# TESTING ORIGINS OF THE HUBBLE SEQUENCE

AUGUST E. EVRARD
Institute of Astronomy, University of Cambridge, and Department of Astronomy, University of California, Berkeley
*Received 1988 August 5; accepted 1988 October 27*

## ABSTRACT

An analytic formalism which allows calculation of separate mass functions for disk, spheroid, and dwarf galaxies in hierarchical clustering models is applied to galaxy formation in the cold dark matter model. The method employs a probabilistic argument which requires theoretical prejudice regarding what physical or statistical criteria broadly define the Hubble types.

Dwarfs are assumed to form within haloes with virial temperatures below a critical value $T_c \simeq 10^{5.4}$ K. Two sets of criteria to differentiate ellipticals from spirals are examined. One model assumes that ellipticals cooled and formed stars rapidly while spirals cooled on longer, nearly hydrostatic time scales. This model fails to reproduce simple observed characteristics of the galaxy distribution: spirals are predicted to be rarer and typically $\sim 5$ times more massive than ellipticals. A second model assumes that ellipticals formed from rare ($\sim 3\,\sigma$) density fluctuations, while spirals arose from more common ($\sim 2\,\sigma$ to $3\,\sigma$) perturbations. These ideas, when tuned to give the proper global abundances of spiral and elliptical galaxies, produce mass functions with shapes in good qualitative agreement with observations. In addition, the model exhibits a "natural" morphological bias which allows enhancement of the elliptical fraction in clusters to levels in agreement with the observed morphology-density relation.

In searching for a physical interpretation of the latter model, a conjecture is made that the transition from the spheroid-forming to disk-forming regimes occurs at a critical *pressure* which is *independent of environment*.

*Subject headings:* dark matter — galaxies: formation — galaxies: structure

## I. INTRODUCTION

Nearly 30 years ago, Eggen, Lynden-Bell, and Sandage (1962) presented conclusive evidence that the birth of the Milky Way involved collapse of a protogalactic object with spatial dimension $\sim 100$ kpc over a time scale of $\sim 10^9$ yr. Since that time, the complexities of the galaxy formation process have been slow to unravel. In particular, an explanation of the physical processes which govern the ultimate morphology of a collapsing protogalaxy—the "keys to the Hubble sequence"—remain elusive.

There are a few well-established ideas for what physics broadly separates the Hubble types. For example, it has long been held that gravitational binding energy plays an important role in distinguishing dwarfs from bright galaxies. The early simulations of Larson (1969, 1974, 1975) exhibited the interplay between gas dissipation, star formation rate, and supernova feedback. The idea that collective supernova explosions could drive a galactic wind (Mathews and Baker 1971) and unbind the gas in smaller galaxies has received recent attention (Dekel and Silk 1986; Yoshii and Arimoto 1987). The constructed models can reproduce the characteristic low metallicities and surface brightnesses of dwarfs when the initial binding energy of the galaxy is insufficient to contain gas heated by early generations of star formation. The implication is that normal galaxies (where the term *normal* in this paper is used to denote all *nondwarf* elliptical, lenticular, and spiral galaxies) are systems with virial temperature above some critical value $T_c$.

Larson (1975) further suggested a physically compelling distinction between the formation of disk and the formation of spheroidal systems. He argued for the star formation rate as the critical element in determining whether or not a significant disk structure could be formed. If stars were formed rapidly during the collapse of a protogalaxy, consuming a large gas

fraction in the process, the collapse of the galaxy would involve mainly violent relaxation of a cold stellar component. N-body simulations have shown that such collapse can reproduce the bulk of observed luminosity profiles in elliptical galaxies (Van Albada 1982; McGlynn 1984). Any remaining, presumably enriched gas would settle to the galactic center, with further star formation serving to enhance the central surface brightness and produce a metallicity gradient in the sense observed (Larson 1974).

A second, empirically motivated distinction between spirals and ellipticals was suggested by Faber (1982) and expanded by Blumenthal *et al.* (1984). Faber pointed out that the observed Tully-Fisher and Faber-Jackson relations ($L \propto \sigma^p$, $p \simeq 4$) would arise naturally in clustering from an initial scale-free fluctuation spectrum with spectral index $n = -2$ [$P(k) \propto k^n$]. Blumenthal *et al.* noted that this is approximately the slope of the cold dark matter spectrum on galactic scales. They also pointed out that the observed narrow scatter and trend of decreasing velocity from early to late systems at a given luminosity would require that ellipticals arise from $\sim 3\,\sigma$ fluctuations while spirals are perhaps $\sim 2\,\sigma$ objects (as depicted in their Fig. 4). Further impetus for this model comes from biased galaxy formation—the idea that galaxies form at $\gtrsim 2.5\,\sigma$ peaks in the density field (Bardeen *et al.* 1986). This biasing is *necessary* in an $\Omega = 1$ cold dark matter model if it is to reproduce the observed kinematics of the galaxy distribution on scales of $1$–$10\,h_{50}^{-1}$ Mpc (Davis *et al.* 1985).

