THE ASTROPHYSICAL JOURNAL, 381:14–32, 1991 November 1

TIDALLY TRIGGERED GALAXY FORMATION. I. EVOLUTION OF THE GALAXY LUMINOSITY FUNCTION

CEDRIC LACEY
Department of Astronomy, University of California, Berkeley, and Department of Physics, Oxford University, Keble Road, Oxford OX1 3RH, England

AND

JOSEPH SILK
Departments of Astronomy and Physics, and Center for Particle Astrophysics, University of California, Berkeley, CA 94720, and Mount Stromlo and Siding Spring Observatories, Australian National University, Canberra
Received 1990 May 22; accepted 1991 May 10

ABSTRACT

Motivated by accumulating evidence that large-scale galactic star formation is initiated and sustained by tidal interactions, we develop a phenomenological model for the galaxy luminosity function, commencing from a galaxy mass function that is predicted by a hierarchical model of structure formation such as the cold dark matter dominated cosmology. The epoch of luminous galaxy formation and the galactic star-formation rate are determined by the environment. Gas cooling and star-formation feedback are incorporated, and we are able to satisfactorily reproduce the present epoch luminosity function of bright galaxies and the distribution of galaxy colors. Biasing, via the preferential formation of luminous galaxies in denser regions associated with groups or clusters, is a natural outcome of our tidally triggered star-formation model. A large frequency is inferred of "failed" galaxies, prematurely stripped by supernova-driven winds, that populate groups and clusters in the form of low surface brightness gas-poor dwarfs, and of "retarded" galaxies, below the threshold for effective star formation, in the field, detectable as gas-rich, extremely low surface brightness objects. Predictions are presented for the evolution with redshift of the distribution of characteristic star formation timescales, galaxy ages, and colors. We also make estimates of galaxy number counts, and suggest that dwarf galaxies undergoing bursts of star formation at $z \sim 1$ may dominate the counts at the faintest magnitudes.

Subject headings: galaxies: formation — galaxies: photometry — galaxies: stellar content — galaxies: structure — luminosity function

1. INTRODUCTION

The luminosity function of galaxies is a fundamental observational datum that has hitherto received inadequate attention in most models of galaxy formation. This is because large-scale structure simulations yield a mass function of condensed objects, to be identified with galaxy clusters and halos. There is suprisingly good agreement between the simulations and simple analytic theory, giving some reason for confidence in the derived mass function. However, incorporation of star-formation theory is necessary in order to obtain the galaxy luminosity function from the mass function. Our goal in this paper is to provide a model for deriving the luminosity function and color distribution of galaxies, and in particular, for studying their evolution with cosmological epoch.

We shall develop a phenomenological model for the evolution of the galaxy luminosity function. The distribution of luminosities and colors of galaxies depends both on the distribution of galaxy masses and on the rate of star formation within galaxies. Thus, by studying the galaxy luminosity function, and related quantities such as galaxy number counts and galaxy colors as a function of redshift, we can hope to learn both about the initial conditions for the growth of structure in the universe, and about the processes controlling the rate of star formation on a galactic scale. Traditionally, two different approaches have been followed in this area. One approach is to calculate the dynamical evolution of collisionless matter starting at some early epoch with initial conditions specified by some theory such as the cold dark matter (CDM) model. From this, one obtains the mass function of virialized objects. The link to the luminosity function is then generally made by simply assuming a contant mass-to-light ratio. This approach is fraught with danger; indeed it fails dramatically for dwarf galaxies.

An alternative approach, followed for instance by Tinsley (1980), Bruzual & Kron (1980), and Guiderdoni & Rocca-Volnerange (1987), is to calculate the luminosity and spectral evolution of galaxies in some detail, but essentially working backwards from the present epoch. A library of stellar types is used that incorporates all important stages of stellar evolution. An initial stellar mass function and rate of star formation are prescribed to reproduce the present-day galaxy luminosity function, taken as input. Star-formation time scales are assigned to different morphological types so as to reproduce their present colors. Galaxies are assumed to all form at some redshift z_f which is taken as a free parameter. The models are then used to predict the photometric properties of the galaxy population at high redshift. This approach makes no direct contact with theoretically prescribed initial conditions, except for the rather vague concept of the "redshift of galaxy formation."

Ideally, one would like to combine the best features of these two distinct approaches to derive the luminosity and color distribution of galaxies as a function of redshift, commencing with an appropriate initial spectrum of fluctuations. Large-scale structure studies of the galaxy distribution, galaxy peculiar velocities, and searches for angular anisotropies in the cosmic microwave background are beginning to drastically

limit the freedom of choice for the primeval density fluctuation spectrum from which present-day structure evolved via the theory of gravitational instability. The part of this procedure which is the most uncertain at present is to relate the star-formation rate in a galaxy, once it has formed, to the properties of the galaxy and its surroundings. Since no predictive theory exists for this, it seems necessary to make various simple hypotheses for the dependence of the star-formation rate on physical parameters, and test these against the available data, in the hope that this will lead to a better understanding of what processes are relevant. One approach is to assume that the star-formation rate is regulated simply by the rate at which hot halo gas can cool, or by a balance between cooling and energy injection by young stars. This has been the approach taken in recent papers by White & Frenk (1991) and Cole (1991). However, while cooling of the gas is clearly a necessary condition for star formation, it may not be sufficient; other dynamical processes may control the rate of star formation within the cool gas component. Our star-formation rules will be motivated by the observation that the galaxies undergoing the most vigorous star formation today are generally found to be dynamically interacting in some way.

In this paper, we investigate a simple model in which star formation in a galaxy is induced by gravitational interactions with neighboring galaxies. The motivations for considering such a model are two: first, there is direct observational evidence for enhanced star-formation rates in interacting galaxies, from optical colors (Larson & Tinsley 1978) and ratios of far-infrared and Hα luminosities to optical luminosities (Lonsdale, Persson, & Matthews 1984; Kennicutt et al. 1987). The latter authors estimate that $\sim 10\%$ of all current star formation is occurring in moderately to strongly interacting systems of the type in their samples. Numerical simulations of tidal interactions and mergers of disk galaxies (Icke 1985; Noguchi & Ishibashi 1986; Noguchi 1988; Hernquist 1989) show that these induce shock compression of the gas, both through the direct gravitational effect of the perturbing object, and/or by inducing the formation of a central bar in the stellar disk, which itself will induce spiral shocks in the gas. Shock compression is likely to be conducive to star formation. As shown by Icke (1985) the gas in a disk galaxy, because of its low velocity dispersion, is susceptible even to fairly distant tidal encounters, which will be much more frequent than actual mergers. A second motivation for such a model is provided by the morphology-density relation for galaxies (Dressler 1980; Postman & Geller 1984), namely that the fraction of elliptical (E) and S0 galaxies relative to spirals (S's) is observed to increase with the local number density of galaxies. While this correlation might in principle result from some internal property of a protogalactic perturbation correlating with the overdensity of its surroundings in the linear regime, an alternative explanation is that it results from interactions between galaxies after they have collapsed, with interactions being more frequent in denser environments. In this picture, a moderate frequency of encounters would lead to an ordinary spiral galaxy, still forming stars today; a higher rate would lead to early conversion of all the gas into stars resulting in an S0, as suggested by Icke (1985); a still higher rate would lead to formation of an elliptical, either as a result of a merger of two disks, or perhaps as a result of close tidal encounters inducing most of the gas to form stars before it has a chance to settle into a disk. (If the gas in a protogalaxy were clumpy, it could take longer than a free-fall time to settle into a disk; while in this phase, stars might be induced to form in a more spheroidal distribution.) If a galaxy suffers no tidal interactions at all, its star-formation time scale might be extremely long, perhaps resulting in a system like Malin 1 (Impey & Bothun 1989).

Our star-formation model assumes that essentially *all* star formation in galaxies is associated in some way with galaxy interactions. This is obviously a fairly sweeping extrapolation from the observational estimate that 10% of star formation occurs in strongly interacting systems, in particular since there are many galaxies, including our own, in which star formation is occurring in spiral arms in disks, without any obvious tidal interaction with another galaxy. To reconcile this with our model, we would have to say that spiral structure is maintained only in the presence of external perturbations from other galaxies, in the absence of which it would decay away. If these perturbations are due to small or distant companions, or if the spiral structure thus induced lasts for significantly longer than the tidal interaction itself, then such galaxies would not appear in the samples of obviously interacting systems on which the observational estimate is based. There is likely to be some star formation even in completely isolated galaxies, driven by internal rather than external processes, such as swing-amplified spiral waves (Toomre 1981). However, our model implicitly assumes that this mode must constitute only a small fraction of the total. In a future paper, we will investigate the effects of allowing for star formation being driven by internal as well as external dynamics.

The specific model we choose to investigate assumes that galaxies interact at a rate proportional to the geometric cross section of the dark halo. Tidal interactions are assumed to lead to star formation; mergers are not explicitly considered in what follows. Using this assumption, we can relate the star-formation time scale in a galaxy after it has collapsed to the overdensity of a peak and of its surroundings in the linear regime. The analytical peaks formalism of Bardeen et al. (1986) is then applied to derive the statistical distribution of peak masses and star-formation time scales in terms of the power spectrum of initial functions, which we take to be that for cold dark matter. We also take account of the constraint that the gas must cool before it can form stars, and include in an approximate way the effects of feedback in the form of energy injection into the interstellar medium from supernova explosions. Details of how the models are constructed are given in § 2. Section 3 presents the results of the models for the luminosity function and color distribution of galaxies at the present epoch, and compares them with observations. Section 4 presents some predictions of the models for higher redshift galaxies. In § 5, we make a preliminary estimate of the expected number counts at faint magnitudes. Finding these to be deficient relative to observations, we propose a model for bursts in dwarf galaxies at moderate redshifts to explain the discrepancy. A detailed calculation of the number counts predicted by the model and the comparison with the observational data will be given in a forthcoming paper (Lacey et al. 1991, hereafter Paper II). Section 6 presents our conclusions.

2. MODEL EQUATIONS

We describe the basic ingredients of our models in parts §§ 2.1–2.6 of this section. We discuss some of our assumptions and compare with those made in other work in § 2.7.

2.1. *Statistics of Peaks and Environment for Linear Density Field*

In the model, galaxies are assumed to arise from peaks in the linear fluctuation field $\delta \equiv \delta\rho/\bar{\rho}$. We assume an $\Omega = 1$ cosmological model, so that in the linear regime, the fluctuation amplitude grows with time as $\delta(t) \propto a(t) \propto t^{2/3}$, $a(t)$ being the cosmological expansion factor. For definiteness, in describing the spatial dependence, we consider the fluctuation field δ extrapolated to $z = 0$ according to this linear regime growth law, and express separations in terms of their $z = 0$ comoving values. The linear fluctuation field is assumed to be Gaussian, specified by a power spectrum $P(k)$, k being the wavenumber. The field $\delta(\boldsymbol{x})$ is smoothed by convolving it with Gaussian filters of various scales R_f. We define moments σ_m of the power spectrum of the smoothed field in the usual way:

$$\sigma_m^2(R_f) \equiv \frac{1}{2\pi^2}\int_0^\infty k^{2m}P(k)\exp(-k^2R_f^2)k^2\,dk\,. \tag{2.1}$$

The zeroth moment just gives the rms value of the smoothed field δ_f: $\sigma_0(R_f) = \langle\delta_f^2\rangle^{1/2}$.

Various statistical properties of peaks in a Gaussian field have been derived by Bardeen et al. (1986, hereafter BBKS). In terms of the normalized peak height $\nu \equiv \delta/\sigma_0$, the differential number density of peaks of the field smoothed on scale R_f is

$$\frac{dn_{\rm pk}}{d\nu} = n_{\rm pk}(R_f)P(\nu;\,R_f)\,, \tag{2.2a}$$

where

$$n_{\rm pk}(R_f) = c_\infty/(2\pi)^2R_*^2 \tag{2.2b}$$

is the total number density of peaks in the field smoothed on scale R_f, and

$$P(\nu;\,R_f) = c_\infty^{-1}e^{-\nu^2/2}G(\gamma,\,\gamma\nu) \tag{2.2c}$$

is the probability distribution of peak height at fixed R_f, normalized so that $\int P(\nu)d\nu = 1$. In the above, $\gamma(R_f)$ and $R_*(R_f)$ are defined in terms of the moments

$$\begin{aligned}\gamma(R_f) &\equiv \sigma_1^2(R_f)/\sigma_0(R_f)\sigma_2(R_f)\,,\\ R_*(R_f) &\equiv \sqrt{3}\,\sigma_1(R_f)/\sigma_2(R_f)\,,\end{aligned} \tag{2.3}$$

and the function $G(\gamma, w)$ is defined by equation (A19) of BBKS. The quantity c_∞ has the numerical value 0.6397. We want to obtain an expression for the number density of peaks which is differential in the peak size as well as peak height. Derivation of a rigorous expression of this type is an unsolved problem in the peaks formalism, so we follow an approximate approach similar to that suggested by Bond (1988): we interpret equation (2.2b) as giving $n_{\rm pk}(>R_f)$, the total number density of peaks with size greater than R_f, and differentiate this to get the number density of peaks in the range $(R_f, R_f + dR_f)$:

$$\frac{dn_{\rm pk}}{dR_f} = \frac{c_\infty}{(2\pi)^2}\frac{3}{R_f^4}\left(\frac{R_f}{R_*}\right)^3\frac{d\ln R_*}{d\ln R_f}\,. \tag{2.4}$$

The number density of peaks differential in the peak size R_f and peak height ν is then taken to be

$$\frac{d^2n_{\rm pk}}{dR_f\,d\nu} = \frac{dn_{\rm pk}}{dR_f}P(\nu;\,R_f)\,, \tag{2.5}$$

Equation (2.5), with equations (2.4) and (2.2c), is equivalent to the expression for $d^2n_{\rm pk}/dR_f\,d\nu$ assumed by Bond (1988), except for the factor $d\ln R_*/d\ln R_f$ in equation (2.4). We note that while integration of equation (2.5) over ν gives back equation (2.4) for an arbitrary fluctuation field, integration over R_f gives back equation (2.2a) only if $P(k)$ is a power law [so that $\gamma(R_f)$ and $P(\nu;\,R_f)$ are independent of R_f]. This is indicative of the approximate nature of our identification between smoothing scale and peak size. However, provided that $\gamma(R_f)$ is only a weak function of smoothing scale R_f (as is the case for the cold dark matter power spectrum, for instance), the discrepancy with the rigorous version of peaks theory should not be serious. One potential problem with using equation (2.5) is the fact that some peaks are actually embedded inside larger peaks (the "cloud-in-cloud" problem). Thus, for a power-law power spectrum, $P(k) \propto k^n$, $R_* \propto R_f$, so if we assign each peak a mass $M_f \propto R_f^3$, we obtain a peak mass function $dn_{\rm pk}/dM_f \propto M_f^{-2}$, whose integral over mass diverges logarithmically if we count all peaks as being independent objects. However, various physical effects limit the range of peak masses that contribute to the luminosity function (see § 3.1), so we do not find this double counting problem to be serious in practice.

