*) are given for redshifts $z = 0$, 1, 3, 7, and 15 (*from bottom to top*). Dashed lines give present-day X-ray luminosities in ergs s^{-1} produced by gas cooling at the given rate in each halo. The predicted temperature of this emission is given on the upper abscissa. (*a*) A cooling function for gas of zero metallicity is assumed. (*b*) A cooling function for gas enriched according to the models of § 4 is assumed. In both cases, $\Omega_b = 0.1$ and the biasing parameter b is 2.5.

properties we predict (Fabricant & Gorenstein 1983; Stewart et al. 1984), the emission from most implies a gas fraction much smaller than $f_g \Omega_b \sim 0.05$–0.1 (Thomas et al. 1986; Fabbiano 1989). We are therefore forced to argue that such galaxies have lost their original gaseous halos, either as a result of interactions with their environment or because early supernovae blew them away. Only the central ellipticals in clusters are at rest with respect to their environment and may have been able to retain "their" gas. Although this is quite plausible, it seems a slightly unsatisfactory patch for the model. A similar patch for spirals would be much less plausible, since their formation probably demands a relatively quiescent environment. Future data will provide an important test of our ideas. Note, however, that energy injection from current star formation may affect the gas structure in spiral halos (see below) and so complicate the interpretation of X-ray observations.

We can take the gas supply rates of equations (19) and (20) and the halo abundances of equation (5) and integrate forward from high redshift to find out what happens if all the cold gas is assumed to turn into stars. For $\Omega_b > 0.05$, much of the gas is used up well before the present. This is unacceptable, since the density contributed by the observed stars in galaxies is less than 1% of the critical value. This problem of overly rapid star formation was noted by WR, who solved it by invoking feedback processes along the lines of an earlier suggestion by Larson (1974b). Larson realized that the energy input from young stars and supernovae could easily drive all the gas out of a small protogalaxy before more than a small fraction of it had been converted into stars. He suggested that this inefficiency might account for the low metallicities of dwarf galaxies. WR found that it could also cure their gas consumption problem, and Dekel & Silk (1986) gave a detailed reworking of the arguments and applied them to the CDM model. The energy available from stellar sources can be estimated fairly reliably. Unfortunately, however, the effective heating of the gas is highly uncertain because it depends on the unknown efficiency with which the energy is deposited in the gas. The gas absorbs the radiative output of young stars very inefficiently; heating from stellar winds and supernovae may be much more effective, but there can be substantial radiative losses while the shocks propagate through residual cold gas surrounding the star-forming region. Such losses are very important in the local interstellar medium, but it is hard to anticipate their size in the much more active environment expected during galaxy formation. We are again forced to adopt an extremely simplistic model.

With a standard stellar initial mass function (IMF), about one supernova is expected for each 100 $M_\odot$ of stars formed. The kinetic energy of the ejecta from each supernova is about 10^{51} ergs. Thus, we write the effective heating rate of the halo gas due to star formation at rate $\dot{M}_*$ as

$$dE/dt = \epsilon_0 \dot{M}_* (700 \text{ km s}^{-1})^2 . \quad (22)$$

The single parameter ϵ_0 hides a multitude of sins, including uncertainties in the IMF, in the energy output of supernovae and stellar winds, and in the efficiency with which this energy is deposited in the hot halo gas. It is unlikely to exceed unity, and could be much smaller. Dekel & Silk (1986) found $\epsilon_0 \sim 0.02$ from a calculation of supernova remnant evolution in a uniform medium. Their detailed assumptions are difficult to justify, and there is observational evidence that supernova-driven winds in some starburst galaxies require much higher efficiencies (Chevalier & Clegg 1985; Heckman, Armus, & Miley 1990). The superwinds discussed by Heckman et al. are particularly relevant; their data suggest that almost all the supernova energy is escaping from the starburst regions. However the inferred temperature and speed of the winds ($\sim 10^8$ K and ~ 2000 km s^{-1}) seem rather large for the energy to be effectively absorbed by the outer halo gas in the way that we assume. For the time being we retain our very simple

model, we admit our inability to calculate ϵ_0, and instead we treat it as a free parameter to be fixed by requiring that our models reproduce the observed luminosity density of the universe. Note that in reality ϵ_0 may well be a function of $\dot{M}_*$, V_c, and z, rather than a constant. Equation (22) with its adjustable coefficient thus represents a significant and poorly justified assumption of our theory.

Stellar energy input will counteract radiative losses in the cooling gas and tend to reduce the supply of gas for further star formation. The energy dissipation rate associated with gas supply processes is $\sim \dot{M}_t V_c^2$, where, as argued above, gas is effectively supplied at a rate $\dot{M}_t = \min(\dot{M}_{\rm cool}, \dot{M}_{\rm inf})$. We will assume that star formation is self-regulating in the sense that $\dot{M}_*$ takes the value required for heating (eq. [22]) to balance dissipation in the material which does *not* form stars (i.e., $dE/dt = (\dot{M}_t - \dot{M}_*)V_c^2$). This produces the following model for the star formation rate:

$$\dot{M}_*(V_c, z) = \epsilon(V_c) \min(\dot{M}_{\rm cool}, \dot{M}_{\rm inf}) ,$$
$$\epsilon(V_c) = 1/[1 + \epsilon_0(700 \text{ km s}^{-1}/V_c)^2] . \quad (23)$$

For large V_c the available gas turns into stars with high efficiency because there is no energy source sufficient to prevent cooling and fragmentation; for smaller objects the star formation efficiency ϵ is proportional to V_c^2 (and is inversely proportional to the efficency ϵ_0 with which supernova energy is thermalized). However, it is possible, and perhaps even likely, that the cooling gas will often settle into a centrifugally supported disk and proceed to form stars on a longer time scale. To the extent that such cold gas eventually makes stars, this possibility does not much affect our model. Such a latency period is probably required if our model is to explain the large amount of cold gas observed at $z \sim 2$ in the form of damped Lyα QSO absorption systems (Wolfe 1989).

The assumption of self-regulation at small V_c seems plausible because the time interval between star formation and energy injection is much shorter than either the sound crossing time or the cooling time in the gaseous halos we are considering. In our model a forming galaxy is effectively a large-scale version of the self-regulating gas clouds invoked by Silk & Norman (1981) in their own dissipational model for galaxy formation. However, other possibilities can be envisaged. For example, Dekel & Silk (1986) suggested that supernovae not only would suppress cooling of the halo gas but would actually expel it altogether. (Note that this is not necessary for their arguments about the surface brightness and metallicity of dwarfs to be valid. A self-regulating model would produce the same results if external influences are assumed to remove the halo gas at some stage.) Ostriker & Cowie (1981) assumed that so many supernovae could go off at once that an explosive shock would be driven to large distances through the surrounding intergalactic medium. While we find these possibilities less plausible than the less energetic self-regulation we assume, they certainly cannot be excluded without more detailed modeling of the cooling and star formation processes. Note that they require *more* star formation than our model because more energy is required to expel the gaseous atmosphere to infinity than just to prevent it from cooling. We do allow for the additional possibility that self-regulation may induce bursting behavior rather than a uniform depression of the star formation rate. We model this by taking stars to form at the maximum rate allowed by the supply of gas, but only during bursts with a short duty cycle. (We do not need to assume a duration for the bursts.) Specifically, we assume that stars form at a rate $\dot{M}_{*,b}$ a fraction F_b of the time, where

$$\dot{M}_{*,b}(V_c, z) = \min(\dot{M}_{\rm cool}, \dot{M}_{\rm inf}) , \quad F_b = \epsilon(V_c) . \quad (24)$$

Averaged over time, equations (23) and (24) obviously give the same star formation rates. However, active systems will be rarer and brighter if star formation takes place in a bursting mode.

It is interesting to compare the star formation rates predicted by equations (23) and (24) with those observed in spiral galaxies. To make definite predictions, we must adopt values for Ω_b, for ϵ_0, and for the metallicity of the cooling gas. Since we use ϵ_0 to set the total stellar luminosity density of our models, it turns out to depend on all of our assumptions including those concerning chemical enrichment and stellar populations which we have not yet discussed. In Figure 3 we show predicted star formation rates at various epochs for six models, two in which heavy elements are assumed never to mix into the cooling gas and four which employ the favored mixing model of § 4 below. We give results for our two standard values of the bias parameter b and for $\Omega_b = 0.05$, 0.1, and 0.2. In each case ϵ_0 is chosen to reproduce the observed luminosity density; the required value increases with metallicity, with fluctuation amplitude, and with Ω_b, and reaches quite large, perhaps implausible values. For each model the solid lines correspond to the nonbursting case (eq. [23]); star formation rates for the bursting case during its *on* state (eq. [24]) are given by the dashed lines (cooling rate) or by the straight dotted line (infall rate), whichever is smaller. Notice that at almost all V_c the star formation rates are inferred to be higher at early epochs. Notice also that we predict feedback effects to be important in the formation of all but the most massive galaxies for which $\dot{M}_* \simeq \dot{M}_{\rm cool}$. (This is indicated by the fact that the solid lines always lie below the corresponding dashed lines.) Only if Ω_b is small and enrichment effects are weak does our required efficiency resemble the values advocated by Dekel & Silk (1986). Notice that the dependence of star formation rate on fluctuation amplitude b is quite weak when the efficiencies are set to produce the observed mean luminosity density. The star formation curves in Figure 3 are all substantially steeper than the infall line, which is itself proportional to halo mass (see eqs. [18] and [19]). This means that stars form preferentially in the most massive systems. The result is a bias which goes part way toward reconciling a flat universe with the observed mass-to-light ratios of galaxy clusters. We discusss this process in more detail below.

Kennicutt (1983) gives estimated star formation rates for spiral galaxies with a wide range of luminosities and Hubble types. We plot his data on top of the predictions in Figure 3, using a standard Tully-Fisher relation to convert from M_B to V_c (Pierce & Tully 1988). It is interesting that these observed rates are quite comparable to those we predict for the nonbursting model. If anything, the observed rates are somewhat high. Thus, it seems that cooling halos of the kind we are discussing can indeed provide sufficient infall of cold gas to solve the apparent depletion problem in the disks of many spirals (e.g., Larson, Tinsley, & Caldwell 1980; Kennicutt 1983). Only the smallest galaxies have star formation rates approaching the predictions of our bursting models. This accords with the common claim that bursting behavior is most prevalent in low-mass, irregular systems but is less important in large galaxies (Gerola, Seiden, & Schulman 1980; Thuan

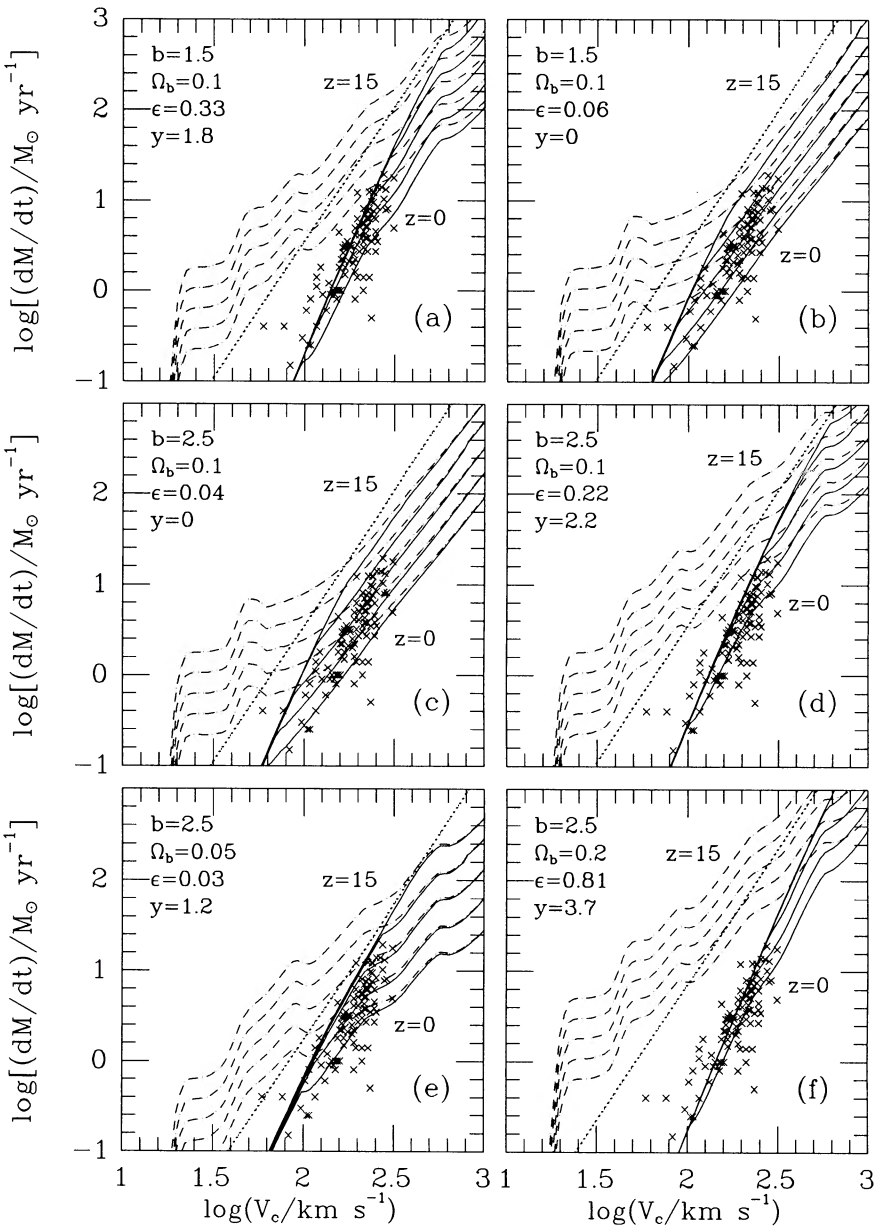

FIG. 3.—Star formation rates in dark matter halos as a function of circular velocity and redshift. Solid lines give the star formation rates in models with continuous star formation (eq. [23]). The corresponding rates in bursting models (eq. [24]) are the minimum of $\dot{M}_{\rm cool}$ (*dashed line*) and $\dot{M}_{\rm inf}$ (*dotted line*). In each case, results are given for redshifts $z = 0$, 1, 3, 7, and 15 (*from bottom to top*). Different panels correspond to different combinations of model parameters as shown. The biasing parameter b and the efficiency ϵ are defined in eqs. (8) and (23), respectively; the parameter y is the yield discussed in § 4. The crosses give star formation rates measured by Kennicutt (1983) for present-day spirals.

1985; Struck-Marcell & Scalo 1987; Arimoto & Tarrab 1990). Notice also that for large Ω_b and efficient enrichment the predicted star formation rates in halos with $V_c \sim 500$ km s^{-1} approach 100 $M_\odot$ yr^{-1} and seem too large to be consistent with observation. This problem may be related to the observed lack of OB stars in present-day cooling flows in clusters; in these systems the initial mass function seems to differ substantially from that in the solar neighborhood. In our models it is tempting to identify this difference with the transition between halos in which feedback effects dominate (low V_c) and ones in which feedback is unimportant (high V_c).