This paper presents a method of determining mass functions for morphological types in order to test the ideas outlined above. The motivation is to make quantitative predictions to compare with observations. Since quantitative predictions require an assumed cosmogony, we consider the specific case of adiabatic fluctuations in a cold dark matter (CDM) dominated universe with cosmological parameters $\Omega = 1$, $\Omega_b = 0.1$, and

26

$H_0 = 50h_{50}$ km s$^{-1}$ Mpc$^{-1}$. Ironically, it is found that the physically appealing disk/spheroid distinction of Larson fails to match some of the crudest global properties of the galaxy distribution. However, the distinction by Faber, which lacks a sound physical basis, is capable of reproducing several interesting observations of the bright galaxy distribution.

Some of the observed properties of the galaxy distribution which will be used as benchmarks for the models are briefly discussed in § II. The basic physics of galaxy formation from gravitational instability in a cold dark matter universe is outlined in § III. The formalism for constructing mass functions for the morphological types is presented in § IV, and global population results are given. Section V discusses the morphological bias expected in clusters. This leads to a possible physical interpretation of a *pressure-dependent* morphology, a subject addressed in the final section which presents a summary and discussion of the results.

## II. THE OBSERVED GALAXY DISTRIBUTION

The Hubble sequence as illustrated, for example, by Sandage (1961) displays a wide range of beautiful morphologies. A detailed classification system has been developed to accommodate subtle differences between the types. *For the purpose of this paper, we consider galaxies as falling into two basic populations—"normal" and "dwarf"* (Dw)—*with normal galaxies being further subdivided into spheroid* (E) *and disk* (Sp) *systems.* Observationally, the distinction between normal and dwarf galaxies is mainly one of surface brightness (Kormendy 1987). Theoretically, the difference is one of binding energy or virial temperature. No attempt is made here to distinguish between gas-rich and gas poor dwarfs. The presence or absence of a significant stellar disk is used to distinguish Sp from E galaxies. Again, galactic gas content is ignored. Therefore, since lenticulars (S0's) possess significant disks, they will be classed in the Sp category (Larson, Tinsley, and Caldwell 1980). Hereafter, unless otherwise explicitly stated, the term Sp will refer to all nondwarf disk systems, both spiral and lenticular.

The detailed study by Sandage, Binggeli, and Tammann (1985) of galaxies in the Virgo Cluster has prompted a reinterpretation of the standard Schechter luminosity function (Binggeli 1987). When the luminosity function (LF) of all galaxies is decomposed into type-dependent LFs, some interesting regularities are uncovered. The LFs of elliptical, lenticular, and spiral galaxies are similar in shape, being roughly bell-shaped with a width of about two decades in luminosity ($-22 \lesssim M_{B_T} \lesssim -17$). The brightest ellipticals are slightly brighter than the brightest lenticulars, which in turn are somewhat brighter than the most luminous spirals. At $M_{B_T} \simeq -18$ ($L \simeq 3 \times 10^9 L_\odot$), dwarf galaxies begin to appear. They dominate the counts within a magnitude fainter; in fact, the dE galaxies appear to comprise 70% of the total cluster population down to the completeness limit of $M_{B_T} \simeq -14$.

Binggeli (1987) has presented evidence, by comparing the Virgo data with both field galaxies and galaxies in the Coma Cluster, that the type-specific galaxy LFs are *independent of environment.* Only the relative proportions of the Hubble types vary with local density. In the field, disk systems tend to outnumber ellipticals (Dressler 1980; Postman and Geller 1984) by a global ratio

$$\left(\frac{\mathrm{Sp}}{\mathrm{E}}\right)_{\mathrm{glob}} \simeq 9 , \tag{1}$$

whereas in clusters of galaxies, roughly twice the average number of ellipticals are found within a cluster optical radius (Oemler 1974; Dressler 1980):

$$\left(\frac{\mathrm{Sp}}{\mathrm{E}}\right)_{\mathrm{clust}} \simeq 4 . \tag{2}$$

Enhancement of the elliptical population does not occur only in massive rich clusters. Hickson, Kindl, and Huchra (1988) find 20% ellipticals in compact groups containing typically only four or five galaxies. Thus, the processes governing morphology appear to be sensitive to *local* conditions, where local means on the scale of a few galaxies. This idea is strengthened by the strong dependence of morphological type on local galaxy density (Dressler 1980; Postman and Geller 1984).