Our model for star formation depends on the properties of the surroundings of a peak as well as of the peak itself. Therefore we consider the properties of the density field smoothed on two scales, a peak scale R_p with overdensity δ_p and a background scale R_b with overdensity δ_b. (We assume $R_b > R_p$). The joint probability distribution for δ_p and δ_b, given that one is at a peak of the field δ_p, is given by equation (E5) of BBKS:

$$P(\nu_p,\,\nu_b;\,R_p,\,R_b) = \frac{1}{c_\infty[2\pi(1-\epsilon^2)]^{1/2}}\,e^{-\nu_p^2/2} \times\exp\left[-\frac{(\nu_b-\epsilon\nu_p)^2}{2(1-\epsilon^2)}\right]G(\tilde{\gamma},\,\tilde{\gamma}\tilde{\nu})\,, \tag{2.6}$$

where $\nu_p \equiv \delta_p/\sigma_{0p}$ and $\nu_b \equiv \delta_b/\sigma_{0b}$. We have also defined

$$\epsilon \equiv \frac{\sigma_{0h}^2}{\sigma_{0p}\sigma_{0b}}\,,\quad r_1 \equiv \left(\frac{\sigma_{1h}\sigma_{0p}}{\sigma_{0h}\sigma_{1p}}\right)^2\,, \tag{2.7}$$

where $\sigma_{mp} = \sigma_m(R_p)$ etc., with $R_h = [(R_p^2 + R_b^2)/2]^{1/2}$, and also

$$\tilde{\gamma} = \gamma_p\left[1 + \epsilon^2\frac{(1-r_1)^2}{(1-\epsilon^2)}\right]^{1/2}\,, \tag{2.8a}$$

$$\tilde{\nu} = \left(\frac{\gamma_p}{\tilde{\gamma}}\right)\frac{(1-r_1)}{(1-\epsilon^2)}\left[\nu_p\frac{(1-\epsilon^2r_1)}{(1-r_1)} - \epsilon\nu_b\right]. \tag{2.8b}$$

The quantity $P(\nu_p, \nu_b)$ is normalized so that

$$\iint P(\nu_p,\,\nu_b)d\nu_p\,d\nu_b = 1.$$

Our procedure for deriving the galaxy mass function and star-formation rate could in principle be applied to any model in which structures form hierarchically by gravitational instability, with small scales collapsing before large scales. The condition for hierarchical collapse is simply that the rms linear overdensity $\sigma_0(R_f)$ decrease with increasing scale R_f. For definiteness, we take the spectrum of mass fluctuations to be that predicted in the standard cold dark matter picture, with $\Omega = 1$ and $\Omega_b = 1 - \Omega_{\rm CDM} \ll 1$, b and CDM referring to the baryonic and cold dark matter components, respectively. In this case, the initial power spectrum is $P(k) = AkT^2(k)$, where for the transfer function $T(k)$ we use the numerical fit given by BBKS:

$$T(k) = \frac{\ln(1+q)}{q}\,[1 + 1.66q + (6.88q)^2 + (2.33q)^3 + (2.87q)^4]^{-1/4}\,, \tag{2.9}$$

where $q \equiv 2.34k/(\Omega h^2\ \mathrm{Mpc}^{-1})$, and $h \equiv H_0/(100\ \mathrm{km\ s}^{-1}\ \mathrm{Mpc}^{-1})$. The spectrum is normalized in terms of the parameter σ_8, the rms mass fluctuation in a sphere of radius $8h^{-1}$ Mpc. Results for the structure of individual dark matter halos and for galaxy clustering and peculiar velocities indicate that this form for the spectrum provides a reasonable fit to the observations for scales in the range $0.1 \lesssim r \lesssim 10h^{-1}\ \mathrm{Mpc}^{-1}$ when the normalization is chosen in the range $\sigma_8 = 0.4$–0.6 (Davis et al. 1985; BBKS); in terms of the "bias" parameter $b \equiv 1/\sigma_8$, this is equivalent to $2.5 \gtrsim b \gtrsim 1.7$. For most of our models we adopt $\sigma_8 = 0.5$ ($b = 2$) as an intermediate value, but also investigate the consequences of varying σ_8 over this range. There are indications that on scales $r \gtrsim 10h^{-1}$ Mpc, the power spectrum may depart from the standard CDM form, but this would not significantly affect our results, which depend on the amplitude of fluctuations on galaxy and group scales.

We will assume a Hubble constant $h = 0.5$, so that the universe has an age of $t_0 = \frac{2}{3}H_0^{-1} = 13$ Gyr. Constraints on the baryon density from the standard model of primordial nucleosynthesis imply $0.03 \lesssim \Omega_b \lesssim 0.10$ for this value of h (Kawano, Schramm, & Steigman 1988), so we will assume a baryonic fraction $f_b = \Omega_b/\Omega = 0.07$, the remainder of the mass being the collisionless cold dark matter.

2.2. *Nonlinear Collapse Model*

Peaks in the initial fluctuation field having $\delta > 0$ are assumed to collapse to bound objects once they have grown to sufficient amplitude, while peaks having $\delta < 0$ (due to being in larger scale underdense regions) are assumed never to collapse. The collapse and virialization lead to formation of a dark halo, within which a baryonic core may condense if cooling allows. We relate the collapse time and postvirialization radius and density of the dark halo to the initial overdensity and size of the corresponding peak in the linear regime using the spherical "top hat" collapse model (e.g., Peebles 1980), which is the exact solution for the nonlinear evolution of a uniform density, spherical, overdense region in the case that the matter exerts no pressure. In this idealized model, the overdense sphere recollapses to a point. We take the formation time t_f of the collapsed object to be the time at which this occurs. For an $\Omega = 1$ cosmology, this is given by

$$t_f = (\pi/H_0)(3/5\delta)^{3/2}\ , \tag{2.10}$$

and corresponds to a redshift z_f given by

$$1 + z_f = \delta/[\tfrac{3}{20}(12\pi)^{2/3}] \equiv \delta/\delta_{\mathrm{cr}}\ , \tag{2.11}$$

where δ_{cr} has the numerical value $\delta_{\mathrm{cr}} = 1.69$. In the above, δ is the overdensity of the top-hat perturbation, extrapolated to $z = 0$ as though it were still in the linear regime. We assume throughout that only the growing mode is present in the linear perturbation field. For simplicity, we assume that the structure of the dark halo after collapse and virialization is a uniform density sphere of radius r_f. By the virial theorem, r_f is equal to half of the radius of the top-hat perturbation at its time of maximum expansion, and so is given by

$$r_f/R = (3/10\delta)\ , \tag{2.12}$$

where R is the $z = 0$ comoving radius of the perturbation. The present mean density in an $\Omega = 1$ universe is $\bar{\rho} = 3H_0^2/8\pi G$, so the final density of the halo is

$$\rho_f = \bar{\rho}(R/r_f)^3 = (3H_0^2/8\pi G)(10\delta/3)^3\ . \tag{2.13}$$

The dynamical time, defined as the time for a halo particle to oscillate once in radius, is

$$t_d = (3\pi/4G\rho_f)^{1/2} = \tfrac{1}{2}t_f\ , \tag{2.14}$$

while the three-dimensional velocity dispersion is given by the virial thorem as

$$v_d = (3GM/5r_f)^{1/2} = H_0 R\,\delta^{1/2}\ , \tag{2.15}$$

M being the total mass. To relate the parameters (δ, R) in the top-hat collapse model to (δ_f, R_f) in the Gaussian peaks theory, we take the mass to be that under a Gaussian window, $M = (2\pi)^{3/2}R_f^3\bar{\rho}$, so that the top-hat radius R corresponding to Gaussian filter radius R_f is $R = (3/4\pi)^{1/3}(2\pi)^{1/2}R_f$, and take the top-hat perturbation amplitude to be the same as the Gaussian-filtered amplitude, $\delta = \delta_f$.

2.3. *Group Collapse and Cooling Condition*

We assume that the gas is shock-heated to the virial temperature of the halo when the galaxy collapses, and that before it can form stars it must cool and condense into a self-gravitating baryonic core at the center of the halo (White & Rees 1978). To calculate the cooling time, we assume that the gas is initially distributed in the same way as the dark matter halo, which is modelled as a uniform density sphere. Then the initial temperature of the gas is

$$T_{\mathrm{vir}} = m_H v_d^2/3Y_T k\ , \tag{2.16}$$

and the cooling time is

$$t_{\mathrm{cool}} = \frac{(3/2)n_T kT_{\mathrm{vir}}}{n_H^2 L(T_{\mathrm{vir}})} = \frac{3Y_T}{2Y_H^2}\frac{kT_{\mathrm{vir}}}{n_b L(T_{\mathrm{vir}})}\ , \tag{2.17}$$

where $n_b = f_b\rho_f/m_H$ is the baryon number density in the halo, and $L(T)$ is the cooling function. We have defined dimensionless factors Y_T and Y_H such that $n_T \equiv Y_T n_b$ is the total particle density, and $n_H \equiv Y_H n_b$ is the number density of hydrogen atoms (ionized + neutral). For most of the models calculated, we assume that the gas has primordial composition, with 76% H, 24% He, so that $Y_T = 1.70$ and $Y_H = 0.76$, and use the Fall & Rees (1985) cooling function, which includes radiation by ionized and neutral hydrogen and helium, but no molecules. If the gas were already contaminated by metals produced by an earlier generation of stars, for instance in dwarf galaxies which underwent mass loss, then the cooling rate would be larger than we have asumed here. We have calculated a few models including cooling by metals. For these, we have used the Gaetz & Salpeter (1983) cooling function for solar metallicity, and interpolated for intermediate heavy element abundances. Treating the halo as a uniform sphere is also obviously a simplification; in reality one expects it to have a density profile roughly of the form $\rho \propto r^{-2}$, so that the inner regions will cool before the outer regions. Our treatment is intended to give the effective cooling time for the halo as a whole.

Each dark matter halo forms part of a clustering hierarchy. If the next scale in the hierarchy (denoted "gr" for "group") collapses before the gas in the halo corresponding to the previous scale (denoted "g" for "galaxy") has had time to cool, we assume that the gas will be shock-heated again and dispersed through the dark halo of the new larger system. Thus the gas can only condense into a baryonic core on scale g if it satisfies the condition (compare to White & Rees 1978)

$$t_{\mathrm{cool}} < t_{f,\mathrm{gr}} - t_{f,\mathrm{g}}\ . \tag{2.18}$$

The formation times $t_{f,g}$ and $t_{f,gr}$ are calculated from the linear overdensities δ_g and δ_{gr} on the galaxy and group scales R_g and R_{gr} using the top-hat collapse model already described. We calculate the joint probability distribution of δ_g and δ_{gr} using the formalism set out in § 2.1, identifying the galaxy scale with the peak (p) and the group scale with the background (b). Note that objects satisfying the cooling condition (2.18) automatically satisfy the weaker condition $\delta_g > \delta_{gr}$ that the peak be higher than its surroundings. It remains to choose the group scale R_{gr} for a given galaxy scale R_g. We make the simplification of assuming a fixed ratio of scales R_{gr}/R_g in the clustering hierarchy (although in reality there will be some dispersion in this ratio, with different galaxies being embedded in groups of different sizes). Specifically, we shall assume $R_{gr}/R_g = 2$, corresponding to a ratio of mass scales $M_{gr}/M_g = 8$. This is fairly consistent with the results of Cole (1989), who found from analysing N-body simulations with CDM initial conditions that halos typically lose their identity after being incorporated into systems ~ 6 times more massive.

Thus, each halo will condense a baryonic core, if the gas can cool fast enough. Once formed, these baryonic cores are assumed to survive subsequent merging of the dark halos without themselves being merged. This "no merger" assumption for the cores can only be an approximation, since some merging of the cores undoubtedly occurs. However, the rate of merging of present day luminous galaxies appears to be small, based on observational estimates (Toomre 1977). Some theoretical support for the no merger approximation comes from the N-body simulations by Carlberg & Couchman (1989), which include the effects of dissipational gas dynamics in a crude way, and find little merging of cores at recent epochs. They also find that the mass distribution of cores is well described by the peaks theory.

2.4. *Model for Star-Formation Rate*

Given that the gas has condensed into a cool self-gravitating core, one needs to know at what rate it will be converted into stars. Our model for the star-formation rate is based on the idea that star formation is induced by tidal interactions between galaxies in a group. The duration of a star formation event may, however, exceed the duration of the galaxy encounter which triggers it, if, for instance, the encounter induces a bar or spiral density wave. Star formation in galaxies when they are in their quiescent state between encounters is assumed to constitute a negligible fraction of the total. As applied to the Milky Way galaxy, this model would require that most of the star formation in the last few Gyr has been due to interactions with the Large Magellanic Cloud, which is the nearest galaxy having significant mass.

The statistics of group and galaxy properties are calculated in the way described in the previous section. Since star formation is assumed to be driven by galaxy interactions, we assume that star formation in a galaxy turns on at the time $t_* = t_{f,gr}$ at which the surrounding group collapses, corresponding to a turn-on redshift

$$1 + z_* = \delta_{gr}/\delta_{cr} . \tag{2.19}$$

Subsequent to this, formation of new stars is assumed to proceed at a rate Ψ proportional to the instantaneous gas mass $M_{gas,g}$:

$$\Psi = M_{gas,g}/\tau_* . \tag{2.20}$$

The initial gas mass is $M_{gas,g} = M_{b,g} = f_b M_g$. If the recycling of gas from dying stars were neglected, this would give a star-formation rate $\Psi(t) = M_{b,g} \exp[-(t - t_*)/\tau_*]$, for star formation beginning at time t_*. In fact, the evolutionary models we use include gas recycling, so the time-dependence is slightly more complicated than this.

We assume that the time scale τ_* is related to the galaxy-galaxy collision time calculated according to the following simple model: Suppose that we have $N_{gr} = M_{gr}/M_g$ identical galaxies of radii r_g in a group of radius r_{gr}, where the radii all refer to the collisionless dark matter, and are postvirialization values. Assigning each galaxy its geometrical cross-section πr_g^2, the collision probability per group crossing time $t_{d,gr}$ would be $N_{gr}(r_g/r_{gr})^2$. We then make the assumption that the time scale for inducing star formation is simply $1/A_*$ times the collision time calculated in this way, where A_* is treated as an adjustable parameter. The effects of having a spectrum of masses for neighboring galaxies, and of the efficiency of converting gas to stars being less than 100% in a single encounter, are absorbed into the parameter A_*. (These effects are discussed in the Appendix, where we attempt an estimate of A_* from first principles.) Thus we obtain

$$\begin{aligned}\tau_* &= \frac{1}{A_*}\frac{t_{d,gr}}{N_{gr}}\left(\frac{r_{gr}}{r_g}\right)^2 \\ &= \left(\frac{\pi}{2}\right)\left(\frac{3}{5}\right)^{3/2}\frac{1}{H_0 A_*}\left(\frac{M_g}{M_{gr}}\right)^{1/3}\left(\frac{\delta_g}{\delta_{gr}}\right)^2\frac{1}{\delta_{gr}^{3/2}} ,\end{aligned} \tag{2.21}$$

where in the second line we have substituted from equations (2.14), (2.10), and (2.12).

We emphasize that we are assuming that galaxies interact, but that the luminous cores do not merge, so that the number of luminous galaxies is conserved, apart from the formation of new galaxies. In reality, the star formation would occur in a series of bursts induced by tidal encounters, with equation (2.20) describing the average behavior, but we ignore this complication in the present models. Another effect we ignore is that the group environment of a galaxy may evolve with time, as the next scale in the hierarchy above the original group scale collapses. This would only affect the star-formation rate long after the galaxy formed.