4. CHEMICAL ENRICHMENT MODELS

Figures 2 and 3 show that the enrichment of the gas in galaxy halos can have a substantial effect on cooling rates and so on the inferred history of star formation. Since our models require substantial heating of the halo gas by star formation activity, it seems plausible that substantial amounts of processed material could be mixed into gas out to the cooling radius. There are at least two observational indications that this kind of process does indeed occur. The [Mg II] absorption-line systems observed at relatively low redshift in

quasar spectra seem almost always to be associated with bright, actively star-forming galaxies (Bergeron 1988). These enriched clouds are typically seen several optical diameters away from the galaxy center, suggesting that halos are enriched to large radius as a consequence of star formation. The second indication is the mean metallicity of the X-ray–emitting gas in rich clusters. Although the metallicity of the gas is a factor of 2 or 3 lower than that of the galaxies, the total metal contents in the two components are comparable, at least within the region where they are observed. Apparently, a substantial fraction of the heavy elements produced by the stars was not retained by the galaxies; even quite bright galaxies must therefore have contaminated the surrounding medium. Since the galaxies have higher metallicity than the gas, they must have held on to some of their metals, for the situation would otherwise resemble the solar neighborhood where the mean metallicity of the gas exceeds that of the stars (e.g., Tinsley 1979). In addition, the systematic increase of galaxy metallicity with luminosity suggests that large galaxies retained more of their heavy elements, a property which is natural to link to their greater star formation efficiency as expressed in equation (23) or equation (24). Models of this type were first proposed by Larson (1974b) and were explored further by Dekel & Silk (1986) among others.

To explore the effects of enrichment on our galaxy formation picture, we will compare models in which it is neglected to models in which it occurs with maximum efficiency. The first case is equivalent to assuming that the heavy elements produced by star formation never leave the galaxy center, and so do not affect gas cooling rates. In this case, the mean metallicity of the stars is equal to the yield (defined as the mass of metals produced per unit mass of long-lived stars formed), and the mean metallicity of the *halo* gas is zero (Tinsley 1980a). In the other extreme case, we assume that the metals are uniformly mixed with the cooling gas, but not beyond it. This leads to the highest plausible metallicity in this gas, and thus to the maximum plausible effect on its cooling rate. For a given halo the enrichment equations can be set up as follows. Let $M_t(t)$ be the total mass of gas available for star formation (i.e., the minimum of $M_{\rm inf}$ and $M_{\rm cool}$ from eqs. [19] and [20]), and let $M_h(t)$ be the mass of gas heated by supernovae. Our hypotheses for star formation are that $\dot{M}_* = \epsilon(V_c)\dot{M}_t$, and that $\dot{M}_h = \dot{M}_t - \dot{M}_*$. Further, for a given halo let Z_0, Z_h, and Z_* be the metal abundances by mass in the initial gas, the heated gas, and the stars, and let y be the yield. Then one finds

$$\frac{d(M_* Z_*)}{dt} = Z_h \frac{dM_*}{dt}$$

and (25)

$$\frac{d(M_h Z_h)}{dt} = Z_0 \frac{dM_t}{dt} + (y - Z_h)\frac{dM_*}{dt} .$$

The mass of metals in stars increases as metals are incorporated from the surrounding gas, while that in gas is increased by expansion of the cooling or accretion radius and by stellar ejecta, but is decreased by the metals lost to stars. The solution of these equations which is relevant to our models is the simple one

$$Z_h(V_c, z) = \epsilon(V_c)y + Z_0(V_c, z) . \quad (26)$$

This is the metallicity we use to estimate cooling rates in equation (16); the zero-enrichment case corresponds to setting $y = 0$. In this rather crude model the metallicity of the stars formed in a given halo depends primarily on V_c and is almost independent of position and time. This is in rought accord with the weak observed correlation between age and metallicity in the solar neighborhood, but it cannot explain the observed metallicity gradients in spiral disks. The spread of metallicities in the Galactic bulge would have to be ascribed to the merging of systems which originally had very different values of V_c. However, it probably does not make sense to try to interpret the model in such detail.

It remains to specify the initial metallicity, $Z_0(V_c, z)$, in equation (26). This depends on the enrichment history of the gas in each particular halo. We address this problem by using equations (11) and (12) to specify how the stars and metals produced by star formation at one epoch are distributed among the halos present at a later time. Thus the mean mass and metallicity of stars in a halo are given by

$$M_*(V_c, z)n(V_c, z) = \int_0^{t(z)} dt' \int_0^\infty dV\, \dot{M}_*(V, z') \times n(V, z')f(V_c, z \mid V, z') , \quad (27)$$

and

$$M_*(V_c, z)Z_*(V_c, z)n(V_c, z) = \int_0^{t(z)} dt' \int_0^\infty dV\, \dot{M}_*(V, z')n(V, z') Z_h(V, z')f(V_c, z \mid V, z') . \quad (28)$$

The fraction of the baryons remaining gaseous (f_g in eq. [3]) and the metallicity Z_0 of this gas can then be obtained from

$$f_g(V_c, z) = 1 - M_*/(\Omega_b M) \quad (29)$$

and

$$Z_0(V_c, z) = (f_g^{-1} - 1)(y - Z_*) , \quad (30)$$

where M is the total halo mass from equation (3).

When metal enrichment is included, these assumptions add the yield, y, as a further parameter of our models. Rather than taking y as fixed by the initial mass function and the theory of stellar nucleosynthesis, we have elected to treat it as a free parameter and to choose its value so that the mean metallicity of all the stars formed by $z = 0$ is about 0.7 times the solar value. This is an arbitrary but plausible choice. Because of equations (27)–(30) the mean metallicities of stars and gas, Z_* and Z_0, and the fraction of the baryons in stars, $f_* = 1 - f_g$, are a function of V_c in present-day halos. Figure 4 illustrates this dependence in the same models for which star formation rates are illustrated in Figure 3. The fraction of the gas converted to stars increases rapidly with V_c and then levels off, reflecting the V_c dependence of the star formation efficiency. The metallicity distributions in Figures 4*b* and 4*c* show a roughly similar behavior, although in the case of Z_* the leveling off occurs only at rather large values of V_c unless Ω_b is small. The various wiggles in these curves reflect similar features in the star formation rates of Figure 3 and can be traced back to features in the cooling functions, $\Lambda(T)$. Notice that the dependence on fluctuation amplitude b is very weak compared with that on baryon density Ω_b or on the treatment of enrichment.

In halos like that of our own Galaxy the curves of Figure 4 predict that a fraction $\sim 0.015/\Omega_b$ of the gas has turned into stars with approximately 70% of the solar metallicity, while the gas is about 2.5 times less metal-rich. In halos which correspond to rich clusters of galaxies the star fraction is similar, the stellar metallicities are up to a factor of 2 higher, and the gas

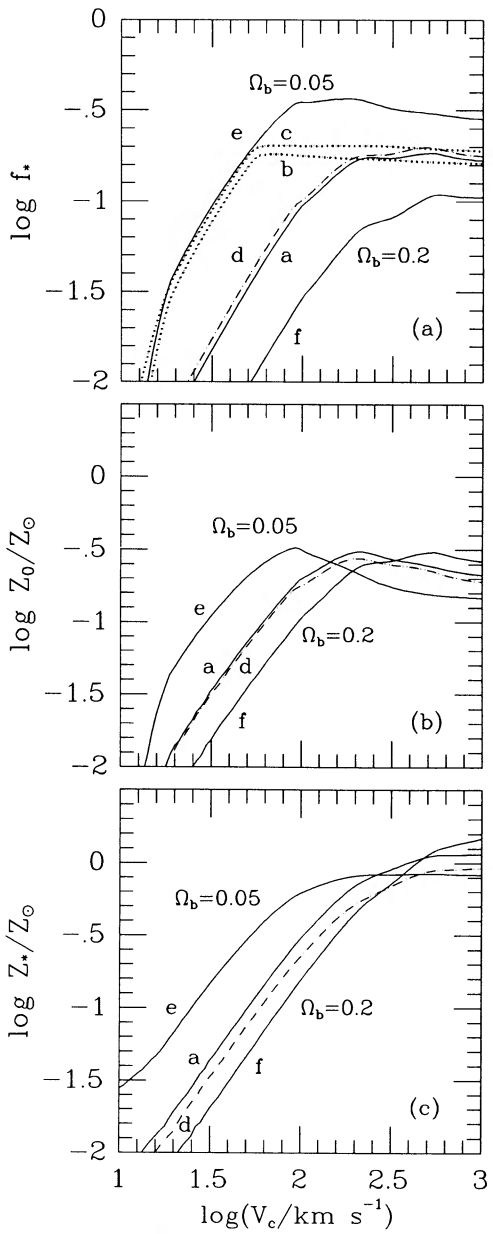

FIG. 4.—(*a*) Fraction of baryons locked up in stars in present-day dark halos as a function of circular velocity. The six curves correspond to the six models of Fig. 3 and are labeled by the panel identifier from that figure. (*b*) Mean metallicity of gas in present-day halos for the metal-enrichment models shown in (*a*). (*c*) Mean metallicity of the stars in the metal-enrichment models shown in (*a*).

metallicities are slightly lower. These metallicity trends are qualitatively similar to those observed, and it seems likely that some fiddling with our model could give quantitative agreement with the rather sparse data available. However, we argued above that data on the Coma Cluster and other rich clusters imply that roughly 30% of the baryons in such systems are in the form of stars, and that roughly 19% of the cluster material is baryonic. It is clear that for $\Omega_b = 0.19$ our models would predict a star fraction about a factor of 3 too small. This is a facet of another problem which we discuss further below. Although the formation of stars is biased toward massive systems in our models, the bias seems smaller than that required observationally if the real universe is, as our models assume, flat. This formation bias must therefore be supplemented by other effects if our models are to match observation.

5. STELLAR POPULATIONS

In order to compare our models directly with the observed properties of galaxies and galaxy clusters, we need to make some assumptions about the kind of stars that form. For this paper we have chosen to assume that stars always and everywhere form with the same initial mass function (IMF). This is a critical and poorly justified assumption with important consequences. For example, we discount the possibility that under some circumstances substantial amounts of baryons might have turned into planet-like objects. As a result we are forced to interpret the small observed fraction of baryons in visible stars ($\Omega_* \lesssim 0.01$) as the consequence of inefficient star formation. The highly effective feedback efficiencies required to produce this could be avoided if most star formation occurred at large radii and produced essentially invisible objects. Such a situation is indeed suggested by the properties of observed cooling flows (e.g., Fabian, Nulsen, & Canizares 1991), and arguments in favor of it are presented by Thomas & Fabian (1990) and Ashman & Carr (1991).

Even after electing for a universal IMF, we need to decide which one to use. For our purposes the properties needed are rather simple. We require a relation between the mass-to-light ratio of a stellar population and its age, for which we adopt the approximate form

$$(M/L)(t) = (M/L)_0(t/10\ \mathrm{Gyr})^{\gamma}\,, \tag{31}$$

and a relation between the UV luminosity of a star-forming region and the star formation rate, which we take as

$$L = L_{\mathrm{uv}}(\dot{M}_*/1\ M_\odot\ \mathrm{yr}^{-1})\,. \tag{32}$$

In this second expression L_{uv} is the power per unit frequency and is roughly constant over the wavelength range 1000–3000 Å. (See the synthetic spectra of star formation bursts in White 1989, 1990). For IMFs of standard form $(M/L)_0$ is determined mainly by the relative abundance of turnoff stars ($M \sim 0.7\ M_\odot$) and the lower mass objects which contain most of the mass, γ depends on the slope of the IMF in the 1–3 $M_\odot$ mass range for the age range (0.5–10 Gyr) which interests us, and L_{uv} reflects the abundance of high-mass stars ($M \gtrsim 10\ M_\odot$) relative to that of the component which dominates the total mass. We have used an updated version of Bruzual's (1983) spectral evolution program (G. Bruzual 1988, private communication) to estimate these parameters for various popular fits to the IMF of the local Galactic disk. The original Salpeter power law [$n(M)dM \propto M^{-1.35}\,dM$] truncated at 0.08 and 75 $M_\odot$ leads to the parameter set

$$((M/L)_0,\ \gamma,\ L_{\mathrm{uv}}) = (20.7,\ 1.09,\ 2.7 \times 10^{27})\,, \tag{33}$$

where the mass-to-light ratio parameters are for the B band, and the units of L_{uv} are ergs s^{-1} Hz^{-1}. Choosing the more detailed IMF fit of Miller & Scalo (1979) gives the parameter set ($6.6,\ 1.15,\ 8.6 \times 10^{27}$), while the revised IMF of Scalo's major (1986) review gives ($4.5, 0.90, 3.6 \times 10^{27}$).

Since Scalo's IMF is undoubtedly the best overall assessment of the available observational data, it would seem a priori to be the one which we should choose. However, in practice, these parameters cause us trouble. The normalization, $(M/L)_0$,

is so small that it leads to a mean galaxy mass-to-light ratio which seems unacceptably low. In our models, this is in part due to the relatively late epoch at which the bulk of the stars form, but the same problem has been noted previously in more traditional galaxy evolution models (e.g., Guiderdoni & Rocca-Volmerange 1987). In addition, Scalo's IMF has a number of marked characteristic mass scales. This seems rather surprising from a theoretical point of view and may reflect the heterogeneous data sources and lines of argument on which he is forced to draw. In any case, we have preferred to use the parameters given by the simple Salpeter model (eq. [33]). These probably overestimate the mass in faint stars, and so the mass-to-light ratios of old systems. This uncertainty in the low-mass end of the IMF has no significant impact on our feedback arguments. The fraction of baryons locked up in stars is small in all our models, so the amount of remaining gas depends only weakly on the IMF. Since this gas must be kept hot by supernovae and we normalize to the current observed luminosity density, the critical parameter is the number of supernovae produced per unit current luminosity. This ratio varies by more than a factor of 3 among the above IMFs, but the effects of this uncertainty are masked in our models by the freedom we allow ourselves in fixing the efficiency and yield parameters, ϵ_0 and y. Similarly, the uncertainty in $(M/L)_0$ scarcely influences our models because we normalize to the observed luminosity density, and so automatically obtain the right overall abundance of stars; the models just have to decide what kind of objects to put them in. Uncertainties in L_{uv} translate directly into uncertainties in the magnitudes we predict for high-redshift galaxies. As a result they affect the comparisons we make below with recent data on counts and spectroscopy of faint objects.

6. RESULTS

The treatment of stellar populations in the last section finally completes the specification of our models. The primary observational datum which we use as a constraint is the luminosity density of the universe. For this parameter we take the value given by Efstathiou, Ellis, & Peterson (1988a),

$$\mathscr{L}_B = 9.7 \times 10^7 \, L_\odot \, \mathrm{Mpc}^{-3} , \qquad (34)$$

which, according to these authors, has an rms uncertainty of about 35%. When we make a specific model, we choose the parameters ϵ_0 (eq. [23]) and y (eq. [25]) so that the final luminosity density matches equation (34) and the mean metallicity of the stars is 70% of solar, except that in models with no enrichment we simply set $y = 0$ and then adjust ϵ_0 alone to fit $\mathscr{L}_B$. With these constraints the only parameters that remain free are the fluctuation amplitude b and the baryon density Ω_b. In addition, we can switch our treatments of enrichment and bursting on or off. Together these two parameters and two switches define the degrees of freedom which we explore below. From a practical point of view it is unfortunate that the best estimates of the luminosity density of the universe are in the B band. In models such as ours, where present star formation rates are not small compared with the past average, a significant fraction of the present luminosity density at B is predicted to come from stars with ages of 10^9 yr or less. This is not ideal, since our goal is to normalize the *total* stellar content of the models. However, equation (34) appears to be the best direct observational constraint available, and, together with the parameters of equation (33), it leads to stellar *mass* densities in our models which are on the range normally assumed.