For the purpose of considering spectral evolution of galaxies, it is convenient to take the fundamental variables to be (M_b, τ_*, z_*). The differential number density of galaxies in these variables is then

$$\begin{aligned}\frac{d^3n}{dM_b\,d\tau_*\,dz_*} &= \left(\frac{d^3n_{pk}}{dR_{f,g}\,dv_g\,dv_{gr}}\right)\left[\frac{\partial(M_b, \tau_*, z_*)}{\partial(R_{f,g}, v_g, v_{gr})}\right]^{-1} \\ &= \left(\frac{dn_{pk}}{dR_{f,g}}\right)\left(\frac{R_{f,g}}{3f_b M_g}\right) \\ &\quad \times \frac{P(v_g, v_{gr}; R_{f,g}, R_{f,gr})}{(\partial\tau_*/\partial\delta_g)(dz_*/d\delta_{gr})\sigma_0(R_{f,g})\sigma_0(R_{f,gr})} .\end{aligned} \tag{2.22}$$

We note that a consequence of our assumption that star formation does not begin in a baryonic core until the surrounding group has collapsed is that there is a significant fraction of "retarded" galaxies, consisting of cool gas, in which star formation is assumed either not to occur at all, or to occur at such a low rate that these galaxies are essentially invisible to direct observation by means of stellar emission. The gas they contain might, however, be visible in absorption against a bright background source, producing some of the observed

quasar absorption lines. Indeed, such failed galaxies might help to account for the number of damped Lyα absorption systems seen at $z \approx 2$—these systems have properties similar to the gas disks in ordinary spiral galaxies, but are overabundant by a factor ~ 5 compared to what one infers from the present-day galaxy luminosity function, assuming no evolution in the numbers or sizes of galaxies (Wolfe 1988). The prediction of large numbers of dark or underluminous galaxies arises in almost any model of hierarchical formation of structure in which one tries to reconcil the observed galaxy luminosity function with having $\Omega = 1$. For instance, in their interpretation of their N-body simulations of the CDM scenario, Frenk et al. (1988) assign luminosities to dark matter halos in a way that guarantees that they reproduce the observed galaxy luminosity function. They then find that most of the mass in the universe is in galaxies having relatively low circular velocities and total M/L ratios more than 10 times those in "normal" galaxies. Our model differs from this in that we predict galaxies of cool gas that are completely dark, that are not exclusively of low circular velocity, and are not related to present day halos in a one-to-one way.

2.5. *Feedback from Star Formation*

There is feedback from star formation, in the form of supernovae, stellar winds, and H II regions, produced by massive young stars. This energy input heats the gas and tends to limit the star-formation rate. It may drive a wind out of the galaxy, and in the case that the escape velocity is sufficiently low, this wind may cause the ejection of most of the gas before it can form into stars (Larson 1974). We model this feedback in a very simple way, in the form of a halo binding energy criterion (the escape velocity from a galaxy being proportional to the halo velocity dispersion): we assume that there is a critical halo velocity dispersion $v_{\rm crit}$, such that for $v_d > v_{\rm crit}$, no mass loss occurs, and star formation is unaffected by feedback, while for $v_d < v_{\rm crit}$, almost complete mass loss occurs. Dekel & Silk (1986) estimate that the critical velocity dispersion below which catastrophic mass loss occurs is $v_{\rm crit} = 100$ km s^{-1}, and we use this estimate in our models. In reality, the (low mass) galaxies with $v_d < v_{\rm crit}$ will form some stars before ejecting their gas, presumably resulting in the low surface brightness dwarfs, but we will postpone a proper treatment of star formation in these galaxies to Paper II, and in this paper only consider the contribution to the luminosity function from galaxies with $v_d > v_{\rm crit}$, identified with galaxies of "normal" surface brightness.

2.6. *Spectral Evolution*

An important innovation in our modelling of galaxy formation, compared to most earlier work, is that we utilize a realistic galactic spectral evolution code in order to be able to express our model results in terms that will be of direct relevance to observers. Given the star-formation rate in a galaxy (specified in terms of M_b, τ_*, z_*), we calculate the evolution of the integrated spectrum of the galaxy, and then derive the apparent magnitude of the galaxy in various bands (e.g., B, V, R) as a function of redshift by integrating over the appropriately redshifted spectrum weighted by the response functions of the filters. The spectra are calculated using a recent (1989) version of Bruzual's program GALAXEV, which is a somewhat updated version of the program described by Bruzual (1981, 1983), and the filter response functions used are those given by Bruzual. These models include the contribution to the integrated light from stars on the main sequence and red giant branch, but not from the horizontal branch or later stages of stellar evolution, and assume that all stars form with solar metallicity. The models do not include emission from ionized gas in H II regions, or reddening of the spectrum by dust within the galaxy. Our prescription for the time-dependence of the star-formation rate, equation (2.20), is equivalent to that for Bruzual's μ-models, where $\mu = 1 - \exp(-1\ {\rm Gyr}/\tau_*)$. For simplicity, the initial mass function (IMF) of the stars, $\phi(m)$ [where $\phi(m)dm$ = number of stars in mass range $(m, m + dm)$ per unit total mass of stars formed] is taken to be the same as that observed in the solar neighborhood; as discussed by Scalo (1986), there is no compelling observational evidence that the large-scale IMF varies. We adopt a piecewise power-law approximation [$\phi(m) \propto m^{-(1+x)}$ in each segment] to the solar neighborhood IMF inferred by Scalo (1986): a fraction $\xi = 0.5$ of the mass in the IMF is assumed to go into luminous stars, with power-law indices $x = 0.25$ for $0.1 < m/M_\odot < 1$, $x = 1.35$ for $1 < m/M_\odot < 2$ and $x = 1.7$ for $2 < m/M_\odot < 80$, the remainder going into some nonluminous form (for the purpose of these models it does not matter what), such as brown dwarfs. (This is identical to the IMF used by Guiderdoni & Rocca-Volmerange 1987). The fraction $1 - \xi$ of nonluminous mass has been chosen to give reasonable values of M/L_B for the stellar populations in present-day galaxies.

2.7. *Discussion of Assumptions*

It is worthwhile at this point to compare the assumptions we make in our model with those made in previous work on the galaxy luminosity function. One of the earliest attempts to derive the galaxy luminosity function from a model for galaxy formation and clustering was that of White & Rees (1978). Many of the ingredients of the White & Rees model have been incorporated also into more recent models, including our own. These include: condensation of baryonic cores in halos which collapse at various redshifts; shock heating and radiative cooling of gas in galaxy halos; the assumption that the baryonic cores, once formed, do not merge; and including a type of feedback to limit star formation. Our model however differs from White & Rees (1978) in certain key assumptions, and in many of the details. For instance, their cooling condition assumed a fixed ratio of time scales between the collapse of different levels of the clustering hierarchy (in our notation, they took $t_{f,\rm gr} = 2t_{f,\rm g}$). They modeled feedback differently. Stellar populations were assigned a constant mass-to-light ratio, with no account taken of stellar evolution. Most importantly, White & Rees assumed that once a baryonic core formed by cooling of halo gas, star formation followed immediately, converting all of the gas into stars, except for that fraction ejected due to feedback.

A recent model which is closer to White & Rees (1978) in its assumptions is that of White & Frenk (1991) (an earlier version of which was described in White 1989). The White & Frenk (1991) model assumes that the star formation rate is controlled by the rate of cooling of gas in an inhomogeneous halo, modified by feedback. It also includes luminosity evolution of stellar populations in a very approximate way, based on fits to Bruzual models, but makes no attempt to calculate galaxy colours. Other work on relating the luminosity function to the model of structure formation has been much more schematic in the procedures used for relating luminosity to mass; this has usually been done by simply assuming a constant M/L either for all presently existing galaxy halos, or for all baryonic cores formed by cooling. Recent papers along these lines include

Schaeffer & Silk (1988), Cole & Kaiser (1989), and Peacock & Heavens (1990). The most detailed treatment of the cooling of baryons within the population of evolving dark matter halos has been that of Cole (1991).

Most of the previous work (including that of White & Rees 1978 and White & Frenk 1991) has been based on using the Press-Schechter formalism (Press & Schechter 1974; Bond et al. 1990), rather than the peaks formalism, for deriving the mass function of condensed objects from the power spectrum of fluctuations. The Press-Schechter formalism has some advantages over the peaks formalism for describing the mass function of galaxy halos at a fixed time: it automatically includes the effect of merging of halos at earlier times, and its integral over mass is equal to the mean density in the universe. In contrast, if one uses our modification of the peaks formalism to calculate the mass function of collapsed peaks, one ends up counting some peaks which are actually inside larger collapsed peaks, and so would not form independent halos. There appears to be no simple method of correcting for this effect. As a consequence, the total mass in collapsed objects estimated in this way may diverge. In practice, this divergence, which occurs at the low mass end, is only logarithmic, and so not too serious. Comparing the present-day mass functions predicted for a CDM spectrum by the Press-Schechter and peaks methods, one finds that they are rather similar, but with the peaks theory giving an exponential cutoff at the high mass end that is shifted down in mass by about a factor of 2 compared to Press-Schechter, and giving a slightly steeper slope at the low-mass end (due to counting peaks which would in reality have merged into larger halos). Comparison with N-body results for the halo mass function from CDM simulations (Frenk et al. 1988; Efstathiou 1991, private communication) indicates that Press-Schechter gives a better fit to the high mass cutoff, while our peaks formalism gives a better fit at intermediate masses. Present N-body simulations do not greatly constrain the form of the mass function at low masses because of their limited dynamic range, but of course it must eventually fall below the peaks prediction.

To calculate the luminosity function, one needs to know more than just the present halo mass function, since the luminosity of a galaxy depends on when it began forming stars, and at what rate. Furthermore, there may not even be a one-to-one correspondence between dark halos and luminous galaxies, if the baryonic cores avoid merging when the halos merge. These are the reasons why in this paper we work with the peaks formalism, which allows us to attach a formation time to each object, and furthermore allows us to deal with the statistics of environmental effects in a crude way, which is necessary given our model for star formation. Use of the peaks formalism is obviously well matched to the assumption of no merging of the baryonic cores, if we suppose that each positive overdensity peak will at some stage collapse to form a dark halo within which a baryonic core can condense if other conditions (cooling, feedback) allow. White & Rees (1978) and White & Frenk (1991) follow a different approach in calculating the galaxy luminosity function. They assume that dark halos of a given mass form at a rate given by dividing the Press-Schechter mass function at that epoch by the value of the cosmic time, and further that each halo then lasts for a length of time equal to its formation time before merging with another halo. During that time, each halo may condense out a baryonic core, which is then assumed to survive subsequent halo merging. Whatever the merits of the original Press-Schechter method for deriving a halo mass function, the White & Rees (1978) extension of it lacks rigorous justification. In effect, White & Rees (1978) assume that different levels in the clustering hierarchy are separated by a constant factor in *formation time*, while we assume that they are separated by a constant factor in *mass* (to relate group and galaxy scales). Furthermore, the White & Rees (1978) procedure can in principle lead to a divergent total mass in baryonic cores, just as ours can, if other physical effects do not intervene.

3. RESULTS OF MODELS FOR $z = 0$

Our model for global star formation predicts the present-day photometric properties of galaxies. In this section, we compare these predictions with observational data on the luminosity and color distributions of galaxies. For our "standard" model, we adopt values for the amplitude of the fluctuation spectrum $\sigma_8 = 0.5$ (i.e., $b = 2$) and for the collision cross-section parameter $A_* = 5$; we assume a baryon fraction $f_b = 0.07$, and include both the cooling criterion with zero metals, and the halo binding energy criterion with $v_{\rm crit} = 100$ km s^{-1}. The value of A_* is chosen to give a reasonable fit to the color distribution, as discussed below.

3.1. *Galaxy Mass Function*

In Figure 1 we show the distribution of baryon masses M_b for peaks at $z = 0$ in our standard model, for various conditions on the peaks. The mass function for all peaks is very close to a power law $dn/dM \propto M^{-2}$, and diverges in its contribution to the total density at both low- and high-mass ends. The mass function for peaks which have collapsed by the present day cuts off at $M_b \gtrsim 10^{13}\ M_\odot$ as the rms amplitude $\sigma_0(M)$ falls below unity. The mass function for collapsed peaks lies below the mass function for all peaks even at very low masses where $\sigma_0(M) \gg 1$, because some fraction of peaks have $\delta < 0$ due to

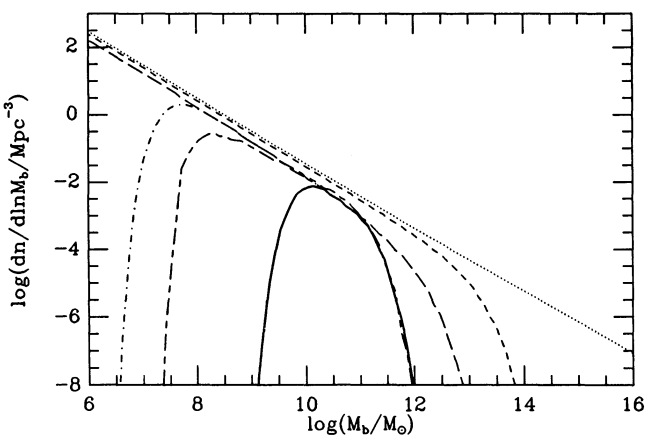

FIG. 1.—Number density of peaks as a function of baryonic mass M_b, at redshift $z = 0$, for CDM spectrum normalized to $\sigma_8 = 0.5$. *Dotted line*: mass distribution for all peaks. *Short-dashed line:* mass function for peaks which have collapsed. *Long-dashed line:* mass function for peaks which have collapsed within groups that have also collapsed (for a mass ratio $M_{\rm gr}/M_g = 8$). *Dotted-short-dashed line:* collapsed peaks within collapsed groups for which the cooling condition is satisfied; these objects would have initiated star formation. *Short-dashed-long-dashed line:* collapsed peaks within which gas has cooled, but for which surrounding group has not yet collapsed; these objects would not yet have initiated star formation. *Solid line:* collapsed peaks within collapsed groups, which satisfy the cooling condition and the condition ($v_d > v_{\rm crit} = 100$ km s^{-1}) to avoid significant mass loss; these objects form normal surface brightness luminous galaxies by the present epoch.

lying in larger scale underdense regions, and so never collapse according to our criterion $\delta > \delta_{cr}$. The mass function for peaks within groups which have collapsed on a mass scale larger by a factor $M_{gr}/M_g = 8$ is further suppressed relative to that for all peaks, since $\sigma_0(M)$ is smaller on the group scale. Imposing the cooling condition $t_{cool} < t_{f,gr} - t_{f,g}$ on peaks within groups which have collapsed leads to the mass function for peaks which can have initiated star formation according to our model; these are the luminous galaxies. This mass function cuts off below $M_b \approx 3 \times 10^9\ M_\odot$. Over the mass range $10^{10} \lesssim$ stood in terms of the cooling argument of Rees & Ostriker (1977) and Silk (1977), and below a mass $M_b \approx 1 \times 10^7\ M_\odot$, because our cooling function assumes that cooling is negligible for $T \lesssim 10^4$ K, and very low-mass objects have virial temperatures below this. Imposing the further condition $v_d > v_{crit}$ on the halo velocity dispersion gives the mass function for peaks which initiate star formation and survive mass loss driven by star-formation feedback effects; these are "normal" surface brightness galaxies in our model. This mass function cuts off below $M_b \approx 3 \times 10^9\ M_\odot$. Over the mass range $10^{10} \lesssim M_b/M_\odot \lesssim 10^{11}$, the number density of star forming galaxies is suppressed relative to that of peaks of the same mass by about a factor of 4, due to the group collapse condition. Thus we have a "natural biasing" of the galaxies relative to the mass. At lower masses, most of the galaxies are of the "low" surface brightness type, which lose most of their baryons after initiating star formation. Finally, we show the mass function for peaks which have collapsed and in which the gas has cooled by the present day ($t_{cool} < t_0 - t_{f,g}$), but in which the surrounding group has not yet collapsed; in our model, these "dark" galaxies have not yet begun to form stars, but might be visible in absorption or H I emission. The number density at a given mass of these dark galaxies is very similar to that of galaxies which have initiated star formation, except that the low-mass cutoff due to cooling is shifted to somewhat higher masses, because imposing the constraint that the overdensity is low on the "group" scale biases one towards lower density peaks on the "galaxy" scale, for which the virial temperature is lower. We note that, since the peaks corresponding to "luminous" galaxies on average have larger overdensities than those corresponding to "dark" galaxies, they will also be more strongly clustered (BBKS).

It is of interest to compute the mass fractions (including the associated dark matter) in objects of various types. The mass fraction in collapsed peaks diverges if integrated down to arbitrarily low masses; it reaches 50% if integrated down to $M_{1/2} \approx 1.9 \times 10^{10}\ M_\odot$, and 100% if integrated down to $M_1 \approx 1.5 \times 10^6\ M_\odot$. In a hierarchical clustering picture, all the mass should be in collapsed objects on some scale, so for consistency one should truncate the mass function below the scale M_1. The mass fraction in collapsed peaks satisfying the group collapse and cooling conditions for initiating star formation is 25%, of which 6.8% is in "normal" surface brightness galaxies and 19% in "low" surface brightness galaxies. The mass fraction of "dark" galaxies, containing cool gas but no stars, is 18%.

3.2. *Distribution of Star-Formation Parameters* (z_*, τ_*)

At a given mass, the star-formation turn-on redshift z_* and the star-formation time scale τ_* for peaks satisfying the conditions for star formation both cover a large range. This is illustrated in Figure 2, which shows a contour plot of $dn_{gal}/dz_*\, d\tau_*$ at fixed mass $M_b = 3 \times 10^{10}\ M_\odot$. The contours show the number density of objects one obtains directly from equation

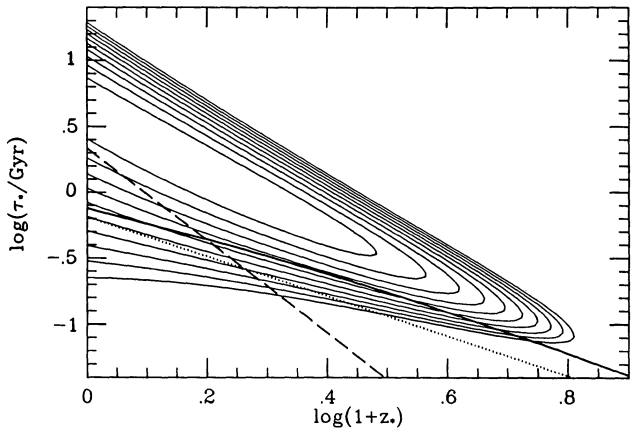

FIG. 2.—Contour plot of $dn_{pk}/d\ln M_b/d\ln(1+z_*)/d\ln\tau_*$, at fixed mass $M_b = 3 \times 10^{10}\ M_\odot$. The contours are at equal logarithmic intervals, at factors 4, 16, 64 ... down from the peak value. Also plotted are the curves representing the following criteria: *Dotted curve:* the condition $\delta_g > \delta_{gr}$. *Solid curve:* the cooling condition $t_{cool} < t_{f,gr} - t_{f,g}$. *Dashed line:* the binding energy condition $v_d > v_{crit}$. Normal surface brightness luminous galaxies correspond to the region of the contour diagram above all three curves.

(2.22), without any additional criteria. We also plot curves corresponding to the criteria: (1) $\delta_g > \delta_{gr}$, (2) $t_{cool} < t_{f,gr} - t_{f,g}$ and (3) $v_d > v_{crit}$; in each case, the region below the line is excluded, so that only the region above all three curves contributes to the luminosity function of "normal" surface brightness galaxies. This plot illustrates the strong correlation between τ_* and z_* at a given mass, with the galaxies that turn on star formation earlier also tending to have shorter star-formation time scales.

In Figure 3, we show the distribution of turn-on redshifts z_* (Fig. 3*a*) and star-formation time scales τ_* (Fig. 3*b*) for presently luminous normal surface brightness galaxies, both for all galaxies taken together, and separately for the baryonic mass ranges 10^9–$10^{10}\ M_\odot$, 10^{10}–$10^{11}\ M_\odot$ and 10^{11}–$10^{12}\ M_\odot$. It can be seen that low-mass galaxies on average have larger z_* and smaller τ_* compared to high mass galaxies. Indeed, for $M_b = 10^9$–$10^{10}\ M_\odot$, the turn-on rate peaks at $z_* \approx 2$, while for $M_b = 10^{11}$–$10^{12}\ M_\odot$, it increases with time up to the present epoch. The most notable feature about the distribution of τ_* is that it is very broad, covering a range ~ 10–10^2 even at a fixed mass. The distribution of z_* is sensitive to the amplitude of the perturbation spectrum, falling off much more steeply with increasing z_* for decreasing σ_8. Decreasing σ_8 also has the effect of broadening the τ_* distribution and shifting it to larger values, but this effect is modest in comparison to the overall width of the distribution, for values in the range $\sigma_8 = 0.4$–0.6.

3.3. *Galaxy Luminosity Function*

Predictions of the models for the distribution of galaxies in absolute B-magnitude $\mathcal{M}_B$ are shown in Figure 4. Figure 4*a* shows the luminosity function for the standard model parameters ($A_* = 5$ and $\sigma_* = 0.5$), both with and without the cooling and binding energy cutoffs. These cutoffs respectively result in high-end and low-end cutoffs in the luminosity distribution. Figure 4*a* also shows the effect of varying A_* over the range $A_* = 2$–10: this mainly affects the luminosity function at the bright end, the number of bright galaxies increasing with A_*, but leaves the number of faint galaxies unchanged. Figure 4*b* shows the effect on the luminosity function of varying σ_8: the number density of galaxies increases with increasing σ_8, this

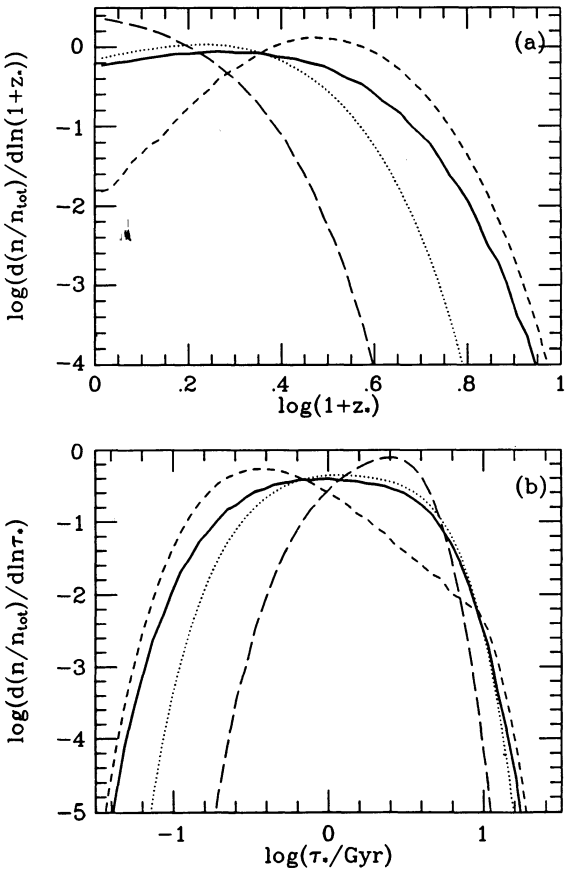

FIG. 3.—Distribution of (*a*) turn-on redshifts z_* and (*b*) star-formation time scales τ_* for normal surface brightness luminous galaxies at $z = 0$ in our standard model. Each of the distributions plotted is normalized to unit total area. *Solid curve:* all galaxies. *Short-dashed, dotted, and long-dashed curves:* galaxies with $10^9 < M_b < 10^{10}\ M_\odot$, $10^{10} < M_b < 10^{11}\ M_\odot$, and $10^{11} < M_b < 10^{12}\ M_\odot$, respectively.

effect being largest in the region of the bright and faint end cutoffs. Figure 4*b* also shows the luminosity function one would derive from the mass function for the standard model if one assumed a *constant* baryonic mass-to-light ratio $M_b/L_B = 5$ in solar units. It can be seen that the actual luminosity function has the same shape as this at the low- and high-mass ends, but is somewhat broader, because the mass-to-light ratio varies, from $M_b/L_B \approx 10$ at low masses to $M_b/L_B \approx 3$ at high masses (although there is a spread in M_b/L_B at any given mass).

In Figure 4*c* we show the effect of varying the baryon fraction f_b in the range $f_b = 0.03$–0.14. The value $f_b = 0.14$ is actually somewhat outside the range favored by primordial nucleosynthesis. The luminosity function is fairly sensitive to f_b. At the bright end, the location of the cutoff due to cooling varies approximately as $L_B \propto M_b \propto f_b^2$, while at lower masses, the luminosity at a given halo mass scales simply as $L_B \propto M_b \propto f_b$. In Figure 4*d*, we show the effect of assuming that the gas from which the galaxies form is already enriched in heavy elements, which affects the cooling rate. As one would expect, this only affects the luminosity function at the bright end, where increasing the metallicity allows somewhat more massive galaxies to cool. We show results for initial metallicities relative to solar $Z/Z_\odot = 0$ (our standard case), 0.1 and 1.0. In the case of our own galaxy, the initial metallicity must have been very low, since one finds stars with metallicities as small as $Z \sim 0.1\ Z_\odot$ in the disk and $Z \sim 10^{-3}\ Z_\odot$ in the spheroid. However, in some environments, for instance galaxy clusters, pre-enrichment might be significant, if dwarf galaxies form first and then eject some of the metals they produce, polluting the gas from which more massive galaxies later form. A pre-enrichment to $Z \sim 0.1\ Z_\odot$ might plausibly arise this way, but values $Z \sim Z_\odot$ seem unlikely. Modeling this effect properly would have complicated the models, and so it has not been included.

The observed galaxy luminosity function can be estimated from redshift surveys of "field" galaxies, which provide a large-scale average, or for individual clusters, in which case only the shape but not the normalization is significant. The field galaxy luminosity function is conventionally fit to a Schechter function, $dn/dL = (\phi_*/L_*)(L/L_*)^\alpha \exp(-L/L_*)$. The most recent determination, by Efstathiou, Ellis, & Peterson (1988), finds that a Schechter function with $\alpha = -1.1$, $\mathcal{M}_B^* = -21.2$, and $\phi_* = 2.0 \times 10^{-3}$ Mpc^{-3} (for $H_0 = 50$ km s^{-1} Mpc^{-1}) gives a good fit over the entire range $-18 \lesssim \mathcal{M}_B \lesssim -23$ for which they have data. Previous determinations, summarized by Binggeli, Sandage, & Tamman (1988) gave similar values for these parameters, with $\mathcal{M}_B^* = -20.7$ to -21.5 and $\alpha = -1.0$ to -1.25. The Efstathiou et al. (1988) luminosity function is plotted in Figures 4*a*–4*d*. Comparing the observed distribution with the model predictions, we see that the predicted luminosity function agrees in amplitude with that observed around the "knee" of the Schechter function at $\mathcal{M}_B \approx \mathcal{M}_B^* \approx -21$, this conclusion being insensitive to the values of σ_8 or A_*, or even to the presence of the cooling and binding energy cutoffs. Without the cooling cutoff, the luminosity function at the bright end ($\mathcal{M}_B \lesssim -21$) falls off far too slowly, while including cooling gives a falloff similar to what is observed. Without the binding energy condition, the predicted luminosity function at the faint end rises much too rapidly, roughly as $dn/dL \propto L^{-2}$, equivalent to $\alpha \approx -2$. Including the binding energy condition results in a flattening of the luminosity function at intermediate luminosities ($-21 \lesssim \mathcal{M}_B \lesssim -17$), in better agreement with the observations. Note that we could have improved the fit to the bright end of the luminosity function by choosing a somewhat larger value for f_b ($f_b \approx 0.1$ instead of $f_b = 0.07$), but that changing f_b by a larger factor would have worsened the fit.

At the faint end of the luminosity function, the main observational evidence comes from the study of galaxy clusters. For the Virgo cluster, Sandage, Binggeli, & Tamman (1985) determined the luminosity function of galaxies separated according to morphological type down to $\mathcal{M}_B \approx -12$. They find that ordinary high surface brightness galaxies (E's, S0's and S's) all have luminosity functions that cut off at $\mathcal{M}_B \gtrsim -16$, while low surface brightness galaxies (dE's and Irr's) only exist at $\mathcal{M}_B \gtrsim -18$. General confirmation of this picture comes from studies of other clusters and of local field galaxies (Binggeli, Sandage, & Tamman 1988; Binggeli, Tarenghi, & Sandage 1990), although the luminosity function in the field may not be identical to that in clusters. Confirmation that dwarf ellipticals form a disjoint family from normal ellipticals comes from the work of Kormendy (1985) on the central surface brightness-core radius-luminosity relation. Our models with $v_{\rm crit} = 100$ km s^{-1} predict a luminosity function that cuts off below $\mathcal{M}_B \approx -16$, and so are consistent with the luminosity function of normal high surface brightness galaxies. The low surface brightness dwarf galaxies presumably result from systems with $v_d < v_{\rm crit}$ which lose most of their gas by supernova-driven winds, as proposed by Larson (1974) and Dekel & Silk (1986). Such

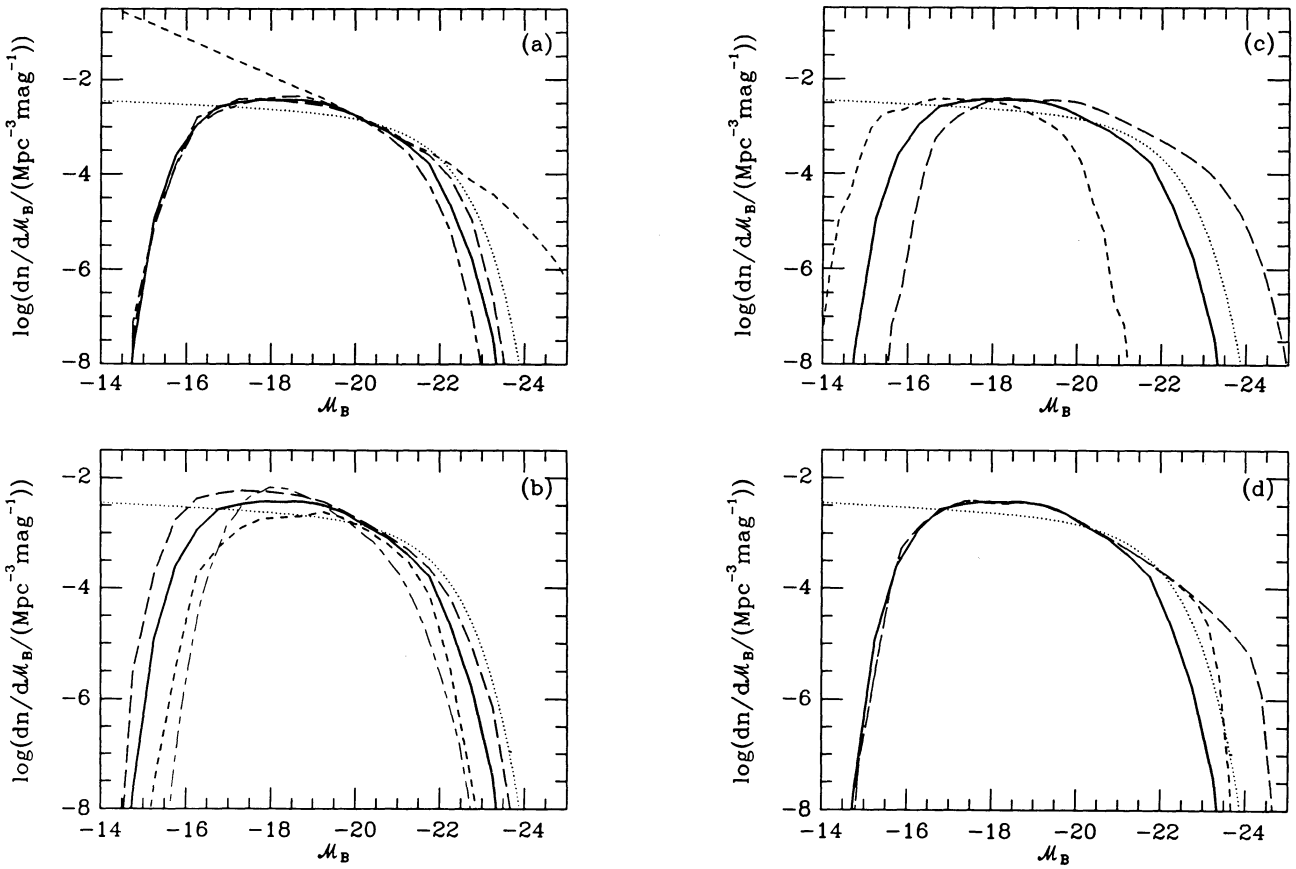

FIG. 4.—Luminosity function of galaxies (number density per unit absolute magnitude $\mathcal{M}_B$) at present epoch ($z = 0$). (*a*) *Short-dashed-long-dashed, solid, and long-dashed curves:* model luminosity distributions for CDM normalization $\sigma_8 = 0.5$ and star-formation cross-section parameter $A_* = (2, 5, 10)$, respectively. Solid curve is our standard model. Short-dashed curve is the luminosity function that would be obtained if the effects of supernova-driven mass loss and the cooling time constraint were neglected. *Dotted line:* observed Schechter luminosity function from Efstathiou et al. (1988). (*b*) *Short-dashed, solid, and long-dashed curves:* model luminosity distributions for $A_* = 5$ and $\sigma_8 = (0.4, 0.5, 0.6)$, respectively. Solid curve is our standard model. *Short-dashed-long-dashed curve:* luminosity function that would be obtained if we assumed a constant baryonic mass-to-light ratio $M_b/L_B = 5$. *Dotted line:* observed Schechter function. (*c*) *Short-dashed, solid, and long-dashed curves:* model luminosity distributions for baryon fraction $f_b = (0.03, 0.07, 0.14)$, respectively, with $A_* = 5$, $\sigma_8 = 0.5$. *Dotted line:* observed Schechter function. (*d*) *Solid, short-dashed, and long-dashed curves:* model luminosity function for initial gas metallicities $Z/Z_\odot = (0, 0.1, 1)$ respectively, with $A_* = 5$, $\sigma_8 = 0.5$.