Before describing the detailed properties of our models, it is perhaps worth reviewing how we calculate them in practice. We pick a logarithmically spaced grid in $(V_c, 1 + z)$-space, usually with 100 points in each direction and with 10 km $s^{-1} < V_c < 3000$ km s^{-1} and $0 < z < 30$. We then integrate forward step by step from high redshift. At each time we calculate the abundance of halos of each V_c from equation (5), and their star formation rates from equation (23). We then evaluate the total mass of stars formed in such halos during the time step, and the mass of metals they contain (eqs. [25] and [26]). Through equations (27) and (28) we distribute these stars, their metals, and their luminosity both among the halos present at the next time step and among those present at $z = 0$. This provides the gas content and gas metallicity of the new halos through equations (29) and (30), so that we have all the information required for the next interval of star formation. At the end of the calculation (i.e., at $z = 0$) we compare the mean luminosity density and metallicity of the stars with the values we require as standard, and we adjust the model parameters ϵ_0 and y iteratively to obtain agreement.

With all our analytic apparatus in place, we now give a systematic exposition of the properties of our models. To explore their sensitivity to the remaining parameters, we have calculated cases with fluctuation amplitudes $b = 2.5$ and $b = 1.5$, with baryon densities $\Omega_b = 0.2$, 0.1, and 0.05, and both with and without enrichment of the halo gas. We also consider both bursting and nonbursting models, although this only affects the properties discussed in § 6.4 below. We begin by describing some of the global properties of the models, in particular their star formation rates as a function of redshift and halo size. We then analyze the star content of present-day halos, producing a luminosity function of "virialized systems" which can be compared directly with observation. This calculation avoids the need for any explicit treatment of the merging of galaxies (as opposed to halos) and so sidesteps the question of how often galaxies merge when their halos coalesce. If such merging is assumed negligible, it is possible to calculate galaxy luminosity functions, and relations between galaxy luminosity, metallicity, and circular velocity. We carry out these calculations and show that while the models reproduce the form of the observed relations between V_c, Z, and L relatively easily, it seems impossible to reproduce the observed luminosity function without substantial loss of faint galaxies through merging. Finally, we calculate luminosity functions for the young star component of our galaxies as a function of redshift, and use them to predict counts and redshift distributions of faint blue galaxies.

Obviously many aspects of our models could be varied within the general framework we have set up. It is, in fact, the ease with which the consequences of various hypotheses can be investigated which is the main strength of our approach. Although we concentrate here on a direct elaboration of the ideas of White & Rees (1978) within a flat CDM universe, most other hierarchical clustering models can be studied with minor variations of our techniques.

6.1. *Global Properties*

In Table 1 we give a variety of properties of our standard set of models. For each Ω_b and b we list the values of ϵ_0 and y which reproduce our standard luminosity density and stellar metallicity. The required feedback efficiency increases strongly with mean baryon density, is substantially increased by the inclusion of the cooling effects of heavy elements, and in

TABLE 1

CHARACTERISTIC PROPERTIES OF GALAXY FORMATION MODELS

Ω_b	b	y	ϵ_0	$(M/L)_*$	Z_*	Z_g	L_h	$V_{c,h}$	$(M/L)_h$	L_g	$V_{c,g}$	z_1	z_{10}	z_{50}
0.20.....	2.5	0	0.22	7.0	0	0	1.1×10^{10}	217	0.60	1.4×10^{9}	173	6.3	3.2	1.0
0.10.....	2.5	0	0.04	8.6	0	0	3.0×10^{9}	146	0.67	2.6×10^{8}	97	7.4	3.9	1.3
0.05.....	2.5	0	0.002	11.3	0	0	6.0×10^{8}	83	0.62	4.1×10^{7}	62	9.0	5.0	1.8
0.20.....	1.5	0	0.45	9.0	0	0	8.4×10^{10}	431	0.62	6.9×10^{9}	329	10.5	5.2	1.5
0.10.....	1.5	0	0.06	11.5	0	0	2.0×10^{10}	289	0.78	4.1×10^{8}	155	12.6	6.6	2.2
0.05.....	1.5	0	0.001	15.6	0	0	3.4×10^{9}	173	0.99	1.8×10^{7}	64	17.5	10.1	3.7
0.20.....	2.5	3.7	0.81	5.5	0.68	0.12	2.9×10^{10}	273	0.45	8.1×10^{9}	245	5.0	2.2	0.7
0.10.....	2.5	2.2	0.22	6.3	0.73	0.15	1.5×10^{10}	218	0.45	3.4×10^{9}	202	5.4	2.6	0.8
0.05.....	2.5	1.2	0.03	7.8	0.70	0.14	4.5×10^{9}	155	0.53	6.7×10^{8}	116	6.6	3.4	1.1
0.20.....	1.5	3.1	1.37	7.8	0.68	0.14	1.5×10^{11}	484	0.49	1.8×10^{10}	410	8.3	3.9	1.1
0.10.....	1.5	1.8	0.33	9.1	0.70	0.16	4.9×10^{10}	364	0.64	4.8×10^{9}	267	9.3	4.6	1.4
0.05.....	1.5	1.05	0.03	11.3	0.71	0.16	1.3×10^{10}	230	0.61	5.5×10^{6}	145	12.1	6.1	2.1

general is higher for the larger fluctuation amplitude (smaller value of b). For the more extreme parameter combinations, feedback must be implausibly efficient if overproduction of stars is to be avoided. On the other hand, with $\Omega_b = 0.05$ and no enrichment it is quite difficult for enough material to cool to make the observed stars, and feedback must be almost negligible. When enrichment is included, the yield which reproduces our fiducial stellar metallicity (70% of solar) is a strong function of the mean baryon density, reflecting, as discussed below, the Ω_b dependence of the epoch when stars form. However, the mean metallicity of the residual gas (Z_g in Table 1) is almost independent of Ω_b and y. The mean mass-to-light ratio of the stars at the end of the calculation decreases with increasing Ω_b, with increasing b, and with the inclusion of cooling by heavy elements. This is simply a reflection of the typical stellar age (see eq. [31]); there is an almost perfect correlation between $(M/L)_*$ and z_{50}, the redshift by which 50% of the final stars have formed.

This median redshift is listed in Table 1 along with the redshifts z_1 and z_{10} corresponding to the epochs when 1% and 10% of the stars have formed. It is clear that stars are formed later in models with large Ω_b, with large b, and in which metal cooling is important. In most of the models, however, $z_{50} \lesssim 1$, and it is worth asking whether this can be consistent with observation. One can argue that roughly half the observed stars in the universe are in galaxy disks (Schechter & Dressler 1987) and that in our own Galactic disk all the stars seem to be quite young, at least compared with globular star clusters (Demarque & McClure 1977; Freeman 1989). Thus, the proposition that most stars formed recently is not a priori absurd. A counterargument may be that bright systems are seen at high redshift which appear to consist almost entirely of old stars. Unfortunately, it is not yet clear how representative these systems are and exactly what limits can be put on their ages (e.g., Chambers & Charlot 1990). Perhaps new extremely deep counts in the near-infrared will begin to show us what fraction of galaxy light had already come from "old" systems by $z = 1$ (Cowie et al. 1990).

The history of star formation is illustrated in more detail in Figure 5, where we show the mean star formation rate as a function of redshift. The peak rate occurs at redshifts between about 2 and 10 and has a value between about 0.1 and 1 $M_\odot$ yr^{-1} Mpc^{-3}. Both the peak rate and the peak redshift depend inversely on Ω_b and y and are decreased by the inclusion of metals. Using the standard IMF which we have adopted, these rates can be converted directly into production rates for ionizing radiation. If these hard photons escape their protogalaxies, they may ionize the diffuse intergalactic medium. It is interesting that our peak star formation rates approach those needed for a medium containing 10% of the critical density to pass present limits from the Gunn-Peterson test, even those derived from high-redshift quasars (Shapiro & Giroux 1989). In our model such photoionization is probably needed because much of the gas is in halos which have virial temperatures too low to ensure sufficient collisional ionization. In fact, we predict that for $b = 2.5$, about two-thirds of the gas has yet to be incorporated in a halo hotter than 10^4 K by a redshift of 3; for $b = 1.5$ the corresponding fraction is 40%. Clearly, ionization by a UV background must be invoked to keep this gas invisible. The nonuniform distribution of gas and its clumping around the sites of UV emission make a proper analysis of this issue quite tricky.

In addition to exploring the distribution of star formation in time, we can study its distribution with respect to halo circular velocity. The top two panels of Figure 6 show the mass of stars formed per unit volume per unit ln V_c. These distributions vary substantially with our model parameters. Stars tend to form in halos with smaller circular velocity when Ω_b is small, when b is large, and when heavy-element cooling is negligible. Thus, for $\Omega_b = 0.05$, $b = 2.5$, and no enrichment, this function peaks at about 50 km s^{-1}, and most stars form in halos with circular velocities between 20 and 100 km s^{-1}. Including enrichment or increasing Ω_b to 0.2 shifts the peak to 100 km s^{-1}, and, in the latter case, many stars form in halos with circular speeds as large as 300 km s^{-1}. For $b = 2.5$ and $\Omega_b = 0.1$ or 0.2, most stars form in halos with $V_c \sim 200$ km s^{-1} when enrichment is included. However, when the fluctuation amplitude is increased to $b = 1.5$, significant numbers of stars form in halos with $V_c > 500$ km s^{-1}, and indeed the peak of the distribution occurs near this velocity for the extreme model with $\Omega_b = 0.2$ and heavy-element cooling. This seems implausible. Bright elliptical galaxies can have measured (one-dimensional) velocity dispersions of order 300 km s^{-1}, and so inferred circular velocities of order 500 km s^{-1}; indeed, the measured circular velocity of M87 is about 660 km s^{-1} at 200 kpc (Fabricant & Gorenstein 1983). However, such galaxies are quite rare, and it seems unlikely that their stars formed in recent cooling flows, as predicted by our model. This difficulty appears related to the well-known problem associated with cluster cooling flows, which we have already mentioned. Gas appears to be condens-

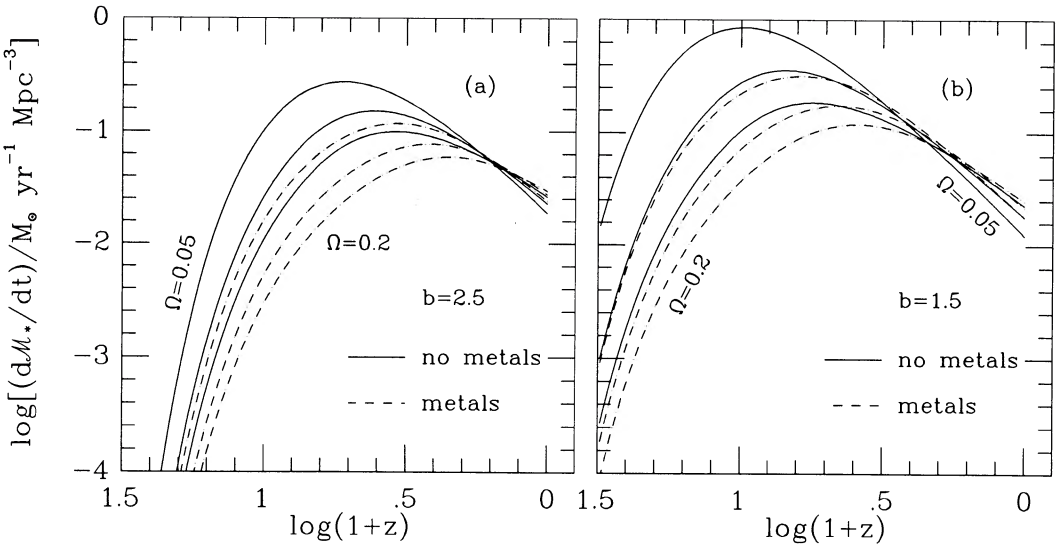

FIG. 5.—Mean star formation rate per unit volume as a function of redshift. Dashed lines correspond to models with metal enrichment, and solid lines to models without enrichment. In each case, the three curves correspond to different values of the mean baryon density: $\Omega_b = 0.05$, 0.1, and 0.2. Values of the other parameters in these models are given in Table 1. (*a*) Biasing parameter $b = 2.5$. (*b*) Biasing parameter $b = 1.5$.

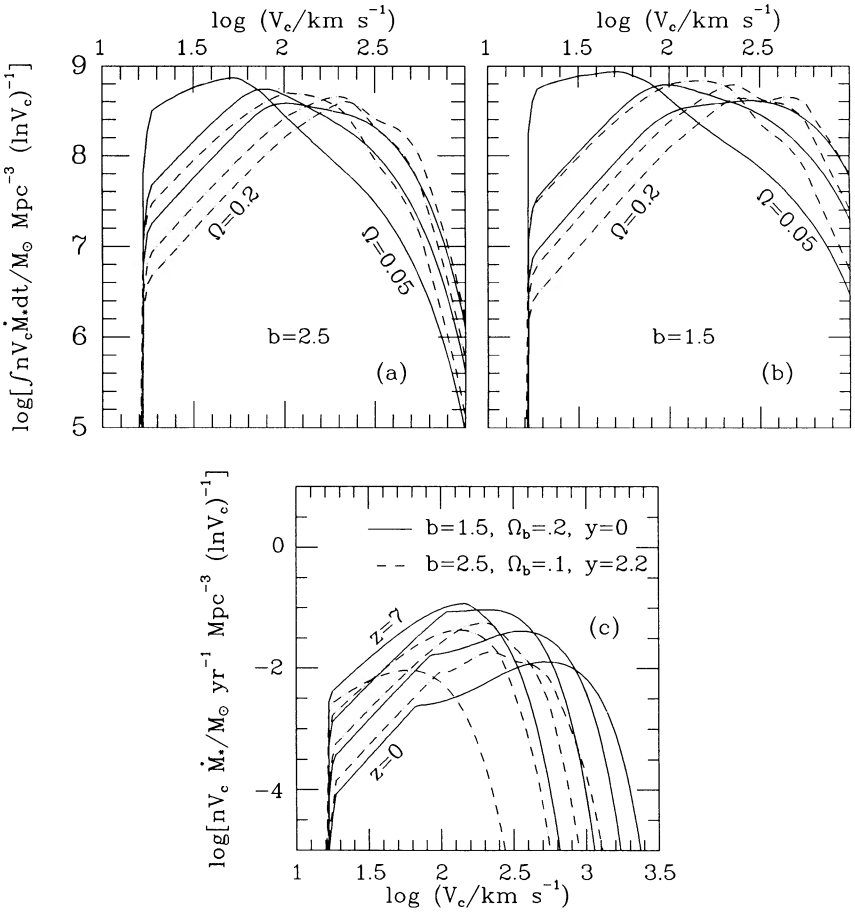

FIG. 6.—(*a*) Mass of stars formed per unit volume as a function of the circular velocity of the object in which they were formed. Dashed lines correspond to models with metal enrichment, and solid lines to models without enrichment. In each case, the three curves correspond to different values of the mean baryon density: $\Omega_b = 0.05$, 0.1, and 0.2. The biasing parameter is $b = 2.5$; values of other parameters are given in Table 1. (*b*) As in (*a*), for $b = 1.5$. (*c*) Mean star formation *rate* per unit volume as a function of halo circular velocity and redshift. Dashed lines correspond to the enrichment model with $\Omega_b = 0.1$, $b = 2.5$, and solid lines to the no-enrichment model with $\Omega_b = 0.2$, $b = 1.5$. In each case, different curves correspond to different redshifts: $z = 7$, 3, 1, and 0.

ing into stars at a substantial rate, and yet there is no sign of the massive stars associated with "normal" star formation (e.g., Fabian et al. 1984).

The last panel of Figure 6 illustrates how star formation shifts progressively to deeper and deeper halos. We pick two models and plot the mean star formation *rate* per unit volume per unit $\ln V_c$ [in units of $M_\odot$ yr^{-1} Mpc^{-3} $(\ln V_c)^{-1}$] at redshifts of 7, 3, 1, and 0. As the redshift decreases, the peak of star formation moves to larger V_c. In the model with $\Omega_b = 0.2$, $b = 1.5$, and no enrichment, the total star formation rate also declines substantially, since $z = 1$ (cf. Fig. 5). Most of the present-day star formation in this model is inferred to be occurring in halos with 200 km s^{-1} $< V_c <$ 1000 km s^{-1}. The upper limit of this range seems very large, and is another manifestation of the "cooling-flow problem" referred to in the last paragraph. Even at $z = 1$ star formation is predominantly in halos with $V_c \sim 400$ km s^{-1}. The situation is less extreme and more plausible for $\Omega_b = 0.1$, $b = 2.5$, and heavy-element cooling. In this case there is rather little star formation for $V_c > 500$ km s^{-1}, and most stars form in halos with $V_c \sim 200$ km s^{-1} both at $z = 1$ and at $z = 0$. Note that in almost all our models the total amount of star formation drops rapidly as the circular velocity decreases below 100 km s^{-1}; the precipitate drop at 16 km s^{-1} is due, of course, to the fact that our cooling functions assume no significant cooling below a temperature of 10^4 K.

6.2. *The Luminous Content of Halos*

In § 4 we discussed how the content of present-day halos can be calculated, and we gave formulae relating halo circular velocity to gas and star content and to metallicity (eqs. [27]–[30]). Once a particular IMF has been chosen, similar methods can be used to obtain the mean present luminosity, L, of halos as a function of V_c:

$$L(V_c)n(V_c, 0) = \int_0^{t_0} dt \int_0^{\infty} dV\, \dot{M}_*(V, z)n(V, z)\, \frac{f(V_c, 0 \mid V, z)}{(M/L)(t_0 - t)}. \quad (35)$$

In this expression t_0 is the present age of the universe, and the stellar mass-to-light ratio (evaluated at $t_0 - t$) is given by equation (31). The luminosity so obtained can be combined with the mass of equation (18) to produce a mass-to-light ratio for halos. We plot this quantity in Figure 7 for our 12 standard models and compare it with the mean M/L of 714 which our adopted luminosity density and Hubble constant imply for the universe as a whole. At large V_c, mass-to-light ratios are approximately independent of circular velocity, implying $L \propto V_c^3$, while for small V_c they vary approximately as V_c^{-2}, implying $L \propto V_c^5$.

We predict that small halos contain rather few stars because of the inefficiency of star formation in such systems and in their progenitors. To compensate, the mass-to-light ratios of larger halos are predicted to be less than that of the universe as a whole. This latter bias is exactly the kind of effect that is needed to reconcile observations of the dynamical properties of groups and clusters of galaxies with a flat universe. However, the amplitude we predict is insufficient. The bias is stronger for models with heavy-element cooling, with large values of Ω_b, and with small fluctuation amplitudes. The first two trends are easily understood as reflecting similar trends in the strength of feedback. The greater the feedback efficiency, the greater the suppression of stars in small systems and so the stronger the bias. The trend with fluctuation amplitude arises because the overall mass distribution shifts toward more massive halos as b decreases; the suppression of star formation in small objects then affects less of the mass and requires less bias to compensate. Dynamical studies of groups and clusters of galaxies suggest mass-to-light ratios which are smaller than that required to close the universe by factors of 3–20, whereas the maximal bias given by the present models is about a factor of 2. This deficiency may be at least partially offset by dynamical segregation effects which lead to a concentration of galaxies toward the centers of their groups, much as galaxy light is concentrated at the center of individual halos (Barnes 1985; Evrard 1987; West & Richstone 1988). Such effects may also produce a systematic difference between the velocities of galaxies and of dark matter particles in clusters (Carlberg 1991).

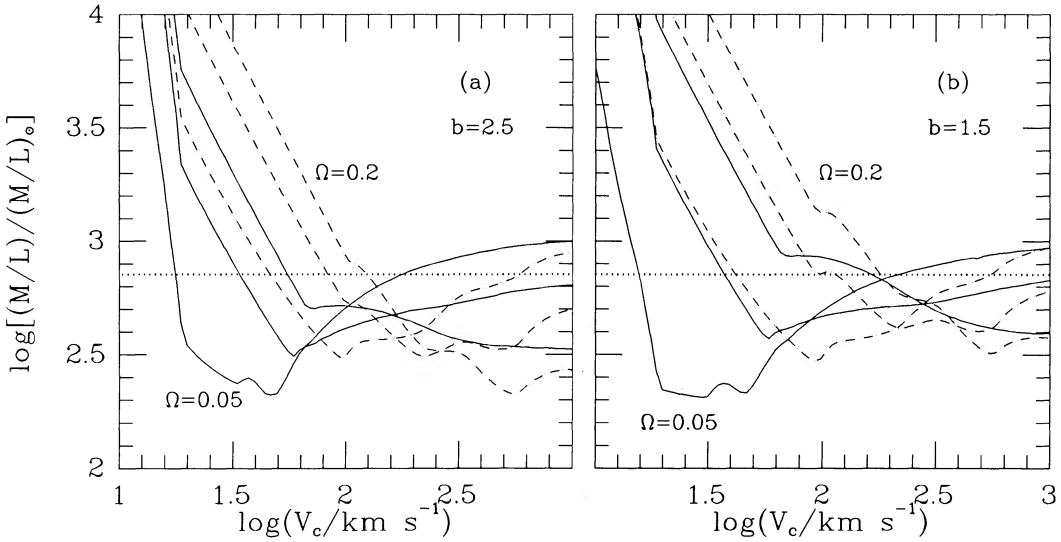

FIG. 7.—Mass-to-light ratio as a function of circular velocity. The mass is defined within the region around the center of each halo for which the mean overdensity is 200. Dashed lines correspond to models with metal enrichment, and solid lines to models without enrichment. The different curves correspond to the standard models described in Table 1. The dotted line gives the mass-to-light ratio of the universe as a whole. (*a*) Biasing parameter $b = 2.5$. (*b*) Biasing parameter $b = 1.5$.

Our model seems to require this kind of help to be compatible with observation.

The relation between circular velocity and luminosity given by equation (35) can be combined with our expression for halo abundance (eq. [5]) to yield a present-day luminosity function for halos:

$$\Phi_{\rm AGS}(L)dL = n(V_c, 0)\frac{dV_c}{dL}dL\,. \tag{36}$$

For large L this gives the abundance of galaxy groups and clusters as a function of their luminosity, whereas for small L the luminosity presumably applies mainly to individual isolated objects. This luminosity function of "all galaxy systems" (AGS) was first measured by Bahcall (1979), and has recently been remeasured for the CfA catalog by Moore, Frenk, & White (1991). It is very useful to us in the present context because we can calculate it without addressing the problem of galaxy merging; when specifying the luminosity of halos through equation (35), the number of separate galaxies among which the stars are divided is immaterial. Parameterizing the luminosity function by the usual Schechter form,

$$\Phi(L) \propto L^{-\alpha}\exp\left(-L/L_*\right), \tag{37}$$

Moore and coworkers found that whereas the luminosity function for *galaxies* in the CfA catalog is best fitted by the parameters $\alpha_{\rm gal} = -1.1$, $L_{*,\rm gal} = 4\times10^{10}\ L_\odot\ {\rm Mpc}^{-3}$, the AGS function is better fitted by $\alpha_{\rm AGS} = -1.35$, $L_{*,\rm AGS} = 10L_{*,\rm gal}$ for $L > 0.2L_{*,\rm gal}$. For fainter objects the AGS function must clearly increase somewhat less rapidly in order to remain compatible with the galaxy luminosity function. At the bright end it cuts off less sharply than the exponential of a Schechter function.

Figure 8 shows the AGS luminosity functions of our standard models and compares them with a smooth fit to the CfA data. The shapes are quite similar for all models. Bumps occur at points where our (L, V_c) relations have sudden changes of slope (cf. Fig. 7) and can be traced back to bumps in the cooling function. The exponential cutoff at the bright end reflects that in equation (5), so the relevant characteristic luminosity is proportional to the cube of the characteristic circular velocity divided by the appropriate M/L ratio. Thus for each value of b the luminosity cutoff brightens as the high-mass M/L values in Figure 7 decrease. Brighter cutoffs occur for larger values of Ω_b and for models including enrichment. Models with $b = 1.5$ have substantially brighter cutoffs than those with $b = 2.5$ because of their larger characteristic value of V_c. Because of the unrealistically large M/L ratios of large halos in Figure 7, most models fail to produce enough objects with the luminosity of rich clusters, even though they all have enough lumps with the appropriate mass and velocity dispersion (Frenk et al. 1990). In this instance the discrepancy cannot be removed by invoking dynamical segregation effects, since these do not alter the total luminosity of a lump. The faint-end slope of all the AGS functions flattens gradually as the bright-end cutoff increases, but in almost all cases the approximate value of "α" lies between -1.5 and -1.7. It is easy to show that a slope slightly flatter than -1.6 is expected for halos where star formation is limited by infall rather than cooling and where feedback is important. For some models, particularly for low b, the agreement with the observed AGS luminosity function appears acceptable. Notice, however, that the observational estimates are not well defined at the small-L end because of the difficulty of assigning galaxies to small groups, and are uncertain at the high-L end because of poor statistics. At small L the true function should lie below the galaxy luminosity function, which is almost certainly flatter than $\alpha = -1.35$ (but see below).

A number of the trends indicated in Figures 7 and 8 are shown more quantitatively in Table 1. The column labeled $V_{c,h}$ gives the circular velocity such that half the luminosity of the universe is in halos with $V_c > V_{c,h}$; the mean luminosity, L_h, and mass-to-light ratio, $(M/L)_h$, of these median halos are also listed, the first in solar luminosities and the second in units of

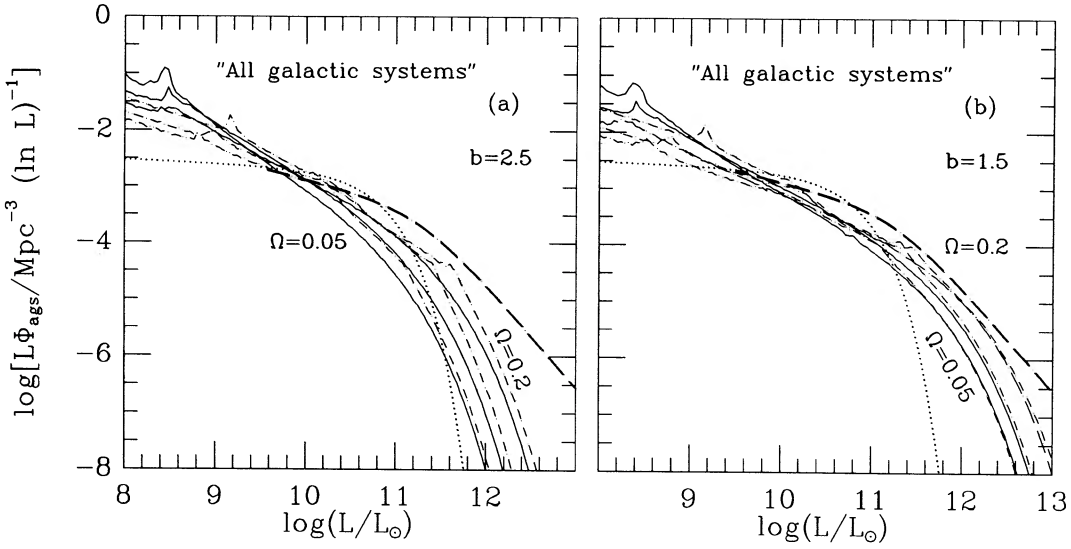

FIG. 8.—Luminosity function of "all galaxy systems." Luminosities are in the B band. Solid and dashed curves correspond to the standard no-enrichment and enrichment models, respectively, described in Table 1. The dotted line represents the Schechter function fit to the *galaxy* luminosity function in the CfA survey. The heavy dashed line is a smooth fit to the AGS luminosity function estimated by Moore et al. (1991), also for the CfA survey. (*a*) Biasing parameter $b = 2.5$. (*b*) Biasing parameter $b = 1.5$.

the overall mass-to-light ratio of the universe. It is interesting to compare the values of $V_{c,h}$ with circular velocities defined in a similar way for the total mass, rather than the luminosity. For $b = 2.5$ half the mass of the universe is in halos with $V_c > 104$ km s^{-1}, while for $b = 1.5$ it is in halos with $V_c > 244$ km s^{-1}. In the models with efficient feedback the luminosity is substantially biased toward the most massive halos, as shown both by $V_{c,h}$ and $(M/L)_h$. For the AGS luminosity function of Moore et al. (1991) the median luminosity is $6.4 \times 10^{10}\ L_\odot$. Without enrichment the models require both a small value of b and a large value of Ω_b to achieve such a large characteristic luminosity. When metallicity effects are included, a $b = 2.5$ model still requires a baryon density somewhat greater than 0.2 to produce bright enough lumps, but the $b = 1.5$ models reach the requisite luminosity for Ω_b slightly more than 0.1. Thus it seems that our models have some difficulty in accumulating the stars into sufficiently bright objects, particularly for the low values of Ω_b preferred by recent reanalyses of the constraints posed by the observed light-element abundances (Olive et al. 1990).

6.3. "*No-Merger*" *Luminosity Functions*

If we wish to calculate a galaxy luminosity function which can be compared directly with observation, we must address the difficult question of galaxy merging. Our models fix the total present-day luminosity in stars, and they specify the distribution of star formation with respect to time and halo circular velocity. However, none of our assumptions so far allow us to specify the total luminosity of the *galaxy* (rather than the halo) in which particular stars are to be found today. Perhaps the simplest possibility is to assume that the stars formed at redshift z in a halo with circular velocity V_c are to be found today in a galaxy with luminosity

$$L_{\rm nm}(V_c, z) = \dot{M}_*(V_c, z)t(z)/(M/L)(t_0 - \bar{t})\ . \qquad (38)$$

(The subscript "nm" here stands for "no merger.") Star formation in each halo is thus taken to continue at a constant rate for a time equal to $t(z)$, the age of the universe at redshift z. The resulting galaxy is assumed to survive until the present day and to have a present mass-to-light ratio corresponding to that of a set of stars formed at time $\bar{t}$. For $t < 2t_0/3$, it seems natural to take $\bar{t} = t$. However, for more recent epochs a time interval of length t, centered at time t, would extend beyond the present. Thus we adopt

$$\bar{t} = t\ , \quad t < 2t_0/3\ ; \qquad \bar{t} = 2t_0/3\ , \quad t > 2t_0/3\ . \qquad (39)$$

We can now apportion the luminosity density of our models into individual galaxies. We obtain a luminosity function by calculating the contribution to the luminosity density from galaxies with luminosity in the interval $(L, L + dL)$ and dividing the result by L. Hence,

$$\Phi_{\rm nm}(L)dL = L^{-1} \int\int_{\mathscr{R}} dt\, dV\, \dot{M}_*(V, z)n(V, z)/(M/L)(t_0 - t)\ , \qquad (40)$$

where the integration region, $\mathscr{R}$, is defined as the set of values for which $L_{\rm nm}(V, z)$ is contained in the interval $(L, L + dL)$. These procedures assume that a typical halo's life span from formation until incorporation into a larger object is, with relatively little scatter, equal to the age of the universe when it is identified; that this characteristic life span is independent of circular velocity; and that the galaxy formed in the halo survives as a separate entity within the larger systems with which the halo merges. Examination of appropriate survival probabilities generated from equation (12) supports the assumptions about mean halo lifetime and its lack of correlation with V_c. This approximation is very similar to that employed by WR to calculate galaxy luminosity functions, and has the virtue of giving a lower limit to the effects of merging. An upper limit is presumably provided by the AGS luminosity functions of the last section.