galaxies are not included in the present model. There will be a very large number of these objects, and we speculate that, at least during their brief and prematurely interupted star-forming phase, such dwarfs may provide a dominant contribution to the very deep galaxy counts. This idea is discussed in § 5. Calculation of the luminosity function for the galaxies with $v_d < v_{\rm crit}$ requires a more detailed treatment of the interrelation between mass loss and star formation, and we defer this to a Paper II.

3.4. *Galaxy Color Distribution*

The distribution of $(B-V)$ colors of our galaxies is shown in Figure 5. Figure 5*a* displays the color distribution for our standard model in various ranges of absolute magnitude: $-23 < \mathcal{M}_B < -21$, $-21 < \mathcal{M}_B < -19$, and $-19 < \mathcal{M}_B < -17$. It can be seen that the bright galaxies ($\mathcal{M}_B \lesssim -19$) cover a range of colors from very blue ($B-V \approx 0.2$) to very red ($B-V \approx 0.9$), while the faint galaxies ($\mathcal{M}_B \gtrsim -19$) are mostly very red ($B-V \approx 0.8$–$1.0$). In fact, galaxies with $-17 < \mathcal{M}_B < -15$ (which we have not plotted) have an even narrower distribution around $B-V \approx 0.9$ than do those with $-19 < \mathcal{M}_B < -17$.

The reason for this relation between color and luminosity is that in our model, most of the faint, low-mass galaxies turned on at $z_* \sim 1$–3, and so are now quite old, while most of the bright, high-mass galaxies turned on at $z_* \lesssim 1$, and have a range of ages down to zero. In addition, because of the correlation between τ_* and z_*, most of the faint galaxies have star formation timescales $\tau_* \lesssim 3$ Gyr, so that star formation has effectively ceased, while many bright galaxies are still forming stars at close to their initial rate. The narrowness of the color distribution for the faintest galaxies in our model is enhanced because we have chosen a sharp cutoff $v_d > v_{\rm crit}$ to represent the effects of mass loss from feedback. This forces low-mass galaxies to be quite dense if they are to avoid mass loss, which requires that they form at relatively high redshift and have short star-formation time scales. A smoother cutoff, allowing for some star formation in mass-losing systems, would allow the existence of some low-mass galaxies (those which collapsed recently) with blue colors, although one would still expect most low-mass galaxies to be very red.

Figure 5*b* shows how the color distribution predicted for "bright" galaxies (defined to be those with $\mathcal{M}_B < -19$), depends on A_* in the range $A_* = 2$–10, for $\sigma_8 = 0.5$. It can be seen that the effect of increasing A_* is to shift the peak of the distribution to redder colors, although the total range of colors covered is not much changed, the latter being more sensitive to the range of turn-on redshifts z_*. Varying σ_8 over the range

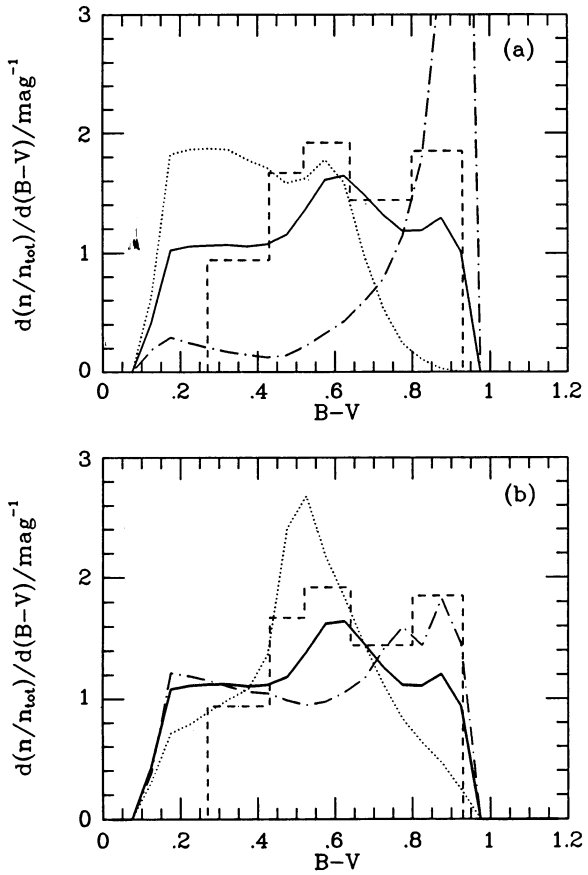

FIG. 5.—Distribution of $B-V$ color for volume-limited sample of normal surface brightness galaxies at present epoch. Each of the distributions is normalized to unit total area. (*a*) Color distributions for standard model, broken down according to absolute magnitude: dotted, solid, and dot-dashed curves are for $\mathcal{M}_B = -23 \rightarrow -21$, $-21 \rightarrow -19$, and $-19 \rightarrow -17$, respectively. *Dashed histogram:* observational color distribution for galaxies with $\mathcal{M}_B < -19$, constructed as described in text. (*b*) Color distribution for bright ($\mathcal{M}_B < -19$) galaxies. *Dotted, solid, and dot-dashed curves:* model results for $\sigma_8 = 0.5$, $A_* = (2, 5, 10)$, respectively. Solid curve is our standard model. *Dashed histogram:* observational distribution, as in Fig. 5*a*.

0.4–0.6 at fixed A_* has a much smaller effect; increasing σ_8 tends to make the distribution flatter and less peaked.

Given the existing observational data, the most precise test of the galaxy color distribution would come from comparing predicted and observed color distributions in an apparent magnitude-limited sample. Since this involves computing galaxy number counts, which is the subject of Paper II, we make a more approximate comparison here, based on the color distribution in a volume-limited and absolute magnitude limited sample. We construct an observational $B-V$ color distribution for bright galaxies to compare with the models by the following procedure: King & Ellis (1985) have determined Schechter luminosity function parameters separately for galaxies of different morphological types. From this, we can calculate the number density fractions of different morphological types at $\mathcal{M}_B < -19$, with the following result: E/S0: 24%, Sab: 23%, Sbc: 23%, Scd: 15%, Sdm: 15%. We assign a $B-V$ color range to each morphological type based on de Vaucouleurs (1977), who gives the mean color and the dispersion in the color for different morphological types, based on the galaxy sample in the Second Reference Catalogue (de Vaucouleurs, de Vaucouleurs, & Corwin 1976), which contains mostly intrinsically fairly bright galaxies. Thus we obtain the following $B-V$ color ranges: E/S0: 0.80–0.93, Sab: 0.64–0.80, Sbc: 0.52–0.64, Scd: 0.43–0.52, Sdm: 0.27–0.43. Combining this with the number fractions as a function of morphological type gives a rough estimate of the color distribution, which is plotted as a histogram in Figures 5*a* and 5*b*. A significant source of uncertainty in the color distribution results from correcting the observed colors for extinction by dust in our Galaxy. Huchra (1977a), who uses the same observational data as de Vaucouleurs but applies a different Galactic extinction correction, finds colors at a given morphological type that are roughly 0.1 mag redder in $B-V$ compared to de Vaucouleurs (1977).

Comparing the predicted with the observed color distributions for bright galaxies, one sees that choosing A_* somewhere in the range $A_* = 5$–10 seems to give the best fit. With this choice, the cutoff in the color distribution at the red end is at the right place ($B-V \approx 0.9$), but there seems to be an excess of very blue galaxies (with $B-V \approx 0.1$–0.3). This excess of blue galaxies results from objects which have only recently initiated vigorous star formation. Since these blue galaxies are fairly gas-rich, extinction by dust within the galaxy (which is not included in the Bruzual spectral evolution models) is expected to be significant, and would remove some or even all of this discrepancy. In fact, Huchra (1977b) finds observationally that internal extinctions $E(B-V) \approx 0.1$ are quite typical for the bluest galaxies. Internal reddening would not affect the very red galaxies in our models, since they are gas-poor.

Allowing for the uncertainties in extinction, the overall range of galaxy colours predicted by the models is in rough agreement with that observed. However, the prediction that more luminous galaxies should have bluer colors is worrying. An alternative way of presenting this correlation is to plot the luminosity functions separately for different colors, which we do in Figure 6, for the $B-V$ ranges 0.1–0.4, 0.4–0.8, and 0.8–1.1. These ranges correspond to Irr's, S's and E/S0's according to the morphological type-color relation of de Vaucouleurs (1977). Figure 6 shows that, in the models, bluer galaxies tend to be more luminous. This is in some contradiction with observations, which indicate that the brightest galaxies tend to be ellipticals (e.g., King & Ellis 1985; Binggeli et al. 1988). This result is subject to the same caveats about internal extinction,

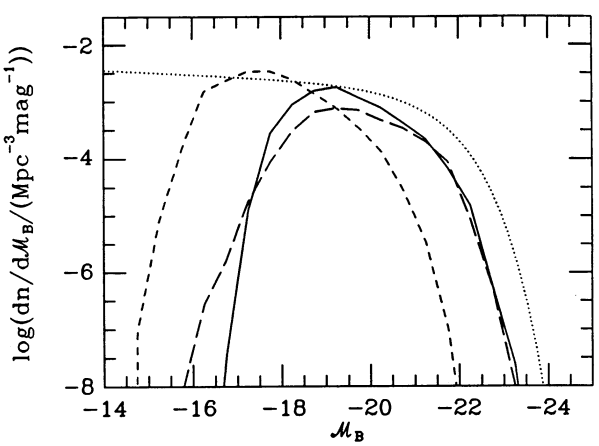

FIG. 6.—Dependence of galaxy luminosity function on $B-V$ color at present epoch. Long-dashed, solid, and short-dashed curves are luminosity functions in the standard model for $B-V = 0.1$–0.4, 0.4–0.8, and 0.8–1.1, respectively. *Dotted curve:* observed Schechter function for galaxies of all colors, as in Fig. 4.

which will tend to reduce the observed luminosities of the bluer galaxies, but such effects seem unlikely to be large enough to completely remove the discrepancy, given that observed elliptical galaxies are red by virtue of the age of their stellar populations rather than because of extinction. The basic source of the problem in our model is simply that a large fraction of massive galaxies initiate star formation at quite low redshift. It seems likely that the model of White & Frenk (1991) will suffer from the same problem with colors, since in that model too, the most luminous galaxies form at low redshifts.

An obvious mechanism to suppress the recent formation of excessively blue galaxies would be to appeal to a low Ω universe, since then the growth of structures in the linear regime would cease for $z \lesssim \Omega^{-1} - 1$. However this would make it even more difficult to understand the evidence for evolution at modest redshifts in either the luminosity function of galaxies or in their comoving number density that is being inferred from a variety of arguments. In addition to the interpretation of the deep number counts of faint galaxies in the B and K bands (Cowie et al. 1991), there is accumulating evidence for significant evolution at redshifts $z \sim 0.1$, from sources as varied as the APM number counts at magnitudes $B \gtrsim 17$ (Maddox et al. 1990), the sparse-sampled *IRAS* redshift survey (Saunders et al. 1990), the incidence of [O II] emission at $z = 0.1$–0.3 (Broadhurst, Ellis, & Shanks 1988; Colless et al. 1990), and the incidence of Mg II absorption line systems in quasars (Bergeron & Boisse 1990). Whether this evolution is primarily in luminosity or in comoving number density is not yet known. Another effect which would reduce the numbers of very luminous blue galaxies would be the fact that massive galaxies may form by the merger of smaller objects which are already partly stellar, while in our models we have assumed that each galaxy forms from pristine gas.

4. RESULTS OF MODELS FOR DISTANT GALAXIES

The prescription for galactic star formation that we have developed readily yields predictions for galaxy properties at higher redshifts. The main quantitative observational data constraining the behavior of high redshift galaxies are the galaxy number counts at faint magnitudes, and perhaps also the distribution of damped Lyα clouds seen in absorption towards distant quasars, the latter depending on the properties of "retarded" or "dark," as well as luminous, galaxies. A detailed comparison with the number counts will be given in Paper II, but we present some preliminary results in § 5.

The number of galaxies, $dn/d\ln M_b$, which have collapsed and begun star formation, per unit *comoving* volume per unit log baryon mass, for redshifts $z = (0, 1, 3, 5)$ and $\sigma_8 = (0.4, 0.5, 0.6)$ is shown in Figure 7. It can be seen that the number density of objects decreases with increasing redshift, but that this evolution is very slow at low masses, and much faster at high masses. This is because low-mass peaks on average have larger overdensities δ than high-mass peaks, and so collapse at higher z. The number density of objects also increases with increasing σ_8, this dependence being much steeper at higher redshift. Figure 8 shows the distribution of star-formation time scales τ_* for star-forming objects at the same redshifts. This distribution becomes narrower and shifts to smaller τ_* with increasing redshift as a result of the shift in the mass distribution to smaller masses.

In Figure 9 we show how the luminosity function in the *rest-frame B* and *R* bands depends on redshift: the comoving number density of galaxies is plotted against the absolute magnitude. Our version of Bruzual's program used for calculating magnitudes and colors gives inaccurate results for small ages when the star-formation time scale is also short. Because of this

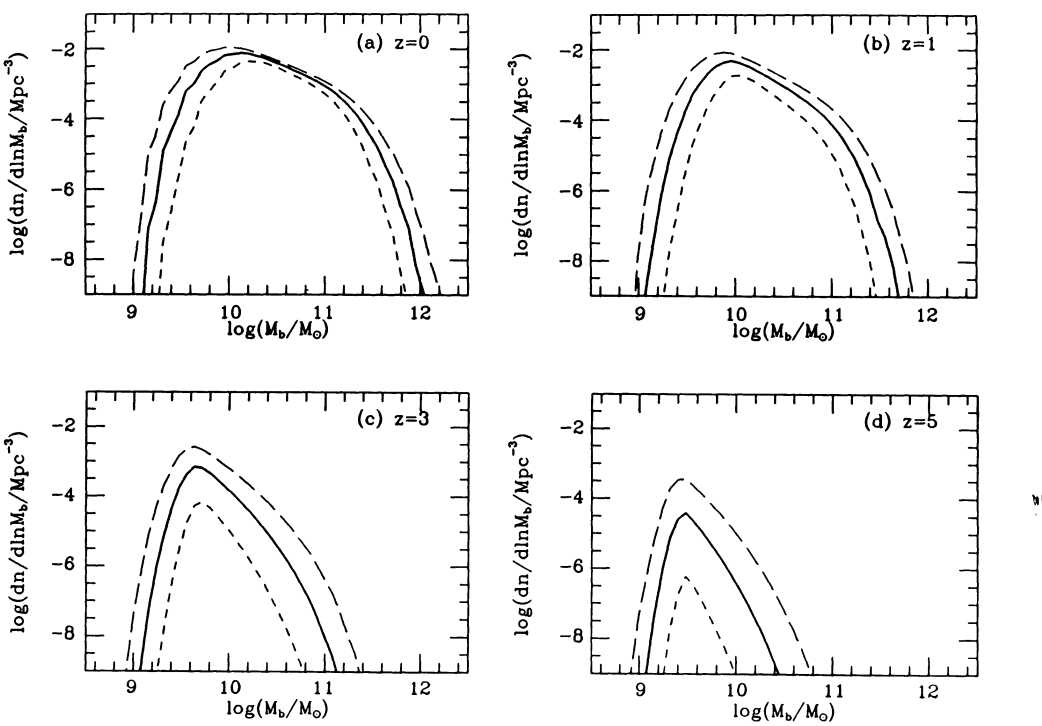

FIG. 7.—Comoving number density of luminous galaxies as function of baryonic mass M_b for various redshifts. In each panel, short-dashed, solid, and long-dashed curves are for $\sigma_8 = (0.4, 0.5, 0.6)$ respectively. Solid curve is the standard model. (*a*) $z = 0$; (*b*) $z = 1$; (*c*) $z = 3$; (*d*) $z = 5$.

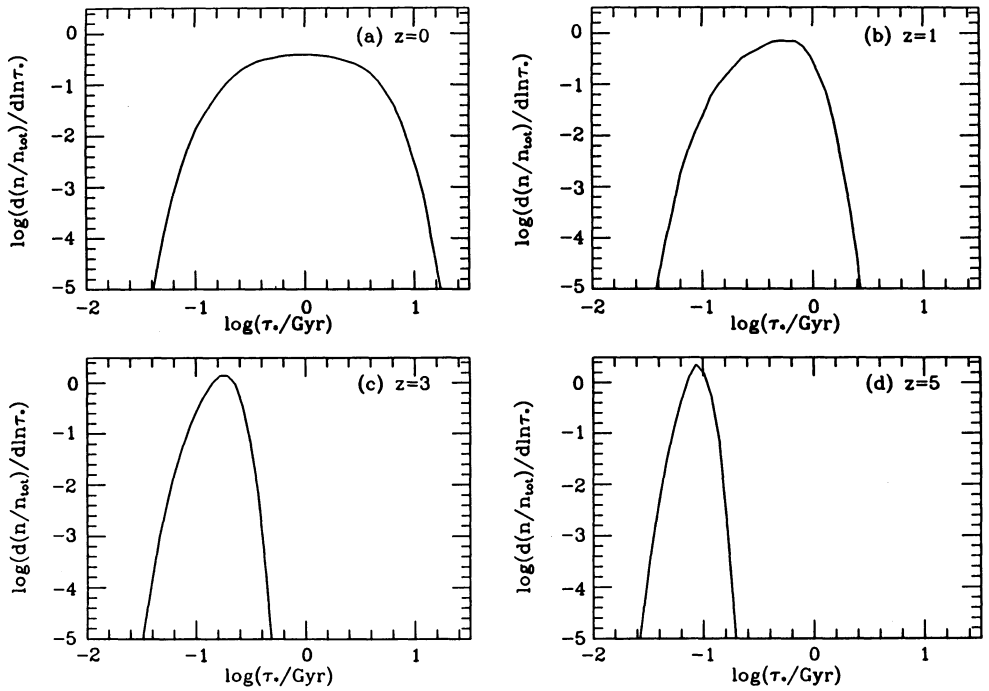

FIG. 8.—Distribution of star-formation time scales τ_* for luminous galaxies for various redshifts, for standard model. All distributions are normalized to unit total area. (*a*) $z = 0$; (*b*) $z = 1$; (*c*) $z = 3$; (*d*) $z = 5$.

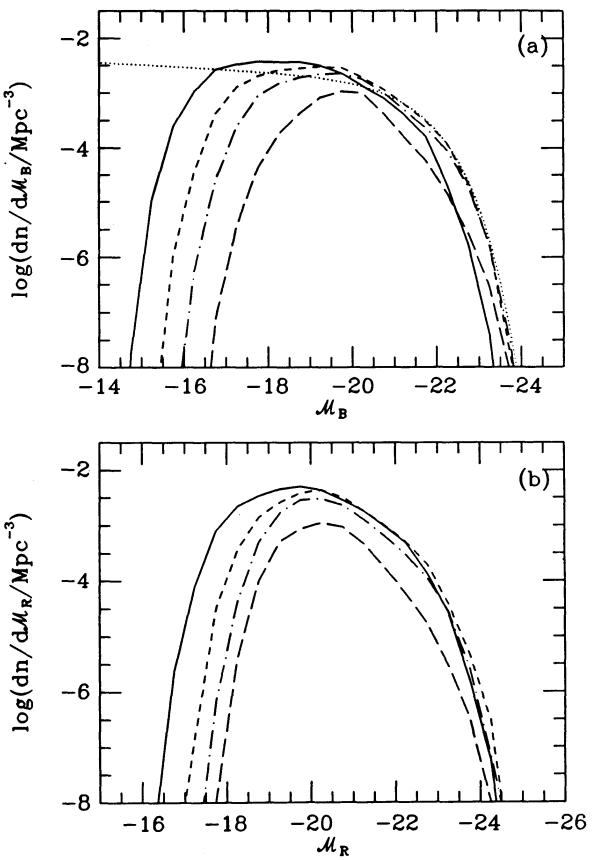

FIG. 9.—Rest-frame luminosity function (comoving number density per unit absolute magnitude) as a function of redshift, in (*a*) *B* band, and (*b*) *R* band. Solid, short-dashed, dot-dashed, and long-dashed curves are for $z = 0$, $z = 0.5$, $z = 1$ and $z = 2$, respectively. Dotted curve in panel (*a*) is the observed $z = 0$ Schechter luminosity function, as in Fig. 4.

limitation (to be remedied in Paper II), we only give results here for redshifts up to 2. In the *B* band, the luminosity function evolves fastest at the *low-luminosity* end, despite the fact that the mass distribution evolves fastest at the *high-mass* end. This is because the *B* band luminosity is sensitive to the presence of more massive stars, and so fades rapidly in time for galaxies in which the star-formation time scale is short, as is the case for the low-mass galaxies in our model. The lower mass galaxies brighten with increasing redshift, tending to compensate for the loss of high-mass galaxies from the mass function. This is the reason why the luminosity function actually *increases* at the bright end between $z = 0$ and $z = 1$. In the *R* band, the luminosity is less sensitive to the presence of high-mass stars, and so this brightening effect is smaller. In the *R* band, the bright end of the luminosity function changes very little between $z = 0$ and $z = 1$.

In Figure 10, we display the *rest-frame B-R* color distributions for the same redshifts, $z = 0$, 0.5, 1, 2 as in Figure 9. The color distributions are plotted separately for different ranges of absolute *B*-magnitude $\mathcal{M}_B$. It can be seen that all galaxies except for the faintest (which form at the highest redshift) tend to become bluer with increasing redshift, as a result of their undergoing more active star formation at early times.

5. GALAXY COUNTS AT FAINT MAGNITUDES: BURSTING DWARF GALAXIES AT $z \sim 1$?

Although a thorough treatment of galaxy number counts is deferred to Paper II, we have made some rough estimates of the counts expected on the basis of the results presented in § 4 for the luminosity function at high redshift. The quantity dN/dm, the number of galaxies per unit solid angle per unit apparent magnitude m, is given by the integral

$$\frac{dN}{dm} = \int_0^\infty \frac{dn}{dm}\frac{dV}{dz}\,dz\,, \qquad (5.1a)$$

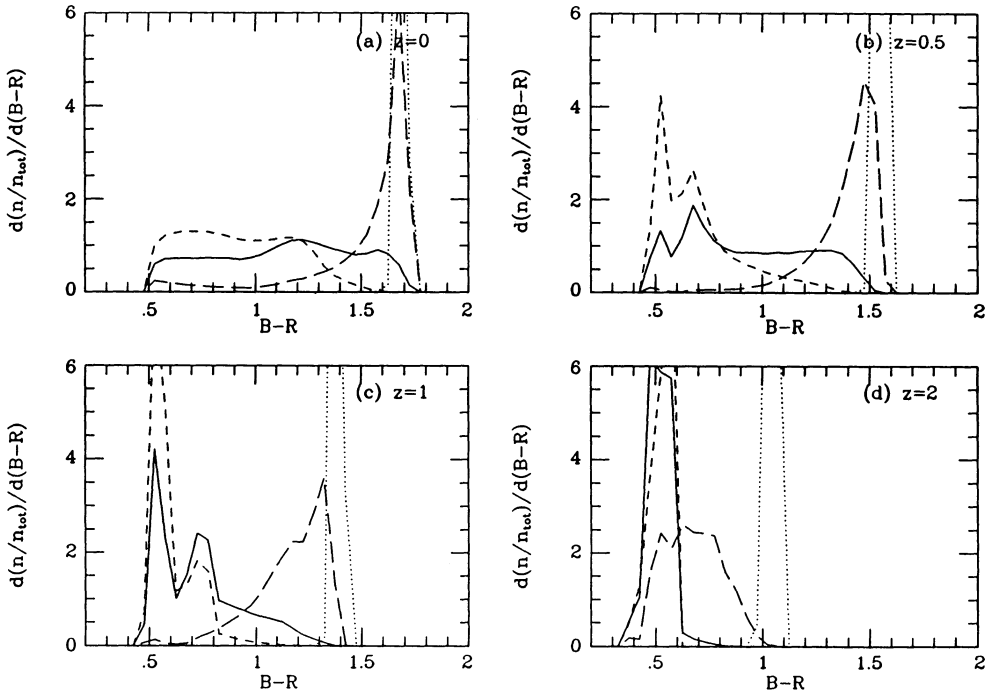

FIG. 10.—Rest frame B-R color distributions as a function of redshift, for (*a*) $z = 0$, (*b*) $z = 0.5$, (*c*) $z = 1$, and (*d*) $z = 2$. The galaxies have been separated according to their rest-frame absolute B-magnitudes: short-dashed, solid, long-dashed, and dotted curves are for $\mathcal{M}_B = -23 \rightarrow -21$, $-21 \rightarrow -19$, $-19 \rightarrow -17$ and $-17 \rightarrow -15$, respectively. All curves are normalized to unit total area.

where dn/dm is the number density of galaxies per unit comoving volume per unit apparent magnitude, and dV/dz is the comoving volume per unit redshift per steradian, which for $\Omega = 1$ is

$$\frac{dV}{dz} = 4\left(\frac{c}{H_0}\right)^3\left(1 - \frac{1}{\sqrt{1+z}}\right)^2 \frac{1}{(1+z)^{3/2}} \qquad (5.1b)$$

(e.g., Weinberg 1972). We have used this to estimate counts in the B band, and the results (including the effects of K-corrections) are plotted in Figure 11. We emphasize that this estimate is a preliminary one, as our calculation of the luminosity function becomes inaccurate for $z \gtrsim 2$, for the reasons given in § 4. However, most of the contribution to the counts in our models seems to come from $z < 2$, even for $B \gtrsim 24$, so that our estimate is probably not in error by more than $\sim 50\%$. Figure 11 also shows the effect on the counts of varying the assumed baryonic fraction from our standard value $f_b = 0.07$; we show results for $f_b = 0.03$ and $f_b = 0.14$. It can be seen that the amplitude of the counts at $B \lesssim 23$ is fairly sensitive to f_b, because of the strong dependence of the bright end of the luminosity function on this parameter.

We have also plotted in Figure 11 a representative selection of the observational data on counts in the B_J band. (The B_J band differs slightly from the standard B band for which we compute our models, but we have not made any correction in plotting the data.) A full comparison with the observations will be given in Paper II, but the following points are already clear: (1) The model counts roughly fit those observed up to $B \lesssim 24$. (2) For $B \gtrsim 24$, the model counts slowly decline, while the observed counts continue to increase up to at least $B \approx 27$. Thus, the counts predicted by the model are too low at very faint magnitudes, this deficiency reaching a factor ~ 30 at $B = 27$. The reason for the decline of the counts in our model at $B \gtrsim 24$ is the relatively late epoch of galaxy formation: most massive galaxies begin star formation only at $z \lesssim 2$.

This same problem of predicting too low number counts at faint magnitudes in an $\Omega = 1$ universe is also found in models that attempt to work backwards from the present epoch, allowing for galaxy evolution but assuming a constant comoving number density of galaxies (e.g., Tyson 1988; Guiderdoni & Rocca-Volmerange 1990; Koo 1990). These models generally assume a large redshift of galaxy formation, $z_f \sim 5$–10, so the discrepancy is not as large as in our models, but it is still a serious problem. One possible solution to this problem consis-

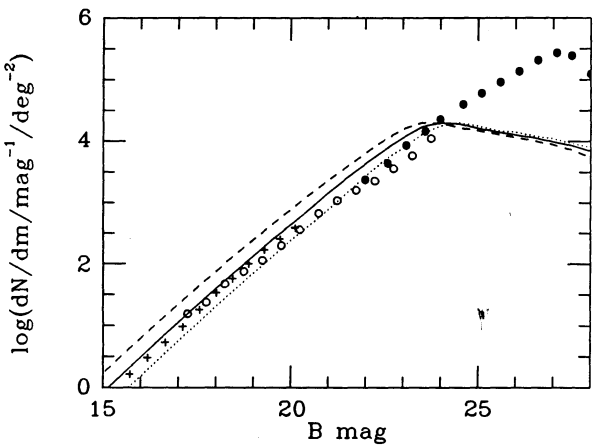

FIG. 11.—Galaxy number counts in the B band. Dotted, solid, and dashed curves are the model predictions for $A_* = 5$, $\sigma_8 = 0.5$, and $f_b = (0.03, 0.07, 0.14)$ respectively. Crosses, open circles, and filled circles are observational data from Maddox et al. (1990), Jarvis & Tyson (1981), and Tyson (1988), respectively. The turnover in Tyson's (1988) data at $B > 27$ depends on the correction made for incompleteness, and may not be real.

tent with having $\Omega = 1$ is to suppose that there has been extensive merging of galaxies since they first started forming stars, so that there were more, but fainter, galaxies at high redshift than predicted by no-merger models. Rocca-Volmerange & Guiderdoni (1990) have shown that the counts can be reproduced assuming a simple phenomenological prescription for the evolution of the galaxy number density, but further work on the dynamics of merging galaxies is required to show whether their prescription is physically reasonable.