We give the "no-merger" luminosity functions for our standard models in Figure 9 and compare them with the Schechter function fit to the observed galaxy luminosity function which we described in § 6.2. As expected, these galaxy luminosity functions always have a fainter cutoff and a steeper faint-end slope than the AGS luminosity functions of the last section.

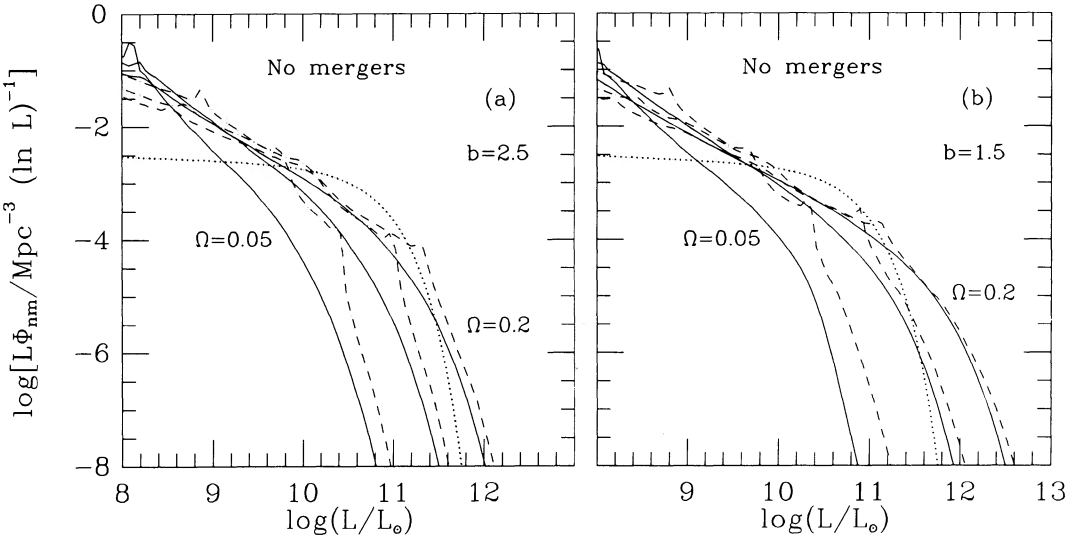

FIG. 9.—Galaxy luminosity function under the "no-merger" hypothesis. Luminosities are in the B band. Solid and dashed curves correspond to the standard no-enrichment and enrichment models, respectively, described in Table 1. The dotted line represents a Schechter function fit to the *galaxy* luminosity function in the CfA survey with parameters given in § 6.2 (as in Fig. 8). (*a*) Biasing parameter $b = 2.5$. (*b*) Biasing parameter $b = 1.5$.

The cutoffs are rather less abrupt than that of the Schechter function, although the shape at bright luminosities is reasonably well fitted by some of the models with metal enrichment. For $b = 1.5$ the cutoff is clearly too bright for the models with $\Omega_b = 0.2$, while it is much too faint for all models with $\Omega_b = 0.05$. At faint magnitudes the slopes are always much steeper than that of the Schechter function, a problem already recognized by WR. It seems plausible that this deficiency might be partially remedied by the loss of faint objects due to merging. However, it seems unlikely that flatter slopes than those of Figure 8 can be achieved by this process. As discussed in § 1, this difficulty is a consequence of the large number of faint galaxies which form at early times, rather than of the steep halo mass function predicted by Press-Schechter theory for a CDM universe. The faint-end slope of the luminosity function could also appear flatter if the extreme effects of feedback which our models postulate cause many small galaxies to have unusually low surface brightness. They might then be missed in the surveys from which field liminosity functions are constructed. The slope of our luminosity functions at faint luminosities could also be made shallower by suppressing star formation in small systems by more than the V_c^2 factor of equation (23). However, the observed value, $\alpha \sim 1.1$, does not seem attainable for any plausible parameters in the context of models where the suppression is a function of V_c. Another possibility is that this slope is affected by bursting behavior (see below; see also Lacey & Silk 1991). Note that the observed galaxy luminosity function is actually quite uncertain at these faint magnitudes. Surveys of dwarf galaxies in the Virgo (Sandage, Binggeli, & Tammann 1985; Impey, Bothun, & Malin 1988) and Fornax (Phillipps et al. 1987; Ferguson & Sandage 1988) clusters produce luminosity functions which rise steeply toward faint magnitudes; Shanks et al. (1991) also find a very steep faint-end slope for the luminosity function of blue galaxies in the field.

As stars form, our models keep track of their metallicity, their location, and their time of formation. We can thus calculate properties of our "no-merger" galaxies using weighted versions of equation (40), e.g.,

$$\langle X \rangle_{\rm nm} \Phi_{\rm nm}(L)\, dL = L^{-1} \int\int_{\mathcal{R}} dt\, dV\, \dot{M}_*(V, z) n(V, z) X(V, z)/(M/L)(t_0 - t)\,, \quad (41)$$

where the property X might be galaxy circular velocity (assumed to equal that of the halo), metallicity, mass-to-light ratio, or the square of any of these quantities (required in order to estimate the scatter around the mean value at each L). Results for our standard models are shown in Figures 10–12.

In models where feedback is important the slope of the relation between luminosity and mean circular velocity is close to $V_c \propto L^{0.25}$ for these "no-merger" galaxies. This is a consequence of our cooling and feedback models (eqs. [20] and [23]), as was noted previously by Cole & Kaiser (1989). We find that the predicted scatter around this "Tully-Fisher" relation is quite small, typically well under 20% rms in V_c at $L = 10^{10}\, L_\odot$ and well under 30% at $L = 10^9\, L_\odot$. At brighter luminosites where feedback becomes negligible the relation steepens toward $V_c \propto L^{0.5}$. In all cases our models predict too large a circular velocity for a given luminosity when compared with the observed Tully-Fisher relation (indicated by a dotted line in Fig. 10). The discrepancy is typically a factor of 2 for $b = 2.5$ and becomes more severe for $b = 1.5$. This is a serious difficulty and is related to the luminosity function problems discussed above; too large a fraction of the stars, is predicted to be in faint galaxies with small circular velocities.

A metallicity-luminosity relation can be calculated for models that include enrichment. We find $Z_* \propto L^{0.3}$ when $\Omega_b = 0.2$, and a weaker dependence for smaller baryon densities (see Fig. 11). This is similar to, but somewhat stronger than, the observed relation (Faber 1973; Visvanathan & Sandage 1977; Persson, Frogel, & Aaronson 1979; see also Tinsley 1978). There are two sources for this relation in our models. The first is the general increase in metallicity with time.

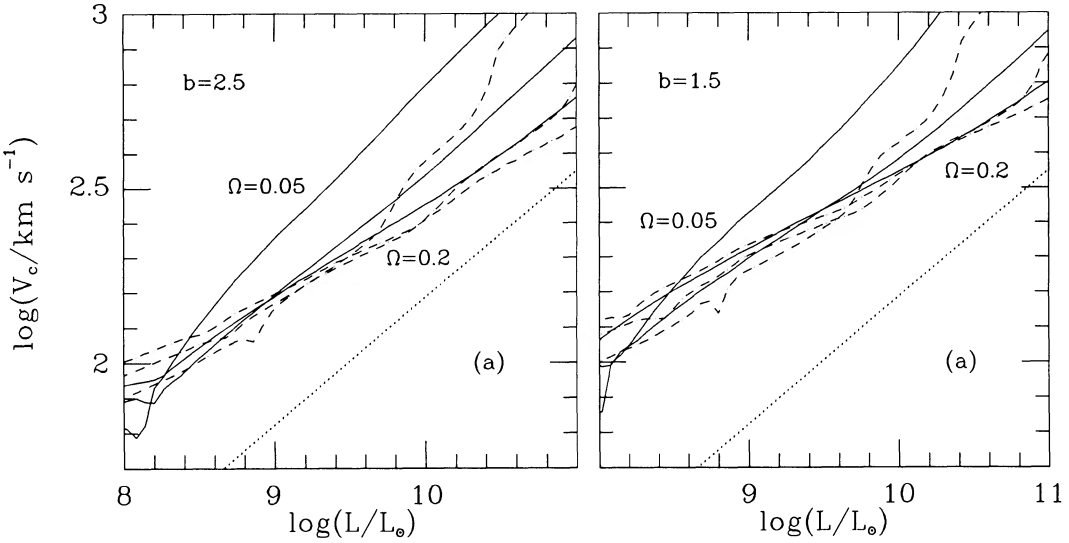

FIG. 10.—Circular velocity of galaxies in the "no-merger" hypothesis vs. their B-band luminosity. Solid and dashed curves correspond to the standard no-enrichment and enrichment models, respectively, described in Table 1. The dotted line shows the observed "Tully-Fisher" relation measured by Pierce & Tully (1988) after correction to $H_0 = 50$ km s^{-1} Mpc^{-1} and to random viewing angle. (*a*) Biasing parameter $b = 2.5$. (*b*) Biasing parameter $b = 1.5$.

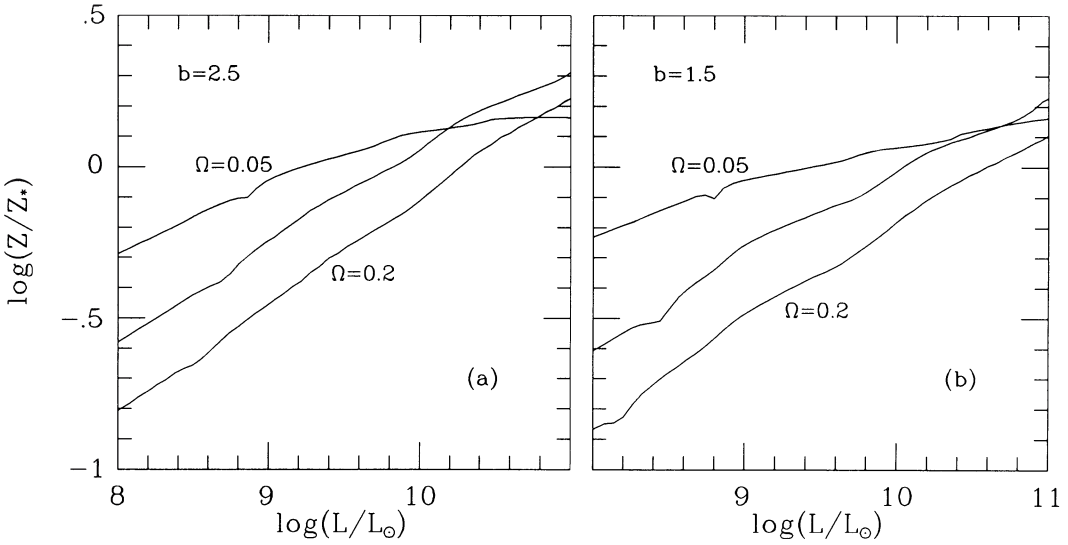

FIG. 11.—Metallicity-luminosity relation for "no-merger" galaxies. The curves correspond to the standard models with metal enrichment described in Table 1. (*a*) Biasing parameter $b = 2.5$. (*b*) Biasing parameter $b = 1.5$.

This results in a higher metallicity in the brighter objects which form later and is a rather weak effect. The more important effect comes from the dependence of $Z_h(V_c, z)$ on $\epsilon(V_c)$ in equation (26). In small objects a relatively small fraction of the gas reservoir is turned into stars, and so a lower metallicity results. The strong Ω_b dependence in Figure 11 is caused by the dependence of feedback efficiency on baryon density. Thus our explanation of the observed metallicity-luminosity relation is very similar to that of Larson (1974b), Dekel & Silk (1986), and Arimoto & Yoshii (1987). The scatter we predict about this relation is relatively small, varying from about 30% at low luminosities to less than 10% at high luminosities.

Our "no-merger" models also predict that the mass-to-light ratios of stellar populations should decrease significantly with galaxy luminosity (Fig. 12), this time with substantial scatter. This trend is a reflection of the greater mean age of stars in small galaxies (cf. eq. [31]). Many such objects are predicted to have been formed at early times when the characteristic mass of clustering was small. It is important to note that equation (41) gives luminosity-weighted averages. These are strongly biased toward recent epochs because the dependence of M/L on age is so strong in the B band. A significant fraction of the luminosity density of the universe is contributed by stars which formed in the last couple of billion years. This strong weighting toward the present has a significant effect on the shape and scatter of the relations in Figures 10–12.

The failure of our "no-merger" hypothesis to give acceptable galaxy luminosity functions is shown quite clearly by the data in Table 1. Here we denote by L_g the luminosity such that half the light of the universe comes from "no-merger" galaxies with $L > L_g$, and by $V_{c,g}$ the mean circular velocity of this median galaxy. For our Schechter function fit to the observed

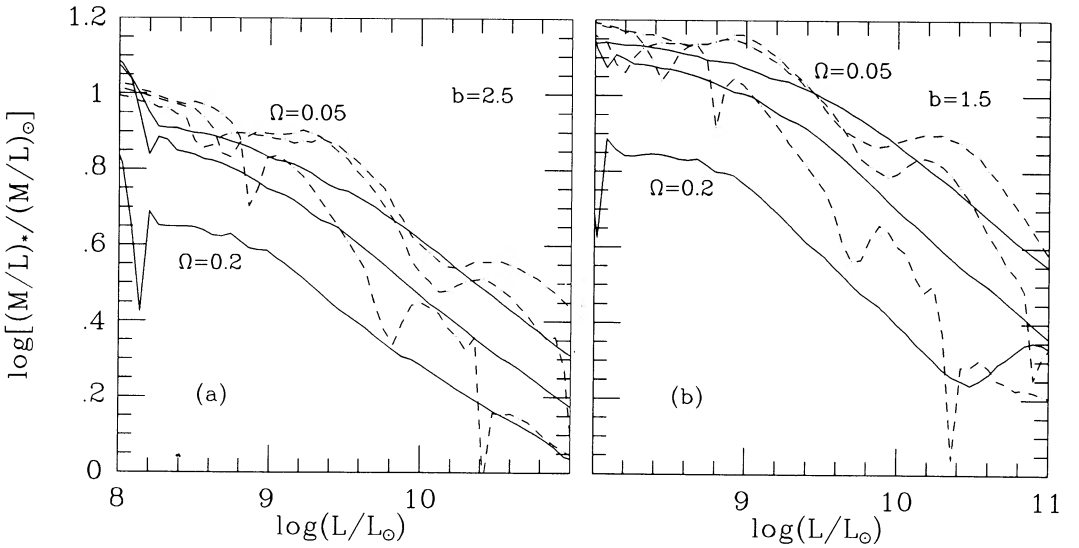

FIG. 12.—Mass-to-light ratio of the stellar populations in "no-merger" galaxies as a function of B-band luminosity. The different curves correspond to the standard no-enrichment and enrichment models, respectively, described in Table 1. (*a*) Biasing parameter $b = 2.5$. (*b*) Biasing parameter $b = 1.5$.

luminosity function the median luminosity is about 2.5×10^{10} $L_\odot$, corresponding to a circular velocity from the standard Tully-Fisher relation of about 215 km s^{-1}. None of our models gives a value of L_g this big, and many of them fail by more than an order of magnitude. For all but the most extreme model with $b = 2.5$ even the median halo luminosity L_h is too small. Thus it seems that it will be hard to fix up the luminosity function of these models by an appeal to merging. In fact, of all our models, only those with enrichment and with $b = 2.5$ and $\Omega_b = 0.2$, or $b = 1.5$ and $\Omega_b = 0.1$, seem moderately close to producing acceptable luminosity functions; the first of these also gives quite a good "Tully-Fisher" relation. However, the faint-end slope of the luminosity function and hence the median luminosity could be significantly altered by the effects discussed earlier which are not included in our models.