We wish here to propose an alternative scenario for explaining the faint counts. The idea is that some fraction of the low-mass galaxies with $v_d < v_{\rm crit}$, whose contribution to the luminosity function we have up to now assumed to be negligible, would undergo bursts of star formation at $z = 1$, the burst being abruptly truncated when a supernova-driven wind cleared the galaxy of gas. The galaxy would be very bright during the burst, but would fade very rapidly thereafter, so that such galaxies might contribute to the faint counts, but be essentially invisible at low redshift. Consider a specific example, of a burst of duration $\tau = 1$ Gyr, during which a mass M_* of stars is formed at a uniform rate, beginning at $z_f = 1.4$, and ending at $z = 1$. During the burst, the observed apparent B-magnitude is roughly constant at the value $B \approx 26.1 - 2.5 \log (M_*/10^9\ M_\odot)$, but after the burst, B declines by 3 mag in a redshift interval of only $\Delta z \approx 0.2$. This type of behavior results because at this redshift, the observed B band light is dominated by stars with very short lifetimes. The principal observational constraint on the redshift at which the bursts occur is that most of the faint ($B \gtrsim 24$) objects are inferred from broad-band colors to have a spectrum which is roughly flat in f_ν in the U through I bands (Songaila, Cowie, & Lilly 1990; Guhathakurta, Tyson, & Majewski 1990), which implies that $1 \lesssim z \lesssim 3$. Tyson's (1988) counts in the range $24 \lesssim B \lesssim 27$ are fitted by the relation $dN/dB \approx 5 \times 10^4 \times 10^{0.4(B-25)}$ deg^{-2} mag^{-1}. For $B \gtrsim 24.5$, this is approximately the excess of the observed counts over that predicted by our model, which we are ascribing to the new population of objects. If we assume that the galaxies producing these counts have a constant comoving number density within a redshift shell $1 < z < 1.4$, then equation (5.1) gives $dN/dB \approx (dn/dB)(dV/dz)\Delta z$, so we infer $dn/dB \approx 1.5 \times 10^{-2} \times 10^{0.4(B-25)}$ Mpc^{-3} mag^{-1} for the comoving number density per unit observed magnitude. If we assume that the galaxies are all undergoing 1 Gyr bursts, then we can convert from B to M_*, giving $dn/d\ln M_* \approx 5 \times 10^{-2}(M_*/10^9\ M_\odot)^{-1}$ Mpc^{-3} for the distribution of burst masses, for the mass range $0.4 \lesssim M_*/10^9\ M_\odot \lesssim 4$. We can compare this with the distribution of baryonic masses in peaks. If we consider peaks of all overdensities, we obtain $dn/d\ln M_b \approx 0.3(M_b/10^9\ M_\odot)^{-1}$ Mpc^{-3}, independent of z, and valid over the entire mass range $10^6 < M_b/M_\odot < 10^{12}$. However, at any given redshift, only some fraction of peaks will have collapsed, and in general only some fraction of these will have begun forming stars. If we apply our model of interaction-induced star formation (without the $v_d > v_{\rm crit}$ condition) to predict which peaks have commenced star formation, then at $z = 1.2$ we find $dn/d\ln M_b \approx 8 \times 10^{-2}(M_b/10^9\ M_\odot)^{-1.2}$ Mpc^{-3} for $0.1 \lesssim M_b/10^9\ M_\odot \lesssim 10$ (assuming $\sigma_8 = 0.5$, $R_{\rm gr}/R_{\rm g} = 2$). Since each burst is assumed to be terminated by ejection of the remaining gas, we assume that the burst mass M_* is related to the total baryonic mass M_b by $M_* = f(M_b)M_b$, with $f = \min [1, (M_b/M_{b\rm crit})^\gamma]$. We obtain consistency between the distributions in M_* and M_b by taking $\gamma = 0.2$ and $M_{b\rm crit} \approx 6 \times 10^9\ M_\odot$. Then we can explain the excess counts in the range $24.5 \lesssim B \lesssim 27$ as arising from galaxies with initial baryonic masses in the range $0.7 \lesssim M_b/10^9\ M_\odot \lesssim 5$. Note that the critical mass $M_{b\rm crit}$ for mass loss that we have inferred in this example coincides approximately with the location of the low-end cutoff in the $z = 0$ mass function obtained with the condition $v_d > v_{\rm crit} = 100$ km s^{-1}. Thus, the idea that the galaxies undergoing the bursts are some subset of those with $v_d < v_{\rm crit}$ is reasonably self-consistent, provided the cutoff in the luminosity function is not too abrupt. Further discussion of this point is postponed to Paper II.

There is some freedom in choosing the parameters of the burst model so as to fit the excess counts. For instance, one could assume that the individual bursting galaxies have the same burst duration τ, but are spread over a redshift interval Δz (corresponding to time interval Δt) which is larger by a factor q, with the same mean z [and still with $\Delta z \ll (1 + z)$]. Then the inferred dn/dB scales as $dn/dB \propto 1/\Delta z \propto 1/q$, as does the number density of bursting galaxies active at any one time. However, we now require a number of successive generations of bursting galaxies equal to $\Delta t/\tau \propto \Delta z \propto q$, so that the total number of galaxies $dn/d\ln M_*$ which undergo a burst at some stage in their life is independent of q. Alternatively, one could keep Δz the same but change τ by a factor q. The star formation rate during the burst would scale as $1/\tau \propto 1/q$ (for fixed M_*), and the B-band flux would scale in approximately the same way, so the inferred burst mass for a given B magnitude scales as $M_* \propto q$. Then the inferred number density of bursts at a given M_* at a given time scales as q (in the case $dN/dB \propto 10^{0.4B}$), but this is compensated by the number of generations of bursts $\propto 1/q$, so that the total value of $dn/d\ln M_*$ at a given value of M_* is unchanged, but the relation between B and M_* is shifted. Thus our burst model is effectively invariant under changes in τ or Δz (subject to the condition $\tau \leq \Delta t$). However, the model breaks down if we increase the mean redshift $\bar{z}$ of the burst epoch to $\bar{z} \gtrsim 2$—then we find that the number density of star-forming peaks is too low to explain the counts in the range $24 \lesssim B \lesssim 27$. However, if for a peak to form stars we only required that the peak itself had collapsed, rather than imposing the stronger constraint that the surrounding group had also collapsed, we would have more freedom to explain the counts with bursting populations at higher redshifts.

One result one can derive that is relatively independent of the details of the burst model is the total amount of star formation involved: integrating over the burst mass function $dn/d\ln M_*$ derived from fitting to the excess counts in the range $24.5 \lesssim B \lesssim 27$, one obtains $\Delta\rho_* \equiv \int M_*\, dn \approx 1 \times 10^8 F(\bar{z})\ M_\odot$ Mpc^{-3}, where $0.7 \lesssim F(\bar{z}) \lesssim 1.3$ for $1 < \bar{z} < 3$. This result is insensitive to τ (in the range $0.1 \lesssim \tau \lesssim 1$ Gyr) and to Δz, for the reasons already given. This is a significant fraction of all the star formation in our model. By $z = 0$, the amount of stars formed in "normal surface brightness" galaxies in our model is $\rho_*(v_d > v_{\rm crit}) \approx 3 \times 10^8\ M_\odot$ Mpc^{-3}, so that the burst population adds another $\sim 30\%$ to this. One can alternatively compare to a value of ρ_* estimated directly from the observations: integrating over the Efstathiou et al. (1988) luminosity function gives a luminosity density $\mathscr{L}_B \approx 1 \times 10^8\ L_\odot$ Mpc^{-3}, and combining with an assumed mean stellar mass-to-light ratio $M_*/L_B = 4$ gives $\rho_* \approx 4 \times 10^8\ M_\odot$ Mpc^{-3}, so that the burst population would account for $\sim 20\%$ of all the star formation in the universe. This conclusion is essentially in agreement with that of Cowie (1988), who finds that the star formation associated with the faint counts would produce

a significant fraction of all the metals seen in present-day galaxies.

If one tries to explain the faint counts in terms of a bursting population at $z \sim 1$, then one must consider whether the relics of these bursts would be visible now. Consider a galaxy which has $B = 26$ at $z \approx 1$ in our $\tau = 1$ Gyr burst model. This would have a burst mass $M_* \approx 1 \times 10^9\ M_\odot$ and a present epoch ($z = 0$) absolute magnitude $\mathcal{M}_B \approx -15$. Tyson finds that galaxies with $B > 26$ typically have half-light radii (assuming an exponential surface brightness profiles) $\theta_{1/2} \approx 0''.4$, corresponding to a physical radius $r_{1/2} \approx 3.5$ kpc. This implies a mean surface brightness within the half-light radius $\bar{\mu}_B \approx 26.5$ mag arcsec^{-2} at $z = 0$, which fortuitously is almost the same as the value at $z \approx 1$, because the effects of evolutionary fading, k-correction and $(1 + z)^{-4}$ bolometric brightening almost cancel. These particular objects (with $B = 26$ at $z \approx 1$) would have a present-day number density $dn/d\mathcal{M}_B \approx 4 \times 10^{-2}$ Mpc^{-3} mag^{-1}. With their low surface brightness, they would probably have avoided detection in most surveys. These estimates depend on the burst lifetime τ that is assumed: the present-day luminosity and surface brightness corresponding to a given apparent magnitude in the faint counts scale approximately linearly with τ.

The model we have sketched to explain the faint counts is obviously incomplete. The parameters we have chosen for the bursting population are somewhat ad hoc. In our model, one expects star-formation time scales $\tau_* \sim 0.1$–1 Gyr for $M_b \sim 10^9\ M_\odot$ objects turning on at $1 \lesssim z_* \lesssim 2$, if no $v_{\rm crit}$ condition is applied. However, many objects of similar mass are then expected to turn on at $z_* < 1$, and we have not explained in detail what would happen to these, although one would expect mass loss to be more effective, since their halos would have lower escape velocity. We have also not explained why the bursting systems have such large radii for their mass, although this could be a result of energy injection from supernovae keeping the gas dispersed even while it is forming stars. More work on this scenario is required. However, the numerical example we have given does indicate the general feasibility of the scenario. We note that the models of White & Frenk (1991), based like ours on a CDM power spectrum with $\Omega = 1$, but with different assumptions about the star formation rate, also produce sufficient faint galaxy counts for certain values of their parameters, with the contributing galaxies having similar values for the halo velocity dispersion, $v_d \sim 50$–100 km s^{-1}, to those in the model we have sketched. However, their models also predict far too many faint galaxies in the present-day luminosity function.

6. DISCUSSION AND CONCLUSIONS

We have seen in § 3 that our model for star formation, when combined with the Bruzual spectral evolution code and the CDM power spectrum, provides a reasonable fit to the luminosity and color distributions of normal surface brightness galaxies at the present epoch, when we choose a suitable value for the parameter A_*. In particular, the number density of L_* galaxies comes out about right without any special tuning of parameters. The model produces a "natural bias," in that only galaxies inside groups which have collapsed by the present epoch will form stars and be luminous. As already mentioned, some form of "bias" seems to be required in any model that attempts to reproduce the luminosity function of bright galaxies and have $\Omega = 1$, but this bias is generally put in in a more ad hoc way, in the form of a threshold condition on the peak height (e.g., BBKS, Davis et al. 1985). In addition, the model will generate a morphology-density correlation of qualitatively the correct form, in that galaxies in dense environments will on average convert their gas to stars on a much shorter time scale than galaxies in low-density environments, although we have not made a quantitative comparison with the observations. Of course, the results for the number density of luminous galaxies would be changed if we chose a ratio of group to galaxy mass greatly different from $M_{\rm gr}/M_{\rm g} = 8$, but a value close to this seems reasonable on a priori grounds if we are interested in interactions between galaxies and their near neighbors.

The comparison of the models with observations does not pin down the value of A_* very precisely, but a value within a factor of 2 of $A_* = 5$ seems to be required to be consistent with the observed color distribution. One can then ask how physically reasonable such a value is. In the Appendix we present an approximate physical derivation of A_* based on the idea that tidal encounters which induce a velocity perturbation in the gas larger than its sound speed induce a burst of star formation. The derived value depends somewhat on the velocity dispersion of the "target" galaxy halo, but for a galaxy similar to ours we obtain $A_* \approx 10 f_*$, where f_* is the fraction of the gas converted into stars in a single encounter. This estimate assumes a mass function of perturbing galaxies $dn/dM \propto M^{-2}$, and it turns out that most of the contribution to A_* comes from perturbers much less massive than the target galaxy. We see that to obtain $A_* \approx 5$, as the models seem to require, we would need $f_* \approx 0.5$. The value of f_* is hard to estimate by purely theoretical means, but its value for spiral galaxies can be constrained by the following observational argument. If the star formation event triggered by each tidal interaction lasts a time Δt_*, then the fraction of the time that a galaxy is actively forming stars (its "duty cycle") is max $[\Delta t_*/f_* \tau_*, 1]$. Since most spirals are observed to be forming stars, we need this fraction to be of order 1, which implies $f_* \lesssim \Delta t_*/\tau_*$. Now Δt_* can be at most a few dynamical times; for our galaxy, we take $\Delta t_* \approx 5 \times 10^8$ yr, and $\tau_* \approx 5$ Gyr, the latter value being required to give the correct present-day ratio of gas to stars. This gives $f_* \lesssim 0.1$, which would imply $A_* \lesssim 1$. This estimate of f_* is also roughly consistent with the values inferred by modeling of interacting galaxies observed to be undergoing bursts of star formation (Kennicutt et al. 1987). Thus the value of A_* that the models require seems to be somewhat larger than one can justify theoretically. However, the calculation in the Appendix neglects the self-gravity of the galactic disk in calculating the perturbation of the gas velocity, which could significantly amplify the response to a small perturbation, and thus increase the value of A_*, so a final judgement on this issue is not yet possible.

We have made a rough estimate of the galaxy number counts predicted by this model, and find approximate consistency with observations at magnitudes $B \lesssim 24$. At fainter magnitudes, the model in its present form, which only includes the contribution of galaxies with halo velocity dispersions $v_d > v_{\rm crit} = 100$ km s^{-1}, predicts many fewer galaxies than are observed. However, we have shown how including the dwarf galaxies, with $v_d < v_{\rm crit}$, may provide an explanation of the faint counts, if some fraction of these dwarfs are assumed to undergo bursts of star formation at $z \sim 1$, before having their remaining gas ejected by supernova-driven winds. The deep K-band counts (Cowie et al. 1991) will be discussed in a sequel to this paper.

The dwarf population that we invoke consists of low σ peaks and should therefore be more weakly clustered than the luminous galaxies. Even allowing for mass loss, the symptotic $dn/dM \propto M^{-2}$ dependence of the mass function means that the low-mass dwarf population will have contributed a significant fraction of the total amount of star formation. They will also contribute to the UV ionizing background at high redshift, if the Lyman continuum photons escape, and to the integrated optical background. The existence of high redshift ($z \gtrsim 3$) galaxies, although discovered via radio selection (Spinrad 1989), tells us that massive galaxies, albeit rare, exist at such large redshifts. Inspection of our predictions of the redshift dependence of the mass function tells us that a model with relatively low biasing consistent with large-scale structure constraints is favored by the observed abundance of these objects ($\sim 10^{-9}$ Mpc^{-3}; Lilly 1990, private communication), presumed to have baryonic masses $\gtrsim 10^{11}\ M_\odot$. Our model does predict that large redshift galaxies have high star-formation rates, which is certainly consistent with the available data.

Our model for star formation thus has considerable success in explaining the observed luminosity and color distributions of galaxies, although there are some problems. Obviously, the model needs to be refined, since the assumptions made in the present version are rather crude. One weakness of the model is that it does not include merging of galaxies. According to N-body simulations, merging of dark halos continues down to $z = 0$, but the luminous baryonic cores of galaxies need not merge when their halos do. The dynamics of this process are at present uncertain, and it is possible that the baryonic cores of galaxies have undergone only a fairly small amount of merging since they began significant star formation, which in our model is at $z \lesssim 3$. It seems that spirals can have suffered little merging since they began forming stars in their disks. However, merging may have been more important for ellipticals. In particular, many of the galaxies which in our model begin forming stars from pristine gas at low redshifts may in reality form by merging of smaller galaxies which are already partly stellar. This would help to alleviate the problem of the brightest galaxies in our model having blue rather than red colors, although it would also modify some of our other predictions. Our assumption that star formation only occurs as a result of galaxy interactions also may be too extreme. Although such interactions undoubtedly enhance the star-formation rate, there is likely to also be some star formation under more quiescent conditions. We will return to these points in future papers.