6.4. *Faint Galaxy Counts*

One of the most striking observational results of recent years has been the detection of faint galaxy images to levels at which they almost cover the sky. The deepest counts have been carried out in the B band and have reached densities in excess of 4×10^5 galaxies per square degree at about 27th magnitude (Tyson 1988). At faint levels, an increasingly large fraction of objects appears blue (Peterson et al. 1979; Kron 1980; Shanks et al. 1984; Koo 1986; Tyson 1988; Majewski 1989). This suggests that they are active star-forming systems at redshifts of about 1–3. (At $z > 3$ the Lyman break shifts through the U band and greatly reduces the flux there.) A moderately high characteristic redshift seems indicated by the sheer number of objects. Indeed, the volume within $z = 4$ is too small to accommodate the number of objects counted to $m_B = 27$, unless the universe is open, the number density of galaxies has decreased substantially since $z \sim 2$, or the local luminosity function seriously underestimates the mean density of galaxies (Koo 1989; Guiderdoni & Rocca-Volmerange 1989). However, this simple expectation is not borne out by spectroscopy of blue images at $m_B \lesssim 23.5$. Although these systems do seem to be actively forming stars, they have so far mostly turned out to be at relatively low redshift, $z \lesssim 0.9$ (Broadhurst, Ellis, & Shanks 1988; Colless et al. 1989; Colless et al. 1990; Cowie et al. 1990). It seems that we may be seeing a population of relatively nearby bursting dwarfs that are absent from the local luminosity function. Further observations are needed to clarify this very interesting paradox.

For objects beyond $z \sim 0.3$ the observed flux at B comes from wavelengths shortward of the Balmer discontinuity and is dominated by relatively young stars. For $1 \lesssim z < 4$ this flux comes from stars so massive that the luminosity is determined by the instantaneous star formation rate. We can thus estimate the faint galaxy counts at B relatively easily. The crude spectral model of equation (32) leads directly to an expression for the apparent B luminosity of a star-forming region seen at redshift z:

$$l_B = f_\nu \Delta\nu (1 + z) L / 4\pi d_L^2 , \tag{42}$$

where $f_\nu(z)$ is the fraction of the B bandwidth which is longward of the redshifted Lyman break, $\Delta\nu = 1.6 \times 10^{14}$ Hz is the effective bandwidth at B, L comes from equation (32), and d_L is the standard luminosity distance,

$$d_L = 2c[1 + z - (1 + z)^{1/2}]/H_0 .$$

This then gives the apparent magnitude at B as

$$m_B = -13.0 - 2.5 \log (l_B / 1 \text{ erg s}^{-1} \text{ cm}^{-2}) . \tag{43}$$

Equations (32), (42), and (43) allow us to solve for the star formation rate, $\dot{M}_*(m_B, z)$, which produces an observed apparent magnitude, m_B, for an object at redshift z. Comparison with equation (23) or equation (24) then gives the circular velocity, $V_c(m_B, z)$, of halos with this star formation rate. The expected count (per steradian) of images brighter than m_B may then be written as

$$N_c(m_B) = \int_0^{4.4} dz \int_{V_c(m_B, z)}^{\infty} dV\, n(V, z) F_b(V) (1 + z)^{-7/2} d_L^2 c / H_0 . \tag{44}$$

Here F_b is the fraction of halos which are bursting at any given time in the model of equation (24); $F_b = 1$ in the nonbursting model. Equation (44) will underestimate the true count, particularly at brighter apparent magnitudes, because we have neglected the contribution of older stars to the observed flux.

Our model for bursting was constructed so as to have no effect on the time-averaged star formation rate within halos of any given circular velocity. However, it substantially modifies the abundance and the luminosity of "active" halos at any given time. As a result, it also changes our predictions for the UV luminosity function of galaxies and so for counts of faint galaxies. This is illustrated in Figure 13, where the abundance of halos as a function of their star formation rate is compared for bursting and nonbursting versions of a model with $b = 2.5$, $\Omega_b = 0.2$, and heavy-element enrichment. We have converted $\dot{M}_*$ to a young-star luminosity using the model of equation (32) with a bandwidth of 1.6×10^{14} Hz. These curves have sharp features which can be traced back to changes in slope in the cooling functions. When bursting is included, the features become narrower and move to brighter luminosity and lower abundance. They thus trace the way in which star-forming regions of a specific circular velocity contribute to the luminosity function in the bursting and nonbursting cases. The overall effect of such shifts is to brighten the characteristic luminosity and flatten the faint-end slope of the luminosity functions. At first sight this seems like a very encouraging change, since the "burst" functions agree much better with observation than the luminosity functions plotted in Figures 8 and 9. Unfortunately, the amount of bursting seems too extreme to be plausible; for example, a galaxy like our own, with $V_c = 220$ km s^{-1}, would have a star formation efficiency $\epsilon = 0.12$, and thus is assumed to be bursting, and so visible, only 12% of the time. Nevertheless, since a major contribution to the faint end of the observed luminosity function comes from irregular galaxies which may form most of their stars in a bursting mode, this kind of effect might go some way toward reconciling our models with observation. Broadhurst et al. (1988) and Colless et al. (1991) interpret their spectroscopic data as indicating a considerable increase in bursting behavior in the fainter part of the luminosity function at redshifts greater than 0.2.

The inclusion or otherwise of bursting has a substantial effect on the counts predicted by equation (44). Results for all our standard models are shown for both the bursting and nonbursting cases in Figure 14, and are compared with a power-law fit to Tyson's (1988) deep blue counts. At brighter magnitudes the integral count increases with increasing Ω_b and with the inclusion of enrichment, and is little affected by chang-

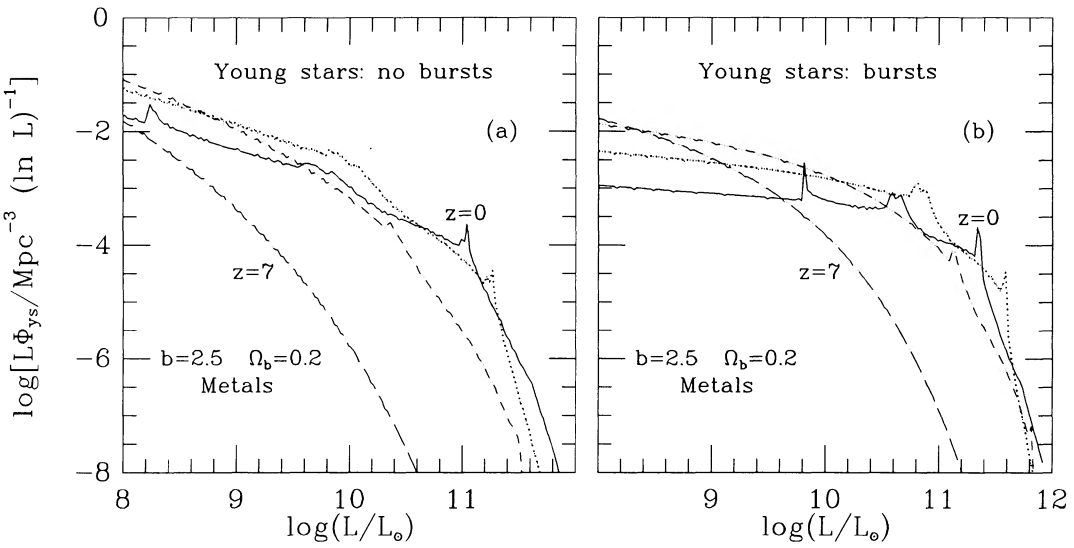

FIG. 13.—Evolution of the luminosity function of actively star-forming regions. Results are shown for the model with $\Omega_b = 0.2$, biasing parameter $b = 2.5$, and metal enrichment. The different curves correspond to redshifts $z = 7$ (*long-dashed line*), $z = 3$ (*short-dashed line*), $z = 1$ (*dotted line*), and $z = 0$ (*solid line*). (*a*) Continuous star formation. (*b*) Star formation in bursts.

ing the fluctuation amplitude or by bursting. For a number of models, particularly those which have a relatively large number of bright objects and include bursting, the slope of the counts flattens significantly at fainter magnitudes. There is no sign of any such change in the real data for $m_B < 26$. A turnover must be present at fainter magnitudes but has yet to be convincingly detected. The apparent luminosity we predict for an object is proportional to the parameter L_{uv} of equation (32); the uncertainty in the zero point of the magnitude axis in Figure 14 is thus at least 1 mag. In addition, a sufficient fraction of the brighter objects are at relatively low redshift (see below) that the contribution of older stars to their apparent

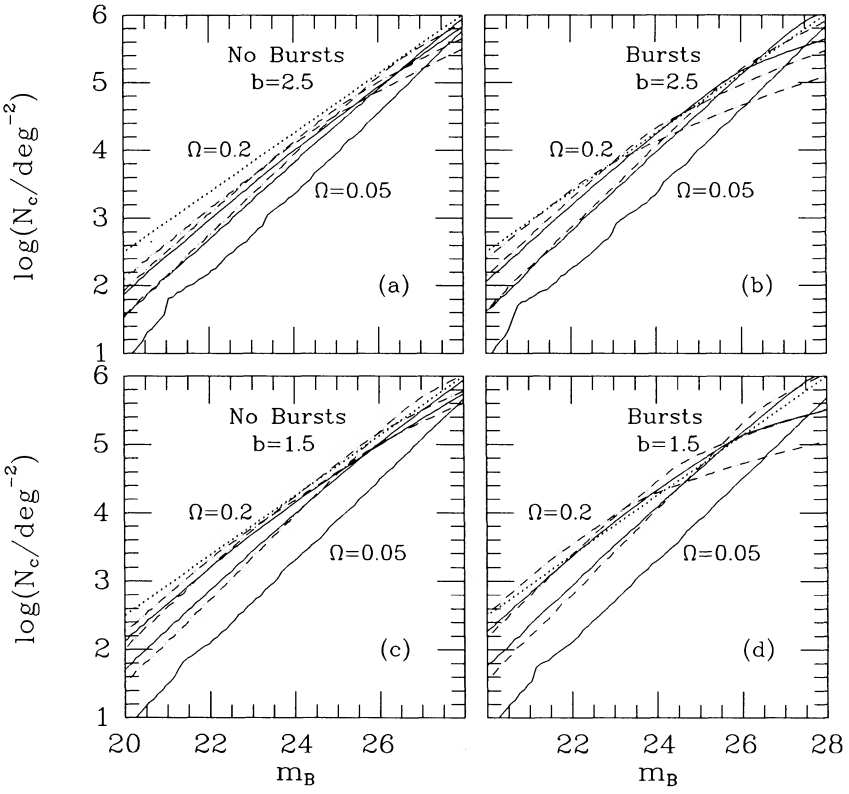

FIG. 14.—Counts of faint galaxies as a function of apparent magnitude in the *B* band. Solid and dashed curves in each panel correspond to the six standard no-enrichment and enrichment models, respectively, described in Table 1. The dotted line is a power-law fit to Tyson's (1988) counts. (*a*) Continuous star formation; $b = 2.5$. (*b*) Star formation in bursts; $b = 2.5$. (*c*) Continuous star formation; $b = 1.5$. (*d*) Star formation in bursts; $b = 1.5$.

magnitudes will be quite substantial. With these uncertainties in mind, it is clear that a number of the models can fit Tyson's counts. Only the most extreme cases seem to be in clear conflict with the data. With a slight brightening of the zero point, the nonbursting model of Figure 13 ($b = 2.5$, $\Omega = 0.2$, and heavy-element enrichment) gives one of the better fits to the observed counts. Its bursting counterpart, on the other hand, has a very different slope than the data.

Our models have no serious difficulty reproducing the total count at faint magnitudes, even though they assume a flat universe. The implication of this result for the arguments of Koo (1989) and Guiderdoni & Rocca-Volmerange (1989) is unclear, since part of this "success" may result from the unacceptable present-day luminosity function of our models. On the other hand, part of the success also undoubtedly comes from the fact that we do not assume galaxies to be conserved, and leave open the exact relation between the population of objects seen at high redshift and observed nearby galaxies. It is possible that a substantial amount of galaxy merging has occurred in the real universe, thus violating the fundamental assumption underlying traditional models for the evolution of the galaxy population (Tinsley 1980b; Guiderdoni & Rocca-Volmerange 1989). Certainly in the CDM cosmogony halos merge profusely (Frenk et al. 1988), although the merging rate of galaxies remains an open question. Notice that the success of our count models hinges on the UV luminosity of young stars being able to escape from each star-forming region. The effects of dust must therefore be relatively weak. This conclusion is independent of our detailed models. It follows from Cowie's (1988) conclusion that the *observed* total flux of faint blue galaxies corresponds to the UV emission from enough massive stars to form a large fraction of the total heavy-element content of the universe. Thus a substantial amount of the UV radiation associated with galaxy formation apparently escapes absorption.

The behavior of the counts in Figure 14 can be understood more clearly by studying the redshift distribution of the counted objects. In Figure 15 we show these distributions for the two models of Figure 13. Recall that the counts for the nonbursting model are close to a power law in magnitude, whereas in the bursting counterpart there are more bright galaxies and fewer faint ones, with a pronounced change in slope near $m_B = 24$. The curves of Figure 15 give redshift distributions for objects in 2 mag bins centered on $m_B = 21$, 23, 25, and 27. The additional counts at brighter magnitudes in the bursting model are due to the larger characteristic luminosity in this model (see the luminosity function of Figure 13). This allows more systems to be counted at high redshift. The count deficit at fainter magnitudes is due to the elimination of distant objects by the Lyman break, in combination with a reduction in the number of nearby objects caused by the shallower faint-end slope of its luminosity function. Indeed, the flatness of the functions in Figure 13*a* causes the curves of Figure 15*b* to coincide at low redshift. The bend in the counts can thus be ascribed to the characteristic distance of a galaxy approaching the limit imposed by the Lyman break in a model where the luminosity function, like that observed, has a relatively small number of faint systems.

Spectroscopic data are now available for samples of blue galaxies down to $m_B = 23.5$. For 187 galaxies with $20 < b_J < 21.5$, Broadhurst et al. (1988) find a median redshift of 0.21. Near $m_B = 23$, Colless et al. (1990) find five redshifts in the range 0.28–0.87 and three featureless spectra in a sample of eight, while Cowie & Lilly (1990) find four redshifts below 0.4 and two flat featureless spectra in another sample of eight. This suggests a median redshift in the range 0.3–0.6. The median redshifts of the first two histograms of Figure 15 are 0.28 and 0.58 for the nonbursting case but 0.62 and 1.12 for the model with bursts. The first of these seems compatible with the data, but the second almost certainly is not. The nonbursting case is also the one which fits the integral counts relatively well. A nonbursting model with $b = 1.5$, $\Omega_b = 0.1$, and enrichment fits both the count data and the observed galaxy luminosity functions slightly better than the $b = 2.5$, $\Omega_b = 0.2$ models of Figures 13 and 15, but it has median redshifts of 0.42 and 0.82 at $m_B = 21$ and 23. These seem too distant to be acceptable.

In Table 2 we give median redshifts for the four histograms of Figure 15 for all our models; both bursting and nonbursting

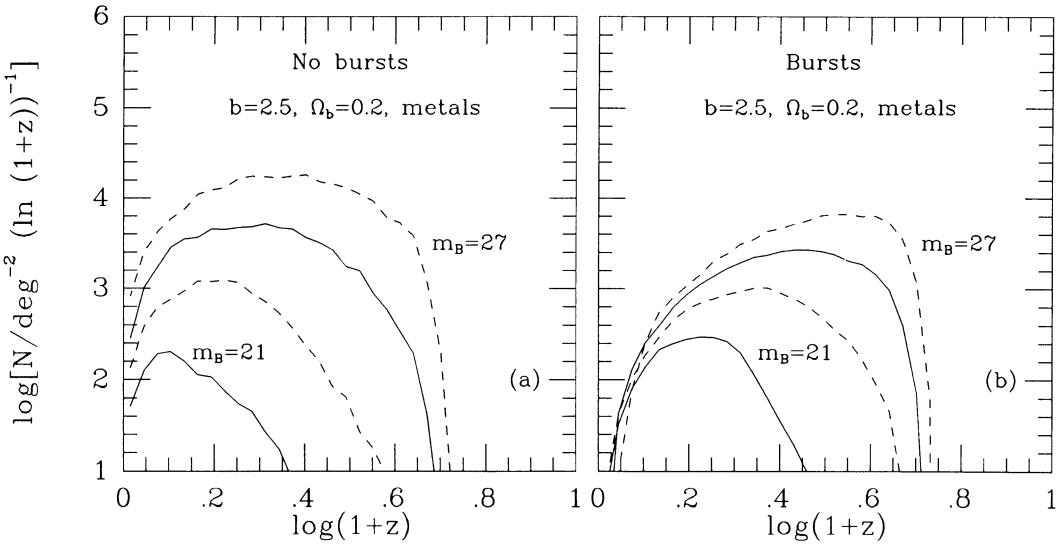

FIG. 15.—Redshift distribution of faint galaxy counts. Results are shown for models with $\Omega_b = 0.2$, biasing parameter $b = 2.5$, and metal enrichment (as in Fig. 13). Each curve gives the distribution for galaxies in two magnitude bins centered on $m_B = 21$, 23, 25, and 27, respectively. (*a*) Continuous star formation. (*b*) Star formation in bursts.

TABLE 2

MEDIAN REDSHIFTS OF FAINT GALAXIES

				z_h (burst)				z_h (nonburst)			
Ω_b	b	y	ϵ_0	21	23	25	27	21	23	25	27
0.20.....	2.5	0	0.22	0.27	0.60	1.44	2.02	0.23	0.45	0.73	1.15
0.10.....	2.5	0	0.04	0.10	0.28	0.64	1.64	0.11	0.26	0.53	1.01
0.05.....	2.5	0	0.002	0.01	0.07	0.17	0.56	0.01	0.06	0.18	0.51
0.20.....	1.5	0	0.45	0.60	1.23	1.99	2.57	0.49	0.93	1.39	1.67
0.10.....	1.5	0	0.06	0.24	0.56	1.11	2.13	0.25	0.59	1.12	1.54
0.05.....	1.5	0	0.001	0.03	0.13	0.31	0.65	0.03	0.12	0.32	0.70
0.20.....	2.5	3.7	0.81	0.62	1.12	1.58	2.02	0.28	0.58	0.89	1.16
0.10.....	2.5	2.2	0.22	0.32	0.84	1.35	1.83	0.21	0.52	0.92	1.20
0.05.....	2.5	1.2	0.03	0.10	0.31	0.92	1.47	0.08	0.28	0.74	1.21
0.20.....	1.5	3.1	1.37	0.94	1.74	2.22	2.51	0.63	0.99	1.48	1.74
0.10.....	1.5	1.8	0.33	0.48	1.24	2.03	2.39	0.42	0.82	1.52	1.78
0.05.....	1.5	1.05	0.03	0.12	0.40	1.25	1.99	0.12	0.38	1.10	1.77

cases are included. At brighter magnitudes the typical redshift of objects increases strongly with Ω_b and with inclusion of heavy elements, and rather more weakly with fluctuation amplitude and with the inclusion of bursting. Bursting only has strong effects in models where the efficiency of feedback is high (cf. eqs. [23] and [24]). All the bursting models with heavy-element cooling seem to give characteristic redshifts which are too high at $m_B = 21$ and 23, but a couple of the models with no enrichment seem acceptable. In the continuous star formation case, models with either value of b, and both with and without enrichment, can give values in the observed range. In general, the models which give acceptable redshifts are also those which give integral counts with approximately the right slope, and have luminosity functions in at least rough agreement with observation. Notice that none of the $\Omega_b = 0.05$ models is acceptable either from the point of view of the faint galaxy data (Fig. 14 and Table 2) or from that of the luminosity functions (Fig. 9 and Table 1).

7. DISCUSSION

Our main goal in this paper has been to set up methods to calculate the characteristics of the galaxy population in a universe where galaxy formation occurs gravitationally as originally envisaged by White & Rees (1978). Gas cools, condenses, and forms stars within a hierarchically clustering distribution of dark matter. We have tried to adopt the simplest assumptions consistent with current understanding of these processes as provided by N-body simulations of dissipationless clustering, and by analytic and numerical studies of radiative and hydrodynamic processes in the gas. Despite this, a disheartening number of ingredients must be assembled to produce a plausibly complete recipe for galaxy formation. The initial conditions require specifying mean densities for the dark matter and residual gas (Ω and Ω_b), and an initial fluctuation distribution, usually taken to have Fourier components with random phase and a given power spectrum, $|\delta_k|^2$. Although this suffices for a full specification of the initial state, treatment of subsequent evolution requires models for the nonlinear growth of the clustering hierarchy, for the radiative cooling and condensation of gas in dark halos, for star formation and its heating and enrichment of surrounding gas, and for the mass function, radiative output, and evolution of the stellar population. While we adopt standard and plausible models for each of these ingredients, the accumulated uncertainties are large, even for a well-specified set of initial conditions such as the flat cold dark matter model on which we concentrate.

The methods we have developed allow us to calculate a number of properties of galaxies. Our basic premise for the evolution of the gas predicts atmospheres surrounding galaxies and galaxy clusters which should be observable in X-rays. Cluster atmospheres are seen, of course, and *ROSAT* should allow a sensitive search for corresponding structures around the most massive isolated galaxies. Infall from these gas halos produces cooling flows in some clusters and may be responsible for gas replenishment in the disks of many spirals. The star formation rates we predict are consistent with those observed in nearby spirals. However, unless Ω_b is very small, this requires substantial heating of the halo gas, presumably as a result of star formation. This feedback may cause significant contamination by processed material, and so may offer explanations for the [Mg II] absorption lines associated with the halos of high-redshift galaxies, and for the heavy-element content of the intergalactic gas in rich clusters. With optimistic assumptions, star formation in our models can contribute substantially to maintaining the ionization of the intergalactic medium. Models of the kind we consider also predict an abundance of massive objects which seems compatible with current estimates of the abundance of high-redshift quasars and gravitational lenses (Efstathiou & Rees 1988; Cole & Kaiser 1989; Narayan & White 1988). A detailed treatment of star formation can predict the present repartition of stars among galaxies or among virialized structures in general; the former calculation is complicated by uncertainties in how galaxies merge. The distribution of star-forming regions at greater distance can be compared rather directly with counts of faint blue galaxies. Thus our models produce luminosity functions for galaxies and for galaxy aggregates, and predict systematic relations between their properties (i.e., luminosity, circular velocity, metallicity, gas content, and total or stellar mass-to-light ratio).

With such a wealth of assumptions and predictions, it is not easy to reach a balanced assessment of any particular model. In the present paper we have given results exclusively for CDM models with $\Omega = 1$ and $H_0 = 50$ km s^{-1} Mpc^{-1}. This narrows down the choice significantly, but still leaves a number of parameters: the fluctuation amplitude or "bias parameter" b, the baryon fraction Ω_b, the efficiency of feedback ϵ_0, and the

heavy-element yield y. The last two are tied to specific models for gas heating and enrichment, and we fixed them a priori by requiring that our models match the observed mean luminosity density and stellar metallicity. We have three further parameters which characterize the evolution of stellar populations, and we study star formation in bursting and nonbursting modes. Within this framework we have been able to isolate the major parameter dependences of the predicted galaxy population. In fact, only for a rather small range of parameters do our models appear within striking distance of success. Furthermore, this range does not include the low values of Ω_b preferred by some recent analyses of cosmological nucleosynthesis.

For the fluctuation amplitudes required to fit galaxy clustering, flat CDM models predict the mass of the universe to be predominantly in relatively massive nonlinear objects. Thus the standard highly biased case, $b = 2.5$, predicts a mean density $\Omega > 0.5$ in halos with circular velocities exceeding 100 km s^{-1}, and $\Omega > 0.1$ for $V_c > 430$ km s^{-1}. For the higher amplitude, $b = 1.5$, favored by some recent papers, the corresponding numbers are $\Omega > 0.5$ in halos with $V_c > 240$ km s^{-1} and $\Omega > 0.1$ for $V_c > 850$ km s^{-1}. For residual baryon fractions as large as those suggested by observations of rich clusters, and in the absence of heating, almost all the gas would cool during the hierarchical buildup of such massive systems. Yet less than 1% of the closure density is observed to be in the form of stars or cold gas. Our models therefore require substantial energy input to keep the gas hot. If this heating is associated with massive stars, it causes star formation to be biased toward massive systems and late times. Such models therefore produce relatively bright galaxies at relatively low redshifts. On the other hand, if Ω_b is as small as suggested by nucleosynthesis arguments, gas can cool efficiently only at early times and in small systems. Feedback must be weak, and such models form stars early and produce galaxies which are much less luminous than those observed. Increasing the fluctuation amplitude increases the typical redshift of star formation but has a rather weak effect on galaxy luminosites. The difficulty is alleviated, but not eliminated, by including heavy elements to enhance cooling rates, even though our models probably exaggerate their effect. Thus our whole paradigm is in trouble for $\Omega_b \lesssim 0.05$: too little gas cools in massive halos to make the observed central galaxy. The situation could perhaps be saved by arguing that gas is concentrated to the centers of halos at an early stage of clustering and remains partially segregated at later epochs. This is not allowed by the assumptions of our current models.

Many aspects of our models work considerably better for larger values of Ω_b. Such models require efficient feedback from star formation. The resulting suppression of star formation in low-mass systems has a number of desirable consequences. Luminosity functions have a shallower faint-end slope and a higher characteristic luminosity. The luminosity–circular velocity and luminosity-metallicity relations for galaxies are of the correct form and have relatively small scatter. Star formation as a whole is concentrated in the highest mass halos, producing a bias in the galaxy distribution in the sense required to reconcile a flat universe with observations of galaxy clustering. Unfortunately, the particular models we have investigated do not produce a sufficiently strong bias, nor can they give a galaxy luminosity function with a faint end as flat as that observed. This latter problem is generic to hierarchical clustering models. Its elimination would seem to require much more efficient suppression of star formation in small objects than can be achieved with the scheme of this paper. The large baryon fractions for which we get best results seem incompatible with cosmological nucleosynthesis. Together with the insufficient bias we find, and the large ratios of baryons to dark matter in observed rich clusters, this could be taken as an argument in favor of an $\Omega < 1$ universe, perhaps with a cosmological constant.

One major property of the galaxy population which we have not addressed explicitly is the existence of clearly separated disk and bulge components. The extreme axial ratio and the highly ordered motions observed in disks argue strongly that they must have been assembled while gaseous, and that they have been at most weakly disturbed since the bulk of the stars formed. Assembly may well occur continuously rather than in a well-defined event, in the manner advocated by Gunn (1982); note that in our own Galaxy the bulk of the disk stars formed comparatively recently. In the context of the theory we have discussed, these properties suggest that disks must settle within a halo which has already reached equilibrium, and form stars on a somewhat longer time scale than that of halo collapse and gaseous dissipation. In addition, disks must have avoided major mergers since the formation of most of their stars (see also White 1991). Systems where stars formed primarily before or during collapse are expected to be seen as ellipticals or as the bulges of spirals. Disks form around bulges from weakly bound residual gas or from newly accreted material, a process which can occur only if the galaxy's halo remains relatively undisturbed and continues to gain mass. In high-density regions interactions will disrupt the outer parts of halos, sweep them clear of gas, and so inhibit or prevent disk formation. In such regions mergers of galaxies can also produce ellipticals, even if the progenitor systems had disks. A qualitative understanding of galaxy morphology and its relation to environment thus seems available in our model; a quantitative treatment should be possible by extension of our methods. It will clearly require supplementary hypotheses to differentiate "disk" and "bulge" formation, to specify how the two are related within a single system, and to allow an explicit treatment of galaxy mergers. Some observed features, for example, the fact that bulges tend to be smaller and older than disks, will clearly be explained naturally by this scheme. Others, such as the fact that the most massive systems tend to be ellipticals rather than disks, may be more difficult.

Although a final assessment of our models still seems beyond reach, there are several obvious areas for further study. One is the problem with excessively steep luminosity functions. As discussed in § 6.4, it may be partially alleviated by the effects of bursting. There is already evidence that bursting is a major mode of star formation in many dwarf galaxies and may dominate observed blue galaxy samples at redshifts beyond 0.2 (Tyson & Scalo 1988; Broadhurst et al. 1988; Colless et al. 1989; Colless et al. 1990). Lacey & Silk (1991) show that the problem may be avoided if star formation terminates entirely in most low V_c galaxies at $z > 1$; they then fade to very low luminosity by the present and could perhaps be identified with low surface brightness dwarf ellipticals. Galaxy merging may also have a substantial effect, although the AGS luminosity functions of § 6.2 show that it cannot cure the problem entirely. To address these questions it is important to tie down the faint end of the observed galaxy luminosity function more securely. This would be especially helpful in the red or near-IR where light from young stars is less significant, but unfortunately

surface brightness selection effects are exacerbated at these wavelengths.

The recent formation of most stars in our models suggests other fruitful areas for comparison with observation. While it seems plausible that stars in disks formed recently, there is mounting evidence for the standard assumption that stars in ellipticals and bulges formed at high redshift (Gunn 1989; Lilly 1989; but cf. Chambers & Charlot 1990). Further study of this question requires spectroscopy of faint samples of randomly selected red galaxies, together with careful modeling of their spectral evolution. In our models (as in any hierarchical model) small galaxies tend to form first. However, where observation sheds any light on the relative ages of galaxies, it suggests that small galaxies are younger. Any resolution of this discrepancy will probably be linked to the merging and bursting which seem required to get an acceptable luminosity function. Wolfe's (1988) discovery that high–column-density absorbing clouds at $z \sim 2$ contain sufficient neutral gas to form *all* the stars seen nearby is another, perhaps related puzzle. While it suggests that much late star formation may indeed occur, such a large quantity of cold gas is unexpected at any redshift in our models and may imply the existence of a latency period between gas cooling and star formation. If future data show that the late galaxy formation predicted by our models is indefensible, a simple generalization would be to allow $\Omega < 1$. Such open universes were considered in the original paper of White & Rees (1978); their evolution is similar to that of the models studied here, but formation activity is shifted to higher redshift. Open models have the additional virtue of eliminating the conflict between the large baryon fractions observed in clusters and required for efficient cooling, and the low values of Ω_b suggested by cosmological nucleosynthesis. However, there are many other plausible variations on the ideas of this paper, and it would clearly be premature to abandon the assumption that $\Omega = 1$. In fact, the rapid and recent evolution predicted in a flat universe is both eminently observable and perhaps already observed. Relevant data are accumulating very quickly, and we can except many lively debates before this issue is resolved.