We are indebted to G. Bruzual for providing us with a copy of his spectral evolution program, and to S. Djorgovski for providing us with some related software. We have greatly benefitted from discussions with B. Rocca-Volmerange and B. Guiderdoni. We thank M. Davis and S. Cole for useful conversations, and the referee, S. White, for useful comments. This research was supported by NSF grant AST-8819802 at Berkeley, and by a SERC Advanced Fellowship awarded to CGL at Oxford.

APPENDIX

A PHYSICAL DERIVATION OF A_*

In this Appendix, we attempt to estimate the value of the coefficient A_* in our expression (2.21) for the star-formation rate by a physical argument. We model the "target" halo as a singular isothermal sphere with circular velocity V_H and truncation radius r_H. These are related to the three-dimensional velocity dispersion v_T and radius r_T of a uniform density ("top hat") halo of the same mass and binding energy by $v_T = V_H$ and $r_T = \frac{3}{5} r_H$. Thus the halo mass is $M_H = V_H^2 r_H / G$. The gas is assumed to be in a disk of radius r_D and sound speed c_s. We consider an encounter between this galaxy and a perturbing galaxy similarly modeled, having halo mass M_p, radius r_p, and moving at relative velocity V_p. Both galaxies are in the same group of mass $M_{\rm gr}$. For the purpose of this calculation, we assume that all virialized groups and galaxy halos satisfy a radius-mass relation $r \propto M^\beta$, and a corresponding velocity-mass relation $V \propto M^{(1-\beta)/2}$. If the power spectrum of density fluctuations can be approximated as a power-law, $P(k) \propto k^n$, then the rms overdensity scales as $\delta \propto M^{-(n+3)/6}$, so $\beta = (n+5)/6$ for the "average" radius-mass relation (see eq. [2.12]). For the CDM spectrum with $M \sim 10^{12}\ M_\odot$, $n \approx -2$, so we take $\beta = \frac{1}{2}$. We assume a ratio of group to galaxy mass $M_{\rm gr}/M_H = 8$, so the typical relative velocity in a collision between two galaxies in the group is $V_{\rm rel} = \sqrt{2}\, V_{\rm gr} = \sqrt{2}(M_{\rm gr}/M_H)^{(1-\beta)/2} V_H = 2^{5/4} V_H$. It will turn out that the interesting range of impact parameters is $p/r_H \sim (1/10)$–$(1/3)$, for which gravitational focusing is only a small effect. Taking this into account, we shall assume that the relative velocity of the perturbing galaxy is $V_p = 3V_H$.

Consider first the case that $r_p < p$, so that the perturber can be treated as a point mass. Then, treating the encounter in the impulse approximation, with the perturbing gravitational field expanded in the tidal approximation, we obtain for the magnitude of the typical velocity impulse given to the gas in its center-of-mass frame

$$\Delta v \approx \frac{2GM_p r_D}{p^2 V_p}, \qquad \left[r_D \lesssim p \lesssim \left(\frac{V_p}{V_H}\right) r_D \right] \tag{A.1}$$

(Binney & Tremaine 1987, § 7.2). The range of validity of the above expression is set by the fact that the tidal approximation breaks down for $p \lesssim r_D$, while the impulse approximation breaks down for "slow" encounters with time scale $p/V_p \gtrsim r_D/V_H$, which gives the second condition. In the case of a stellar system, the net perturbation for slow encounters is exponentially suppressed because of adiabatic invariance. We will assume that slow encounters may similarly be neglected for perturbations of a gaseous system, although this may be unduly pessimistic, in which case we will underestimate A_*. Encounters with $\Delta v \gtrsim c_s$ are assumed to cause shocking of the gas, leading to a star-formation event in which a fraction f_* of the gas is converted into stars (see Icke 1985). We will

take f_* to be constant, although a value that increases with Δv might be more realistic. Combining the condition $\Delta v > c_s$ with equation (A.1) leads to a critical impact parameter p_{cr} for inducing star formation which depends on the mass ratio $\mu \equiv M_p/M_H$:

$$\begin{aligned} p_{cr} &= p_{ad} && (\mu > \mu_1) \\ &= p_{ad}\,\mu^{1/2} && (\mu_2 < \mu < \mu_1) \\ &= 0 && (\mu < \mu_2)\,, \end{aligned} \tag{A.2a}$$

where

$$\begin{aligned} p_{ad} &= \left(\frac{V_p}{V_H}\right) r_D\,, \\ \mu_1 &= \frac{1}{2}\left(\frac{V_p}{V_H}\right)^3 \left(\frac{c_s}{V_H}\right)\left(\frac{r_D}{r_H}\right), \\ \mu_2 &= \left(\frac{V_H}{V_p}\right)^2 \mu_1\,. \end{aligned} \tag{A.2b}$$

The effective cross-section for inducing star formation is then $\sigma_*(M_p) = \pi p_{cr}^2 f_*$. This must then be integrated over the distribution of perturber masses M_p in the group. For the latter we assume $dn/dM = (n_H/M_H)(M/M_H)^\alpha$, where n_H is the nominal number density of halos of mass M_H in the group which we used in § 2.4. Our peaks theory for the mass function of dark halos gives $\alpha \approx -2$, while the Press-Schechter theory with $n = -2$ gives $\alpha = -11/6$, so we adopt $\alpha = -2$. Integrating over the mass distribution gives

$$\begin{aligned} \langle n\sigma_* \rangle &= \int_0^\infty \sigma_*(M)\,\frac{dn}{dM}\,dM \\ &= n_H\,\pi r_D^2 \left(\frac{V_p}{V_H}\right)^2 \left(\frac{\xi}{\mu_1}\right) f_*\,, \end{aligned} \tag{A.3a}$$

where

$$\xi = \ln(\mu_1/\mu_2) + 1\,. \tag{A.3b}$$

For $\mu_1/\mu_2 \ll 1$, most of the contribution to $\langle n\sigma_* \rangle$ comes from fairly low perturber masses, $\mu_2 \lesssim M_p/M_H \lesssim \mu_1$, because of the steeply rising mass function. Comparing to the way that A_* was defined in § 2.3, we find that its effective value is

$$A_* = \frac{\langle n\sigma_* \rangle}{n_H\,\pi r_T^2} \approx \frac{50}{9}\,\xi f_* \left(\frac{r_D}{r_H}\right)\left(\frac{V_H}{V_p}\right)\left(\frac{V_H}{c_s}\right). \tag{A.4}$$

Taking numerical values typical for a galaxy similar to ours, $V_H = 200$ km s^{-1}, $c_s = 10$ km s^{-1}, $V_p/V_H = 3$ and $r_D/r_H = (1/10)$, gives $A_* \approx 10 f_*$.

The analysis so far assumed that the perturber could be treated as a point mass. In an encounter with $p < r_p$, the extended nature of the mass distribution must be taken into account. We approximate this by replacing M_p in equation (A.1) by an effective mass $M_p' = M_p(p/r_p)$. The perturber radius is assumed to be given by $r_p = r_H(M_p/M_H)^\beta$. Treating the perturber as being extended, we then find that the critical impact parameter to have $\Delta v \geq c_s$ is $p_{cr} = p_{ad}$ for M_p large, and $p_{cr} \propto M_p^{1-\beta}$ for M_p small. Thus for $\beta = \frac{1}{2}$, we get the same dependence of p_{cr} on M_p as in equation (A.2) for the point mass case, except that μ_1 and μ_2 are replaced by the values μ_1' and μ_2' respectively, given by

$$\begin{aligned} \mu_1' &= \frac{1}{4}\left(\frac{V_p}{V_H}\right)^4 \left(\frac{c_s}{V_H}\right)^2, \\ \mu_2' &= \left(\frac{V_H}{V_p}\right)^2 \mu_1'\,, \end{aligned} \tag{A.5}$$

Since the actual value of p_{cr} at any given value of M_p is the smaller of those obtained by assuming $M_p' = constant$ or $M_p' \propto p$, p_{cr} is given by either the "pointlike" or "extended" formulae according as $\mu_1 > \mu_1'$ or $\mu_1 < \mu_1'$ respectively. Setting $\mu_1 = \mu_1'$ defines a critical value of V_H,

$$V_{Hc} = \frac{1}{2}\left(\frac{V_p}{V_H}\right)\left(\frac{r_H}{r_D}\right) c_s\,, \tag{A.6}$$

so that encounters are effectively pointlike or extended according as $V_H > V_{Hc}$ or $V_H < V_{Hc}$. For the numerical values assumed previously (assuming c_s, V_p/V_H and r_D/r_H to be independent of V_H), $V_{Hc} = 150$ km s^{-1}. In the extended case, we obtain

$$A_* = \frac{100}{9}\,f_*\,\xi' \left(\frac{r_D}{r_H}\right)^2 \left(\frac{V_H}{V_p}\right)^2 \left(\frac{V_H}{c_s}\right)^2 \tag{A.7}$$

where ξ' is related to μ_1'/μ_2' in the same way as ξ to μ_1/μ_2. We see that A_* scales as $A_* \propto V_H^2$ for $V_H < V_{Hc}$ and $A_* \propto V_H$ for $V_H > V_{Hc}$. In the present paper, we have taken $A_* = constant$ for simplicity, but it might be interesting to investigate the effects of this scaling with velocity in a future paper. This same calculation of collision-induced star formation can also be applied to a galaxy in which the gas is pressure supported in a roughly spherical distribution, rather than being rotationally supported in a disk. The condition $\Delta v > c_s$ still seems a reasonable one for inducing star formation, but in this case we have $c_s \sim V_H$, so a much larger perturbation is in general required.

REFERENCES

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15
Bergeron, J., & Boisse, P. 1990, preprint
Binggeli, B., Sandage, A., & Tamman, G. A. 1988, ARA&A, 26, 509
Binggeli, B., Tarenghi, M., & Sandage, A. 1990, A&A, 228, 42
Binney, J., & Tremaine, S. 1987, Galactic Dynamics (Princeton: Princeton University Press)
Bond, J. R. 1988, in The Early Universe, ed. W. G. Unruh & G. W. Semenoff (Dordrecht: Reidel), 283
Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1990, CITA preprint
Broadhurst, T. J., Ellis, R. S., & Shanks, T. 1988, MNRAS, 235, 827
Bruzual, G. 1981, Ph.D. thesis, University of California, Berkeley
———. 1983, ApJ, 273, 105
Bruzual, G., & Kron, R. G. 1980, ApJ, 241, 25
Carlberg, R. G., & Couchman, H. M. P. 1989, ApJ, 340, 47
Cole, S. 1989, Ph.D. thesis, Cambridge University
Cole, S. C. 1991, ApJ, 367, 45
Cole, S., & Kaiser, N. 1989, MNRAS, 237, 1127
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. N. Lasenby (Dordrecht: Kluwer), 1
Cowie, L. L., Gardner, J. P., Wainscoat, R. J., & Hodapp, K. W. 1991, preprint
Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371
de Vaucouleurs, G. 1977, in The Evolution of Galaxies and Stellar Populations, ed. B. M. Tinsley & R. B. Larson (New Haven: Yale University Observatory), 43
de Vaucouleurs, G., de Vaucouleurs, A., & Corwin, H. G. 1976, Second Reference Catalogue of Bright Galaxies (Austin: University of Texas Press)
Dekel, A., & Silk, J. 1986, ApJ, 303, 39
Dressler, A. 1980, ApJ, 236, 351
Efstathiou, G., Ellis, R. S., & Peterson, B. A. 1988, MNRAS, 232, 431
Fall, S. M., & Rees, M. J. 1985, ApJ, 298, 18
Frenk, C. S., White, S. D. M., Davis, M., & Efstathiou, G. 1988, ApJ, 327, 507
Gaetz, T., & Salpeter, E. E. 1983, ApJS, 52, 155
Guiderdoni, B., & Rocca-Volmerange, B. 1987, A&A, 186, 1
———. 1990, A&A, 227, 362
Guhathakurta, P., Tyson, J. A., & Majewski, S. R. 1990, ApJ, 357, L9
Hernquist, L. 1989, Nature, 340, 687
Huchra, J. P. 1977a, ApJS, 35, 171
———. 1977b, ApJ, 217, 928
Icke, V. 1985, A&A, 144, 115
Impey, C., & Bothun, G. 1989, ApJ, 341, 89
Jarvis, J. F., & Tyson, J. A. 1981, AJ, 83, 1549
Kawano, L., Schramm, D., & Steigman, G. 1988, ApJ, 330, L1
Kennicutt, R. C., Keel, W. C., van der Hulst, J. M., Hummel, E., & Roettiger, K. A. 1987, AJ, 93, 1011
King, C. R., & Ellis, R. S. 1985, AJ, 288, 456
Koo, D. C. 1990, in The Evolution of the Universe of Galaxies: The Edwin Hubble Centennial Symposium, ed. R. G. Kron (ASP Conf. Ser.), 268
Kormendy, J. 1985, ApJ, 295, 73
Lacey, C. G., Silk, J., Guiderdoni, B., & Rocca-Volmerange, B. 1991, in preparation (Paper II)
Larson, R. B. 1974, MNRAS, 169, 229
Larson, R. B., & Tinsley, B. M. 1978, ApJ, 219, 46
Lonsdale, C. J., Persson, S. E., & Matthews, K. 1984, ApJ, 287, 95
Maddox, S. M., Sutherland, W. J., Efstathiou, G., Loveday, J., & Peterson, B. A. 1990, MNRAS, 247, 1P
Noguchi, M. 1988, A&A, 203, 259
Noguchi, M., & Ishibashi, S. 1986, MNRAS, 219, 305
Peacock, J. A., & Heavens, A. F. 1990, MNRAS, 243, 133
Peebles, P. J. E. 1980, The Large-Scale Structure of the Universe (Princeton: Princeton University Press)
Postman, M., & Geller, M. J. 1984, ApJ, 281, 95
Press, W. H., & Schechter, P. 1974, ApJ, 187, 425
Rees, M. J., & Ostriker, J. P. 1977, MNRAS, 179, 54
Rocca-Vomerange, B., & Guiderdoni, B. 1990, MNRAS, 247, 166
Sandage, A., Binggeli, B., & Tamman, G. A. 1985, AJ, 90, 1759
Saunders, W., Rowan-Robinson, M., Lawrence, A., Efstathiou, G., Kaiser, N., Ellis, R. S., & Frenk, C. S. 1990, MNRAS, 242, 318
Scalo, J. M. 1986, Fundamentals of Cosmic Physics, 11, 1
Schaeffer, R., & Silk, J. 1988, ApJ, 332, 1
Silk, J. 1977, ApJ, 211, 638
Songaila, A., Cowie, L. L., & Lilly, S. J. 1990, ApJ 348, 371
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
Tinsley, B. M. 1980, ApJ, 241, 41
Toomre, A. 1977, in Evolution of Galaxies and Stellar Populations, ed. B. M. Tinsley & R. B. Larson (New Haven: Yale University Observatory), 401
Toomre, A. 1981, in The Structure and Evolution of Normal Galaxies, ed. S. M. Fall & D. Lynden-Bell (Cambridge: Cambridge Univ. Press), 111
Tyson, J. A. 1988, AJ, 96, 1
Weinberg, S. 1972, Gravitation and Cosmology (New York: Wiley)
White, S. D. M. 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
White, S. D. M., & Frenk, C. S. 1991, ApJ, in press
White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341
Wolfe, A. M. 1988, in QSO Absorption Lines, ed. J. C. Blades, D. Turnshek, & C. Norman (Cambridge: Cambridge Univ. Press), 297