We thank Ed Bertschinger, Rob Kennicutt, Joel Primack, and Martin Rees for helpful discussions. This work was supported by grants from NASA's Astrophysical Theory Program and the NSF, by the SERC, and by a NATO travel award. S. D. M. W. is grateful to the Institute for Advanced Study of the Hebrew University of Jerusalem for their hospitality during the completion of this paper.

REFERENCES

Aarseth, S. J., Gott, J. R., & Turner, E. L. 1979, ApJ, 236, 43
Applegate, J. H., & Hogan, C. J. 1985, Phys. Rev. D, 31, 3037
Arimoto, N., & Tarrab, I. 1990, A&A, 228, 6
Arimoto, N., & Yoshii, Y. 1987, A&A, 173, 23
Ashman, K., & Carr, B. J. 1991, MNRAS, 249, 13
Bahcall, N. A. 1979, ApJ, 232, 689
Bardeen, J. M. 1986, in Inner Space/Outer Space, ed. E. W. Kolb, M. S. Turner, D. Lindley, K. Olive, & D. Seckel (Chicago: Univ. of Chicago Press), 212
Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15
Barnes, J. 1985, MNRAS, 215, 517
———. 1990, in Dynamics and Interactions of Galaxies, ed. R. Wielen (Berlin: Springer-Verlag), 186
Baron, E., & White, S. D. M. 1987, ApJ, 322, 585
Bergeron, J. 1988, in The Post Recombination Universe, ed. N. Kaiser & A. Lasenby (Dordrecht: Kluwer), 201
Bertschinger, E. 1989, ApJ, 340, 666
Binney, J. 1977, ApJ, 215, 483
Binney, J., & Tremaine, S. 1987, Galactic Dynamics (Princeton: Princeton Univ. Press)
Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 527
Blumenthal, G. R., & Primack, J. 1983, in Fourth Workshop on Grand Unification, ed. H. A. Weldon, P. Langacker, & P. J. Steinhardt (Boston: Birkhauser), 256
Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, in press
Bower, R. J. 1991, MNRAS, 248, 332
Broadhurst, T. J., Ellis, R. S., & Shanks, T. 1988, MNRAS, 235, 827
Bruzual, G. 1983, ApJ, 273, 105
Butcher, H., & Oemler, A. 1978, ApJ, 219, 18
Carlberg, R. G. 1991, ApJ, 367, 385
Carlberg, R., & Couchman, H. 1989, ApJ, 340, 47
Chambers, K. C., & Charlot, S. 1990, ApJ, 348, L1
Chambers, K. C., Miley, G. K., & Joyce, R. R. 1988, ApJ, 329, L75
Chevalier, R. A., & Clegg, A. W. 1985, Nature, 317, 44
Cole, S. 1991, ApJ, 367, 45
Cole, S., & Kaiser, N. 1989, MNRAS, 237, 1127
Colless, M., Ellis, R. S., & Taylor, K. 1989, in The Epoch of Galaxy Formation, ed. C. S. Frenk, R. S. Ellis, T. Shanks, A. F. Heavens, & J. A. Peacock (Dordrecht: Kluwer), 359
Colless, M., Ellis, R. S., Taylor, K., & Hook, R. N. 1990, MNRAS, 244, 408
Cowie, L. L. 1988, in The Post Recombination Universe, ed. N. Kaiser & A. Lasenby (Dordrecht: Kluwer), 1
Cowie, L. L., Gardner, J. P., Lilly, S. J., & McLean, I. 1990, ApJ, 360, L1
Cowie, L. L., & Lilly, S. J. 1990, in Evolution of the Universe of Galaxies, ed. R. Kron (ASP Conf. Ser. 10; San Francisco: ASP), 212
Davis, M., Efstathiou, G., Frenk, C., & White, S. 1985, ApJ, 292, 371
Dekel, A., & Silk, J. 1986, ApJ, 303, 39
Demarque, P., & McClure, R. D. 1977, in The Evolution of Galaxies and Stellar Populations, ed. B. M. Tinsley & R. B. Larson (New Haven: Yale Univ. Obs.), 199
Eder, J. A., Schombert, J. M., Dekel, A., & Oemler, A. 1989, ApJ, 340, 29
Efstathiou, G., Ellis, R. S., & Peterson, B. A. 1988a, MNRAS, 232, 431
Efstathiou, G., Frenk, C. S., White, S. D. M., & Davis, M. 1988b, MNRAS, 235, 715
Efstathiou, G., & Rees, M. J. 1988, MNRAS, 230, 5P
Einasto, J., Kaasik, A., & Saar, E. 1974, Nature, 250, 309
Ellis, R. S. 1987, in Observational Cosmology, ed. A. Hewitt, G. Burbidge, & L.-Z. Fang (Dordrecht: Reidel), 367
Evrard, A. E. 1987, ApJ, 316, 36
———. 1989, ApJ, 341, L71
Evrard, A. E., & Davis, M. 1988, Nature, 333, 355
Fabbiano, G. 1989, ARA&A, 27, 87
Faber, S. 1973, ApJ, 179, 731
———. 1982, in Astrophysical Cosmology, ed. H. A. Bruck, G. V. Coyne, & M. S. Longair (Vatican City: Ponti. Acad. Sci.), 191
Faber, S., & Burstein, D. 1988, in Large-Scale Motions in the Universe, ed. V. C. Rubin & G. V. Coyne (Princeton: Princeton Univ. Press), 116
Fabian, A. C., Nulsen, P. E. J., & Canizares, C. R. 1984, Nature, 310, 733
———. 1991, Astr. Ap. Rev., 2, 191
Fabricant, D., & Gorenstein, P. 1983, ApJ, 267, 535
Fall, S. M. 1979, Nature, 281, 200
Fall, S. M., & Efstathiou, G. 1980, MNRAS, 193, 189
Fall, S. M., & Pei, Y. C. 1989, ApJ, 337, 7
Ferguson, H., & Sandage, A. 1988, AJ, 96, 1520
Freeman, K. C. 1989, in The Epoch of Galaxy Formation, ed. C. S. Frenk, R. S. Ellis, T. Shanks, A. F. Heavens, & J. A. Peacock (Dordrecht: Kluwer), 331
Frenk, C. S., White, S., Efstathiou, G., & Davis, M. 1985, Nature, 317, 595
———. 1988, ApJ, 327, 507
———. 1990, ApJ, 351, 10
Gerola, H., Seiden, P. E., & Schulman, L. S. 1980, ApJ, 242, 517
Gilmore, G., Wyse, R. F. G., & Kuijken, K. 1989, ARA&A, 27, 555
Green, R. 1989, in The Epoch of Galaxy Formation, ed. C. S. Frenk, R. S. Ellis, T. Shanks, A. F. Heavens, & J. A. Peacock (Dordrecht: Kluwer), 121
Groth, J. E., Juszkiewicz, R., & Ostriker, J. P. 1989, ApJ, 346, 558
Guiderdoni, B., & Rocca-Volmerange, B. 1987, A&A, 186, 1
———. 1989, in The Epoch of Galaxy Formation, ed. C. S. Frenk, R. S. Ellis, T. Shanks, A. F. Heavens, & J. A. Peacock (Dordrecht: Kluwer), 367
Gunn, J. E. 1982, in Astrophysical Cosmology, ed. H. A. Bruck, G. V. Coyne, & M. S. Longair (Vatican City: Pontif. Acad. Sci.), 233
———. 1989, in The Epoch of Galaxy Formation, ed. C. S. Frenk, R. S. Ellis, T. Shanks, A. F. Heavens, & J. A. Peacock (Dordrecht: Kluwer), 167
Heckman, T., Armus, L., & Miley, G. 1990, ApJS, 74, 833
Hughes, J. P. 1989, ApJ, 337, 21
Impey, C. D., Bothun, G. D., & Malin, D. F. 1988, ApJ, 330, 634
Kaiser, N. 1984, ApJ, 284, L9
———. 1986, in Inner Space/Outer Space, ed. E. W. Kolb, M. S. Turner, D. Lindley, K. Olive, & D. Seckel (Chicago: Univ. Chicago Press), 258
Kaiser, N., & Lahav, O. 1989, MNRAS, 237, 129
Kennicutt, R. C. 1983, ApJ, 272, 54
Koo, D. 1986, ApJ, 311, 651

Koo, D. 1989, in The Epoch of Galaxy Formation, ed. C. S. Frenk, R. S. Ellis, T. Shanks, A. F. Heavens, & J. A. Peacock (Dordrecht: Kluwer), 71
Kron, R. 1980, ApJS, 43, 305
Kurki-Suonio, H., Matzner, R. A., Olive, K. A., & Schramm, D. N. 1990, ApJ, 353, 406
Lacey, C., & Silk, J. 1991, ApJ, in press
Larson, R. B. 1974a, MNRAS, 166, 585
———. 1974b, MNRAS, 169, 229
Larson, R. B., Tinsley, B. M., & Caldwell, C. N. 1980, ApJ, 237, 692
Lilly, S. 1988, ApJ, 333, 161
———. 1989, ApJ, 340, 77
Lynden-Bell, D., Faber, S. M., Burstein, D., Davies, R. L., Dressler, A., Terlevich, R. J., & Wegner, G. 1988, ApJ, 326, 19
Majewski, S. R. 1989, in The Epoch of Galaxy Formation, ed. C. S. Frenk, R. S. Ellis, T. Shanks, A. F. Heavens, & J. A. Peacock (Dordrecht: Kluwer), 85
McCarthy, P. J., van Breugel, W., Spinrad, H., & Djorgovski, S. 1987, ApJ, 321, L29
McKee, C., & Ostriker, J. P. 1977, ApJ, 218, 148
Merritt, D. 1987, ApJ, 313, 121
Miller, G. E., & Scalo, J. M. 1979, ApJS, 41, 513
Moore, B., Frenk, C. S., & White, S. D. M. 1991, Durham preprint
Narayan, R., & White, S. D. M. 1988, MNRAS, 231, 97P
Olive, K. A., Schramm, D. N. Steigman, G., & Walker, T. P. 1990, Phys. Letters B, 236, 454
Ostriker, J. P., & Cowie, L. 1981, ApJ, 338, 579
Ostriker, J. P., Peebles, J. P., & Yahil, A. 1974, ApJ, 193, L1
Peacock, J., & Heavens, A. 1989, MNRAS, 243, 133
Peebles, P. J. E. 1970, AJ, 75, 13
———. 1971, Ap&SS, 11, 443
———. 1973, PASJ, 25, 291
———. 1974, ApJ, 189, L51
———. 1980, The Large-Scale Structure of the Universe (Princeton: Princeton Univ. Press)
———. 1982, ApJ, 258, 415
Peebles, P. J. E., & Dicke, R. 1968, ApJ, 154, 891
Persson, S. E., Frogel, J. A., & Aaronson, M. 1979, ApJS, 39, 61
Peterson, B. A., Ellis, R. S., Kibblewhite, E. J., Bridgeland, M. T., Hooley, T., & Horne, D. 1979, ApJ, 233, L109
Pettini, M., Boksenberg, A., & Hunstead, R. W. 1989, in The Epoch of Galaxy Formation, ed. C. S. Frenk, R. S. Ellis, T. Shanks, A. F. Heavens, & J. A. Peacock (Dordrecht: Kluwer), 107
Phillipps, S., Disney, M. J., Kibblewhite, E. J., & Cawson, M. G. M. 1987, MNRAS, 229, 505
Pierce, M. J., & Tully, R. B. 1988, ApJ, 330, 579
Press, W. H., & Schechter, P. L. 1974, ApJ, 187, 425
Rees, M. J., & Ostriker, J. P. 1977, MNRAS, 179, 541
Sandage, A., Binggeli, B., & Tammann, G. A. 1985, AJ, 90, 1759
Sarazin, C. L. 1986, Rev. Mod. Phys., 58, 1
Scalo, J. M. 1986, Fund. Cosmic Phys., 11, 1
Schaeffer, R., & Silk, J. 1985, ApJ, 292, 319
Schechter, P. L., & Dressler, A. 1987, AJ, 94, 563
Shanks, T., Metcalfe, N., Fong, R., & Hale-Sutton, D. 1991, in preparation
Shanks, T., Stevenson, P. R. F., Fong, R., & McGillivray, H. T. 1983, MNRAS, 206, 767
Shapiro, P. R., & Giroux, M. L. 1989, ApJ, 321, L107
Silk, J. 1977, ApJ, 211, 638
Silk, J., & Norman, C. A. 1981, ApJ, 247, 60
Spinrad, H. 1989, in The Epoch of Galaxy Formation, ed. C. S. Frenk, R. S. Ellis, T. Shanks, A. F. Heavens, & J. A. Peacock (Dordrecht: Kluwer), 39
Stewart, G. C., Canizares, C. R., Fabian, A. C., & Nulsen, P. E. J. 1984, ApJ, 278, 536
Struck-Marcell, C., & Scalo, J. M. 1987, ApJS, 64, 39
The, L. S., & White, S. D. M. 1986, AJ, 92, 1248
———. 1988, AJ, 95, 15
Thomas, P. A., & Fabian, A. C. 1990, MNRAS, 246, 156
Thomas, P. A., Fabian, A. C., Arnaud, K. A., Forman, W., & Jones, C. 1986, MNRAS, 222, 655
Thuan, T. X. 1985, ApJ, 299, 881
Tinsley, B. M. 1978, ApJ, 222, 44
———. 1979, ApJ, 229, 1046
———. 1980a, Fund. Cosmic Phys., 5, 287
———. 1980b, ApJ, 241, 41
Tyson, J. A. 1988, AJ, 96, 1
Tyson, J. A., & Jarvis, J. F. 1979, ApJ, 320, L153
Tyson, N. D., & Scalo, J. M. 1988, ApJ, 329, 618
Visvanathan, N., & Sandage, A. 1977, ApJ, 216, 214
West, M. J., & Richstone, D. O. 1988, ApJ, 335, 532
White, S. D. M. 1976, MNRAS, 177, 717
———. 1986, In Inner Space/Outer Space, ed. E. W. Kolb, K. Olive, & D. Seckel (Chicago: Univ. Chicago Press), 228
———. 1989, in The Epoch of Galaxy Formation, ed. C. S. Frenk, R. S. Ellis, T. Shanks, A. F. Heavens, & J. A. Peacock (Dordrecht: Kluwer), 15
———. 1990, in The Interstellar Medium in Galaxies, ed. H. A. Thronson, J. M. Shull (Dordrecht: Kluwer), 371
———. 1991, in The Dynamics of Galaxies and the Molecular Cloud Distribution, ed. F. Combes & F. Casoli (Dordrecht: Kluwer), 383
White, S. D. M., Davis, M., Efstathiou, G., & Frenk, C. S., 1987a, Nature, 330, 451
White, S., Frenk, C., Davis, M., & Efstathiou, G. 1987b, ApJ, 313, 505
White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341 (WR)
White, S. D. M., Tully, R. B., & Davis, M. 1988, ApJ, 333, L45
Wolfe, A. 1988, in QSO Absorption Lines: Probing the Universe, ed. C. Blades, D. Turnshek, & C. Norman (Cambridge: Cambridge Univ. Press), 297
———. 1989, in The Epoch of Galaxy Formation, ed. C. S. Frenk, R. S. Ellis, T. Shanks, A. F. Heavens, & J. A. Peacock (Dordrecht: Kluwer), 101
Yang, J., Turner, M. S., Steigman, G., Schramm, D. N., & Olive, K. 1984, ApJ, 281, 493