Although the evidence for universality of the type-specific LFs is still tentative, if it is confirmed the implication would be that the processes which shape galaxy morphology do operate in the same manner in both the cluster and the field, i.e., locally. However, the frequencies with which they operate may be modulated by the large-scale density. The increased elliptical fraction within rich clusters would then be interpreted as due to collective enhancement of the local effects which govern spheroid formation—a form of "natural morphological biasing" analogous to the original idea of biasing proposed by Kaiser (1984).

Looking at the total galaxy population, the space number density per unit luminosity of galaxies with luminosity $L$ is well parameterized by the Schechter function (Yahil, Sandage, and Tammann 1980; Kirshner *et al.* 1983; Geller and Huchra 1983),

$$\Phi(L)dL = \Phi_*(L/L_*)^{-\alpha} \exp{(-L/L_*)}dL , \tag{3}$$

with $L_* \simeq 3 \times 10^{10}h_{50}^{-2} L_\odot$ and faint end slope $\alpha \simeq -1.25$. Most normal galaxies are brighter than $M_{B_T} \simeq -17$ ($L \gtrsim L_*/30$), with dwarfs dominating fainter counts. The fact that normal galaxies appear to have a low-luminosity cutoff means that we can estimate the *total* number density of normal galaxies by integrating the Schechter function down to their limiting luminosity, taken as $M_{B_T} = -17$. Although values of the parameters $\alpha$, $L_*$, and $\Phi_*$ vary somewhat from sample to sample, the integration is not particularly sensitive; a mean value of the three samples listed above is

$$n_{\mathrm{gal}} \simeq 0.008h_{50}^3 \ \mathrm{Mpc}^{-3} . \tag{4}$$

We can estimate a "characteristic" total mass to associate with normal galaxies by combining the luminosity $L_*$ with a typical mass-to-light ratio within the luminous extent of galaxies $M/L \simeq 10h_{50} M_\odot/L_\odot$. The result is a mass per galaxy

$$M_{\mathrm{gal}} \simeq 3 \times 10^{11}h_{50}^{-1} M_\odot . \tag{5}$$

The above two quantities yield an estimate of the fraction of mass in an $\Omega = 1$ universe presently associated with normal galaxies

$$\Omega_{\mathrm{gal}} = \frac{n_{\mathrm{gal}} M_{\mathrm{gal}}}{\rho_0} \simeq 0.03 . \tag{6}$$

This small fraction could reflect the fact that galaxy formation, or rather *concentrated star formation*, is an inefficient process—perhaps lending credence to biased galaxy formation models. However, the strength of this statement is limited by the uncertainties of associating mass with luminosity.

### III. MODELING GALAXY FORMATION PHYSICS

We are assuming that structure in the universe formed via gravitational instability of density perturbations imprinted in the early universe. In particular, we consider perturbations arising in a universe dominated by cold dark matter (Peebles 1982; Bond and Efstathiou 1984; Blumenthal *et al.* 1984). Galaxy formation in this model is envisioned as a two-stage process (White and Rees 1978) whereby dissipationless clustering of the dark matter creates halo potential wells into which the baryons can subsequently collapse and dissipate through radiative cooling to form stars.

To make galaxies "by hand" (without detailed three-dimensional numerical modeling) requires several ingredients. One is a way of propagating initial density perturbations forward in time to yield both the spatial number density and characteristic internal properties (densities, temperatures, etc.) of collapsed systems at a given mass. The Press-Schecter (1974) formula is used to provide number densities $n(M, z)$ while a modified spherical "top-hat" prescription (Gunn and Gott 1972; Gunn 1977) is used to determine characteristic internal properties. Another necessary ingredient is a criterion to judge which collapsed systems form galaxies and which do not. The ability to cool on a physically interesting time scale is the classic criterion (Rees and Ostriker 1977; Silk 1977). More complicated critiera determining morphology will be addressed in the next section.

#### a) Gravity

Consider at some early postrecombination epoch $z_i$ a spherical density perturbation with a mean overdensity profile $\delta_i(r) \equiv [M(r) - \bar{M}(r)]/\bar{M}(r)$. The top-hat prescription assumes that a mass shell $M$ will expand to a turnaround radius $r_{\rm turn}(M) \propto \delta^{-1}(M)M^{1/3}$ and collapse by a factor of 2 before virializing through shock heating or violent relaxation. Straightforward calculation (see the Appendix) leads to the final virialized characteristics of the system as a function of mass shell,

$$1 + z_{\rm vir}(M) = \delta_0(M)/1.68 , \tag{7a}$$

$$r_{\rm vir}(M) = 450\,\delta_0^{-1}(M)M_{12}^{1/3}h_{50}^{-2/3}\ {\rm kpc} , \tag{7b}$$

$$D_{\rm vir}(M) \equiv \rho_{\rm vir}(M)/\rho_0 = 37.3\beta^{-1}(M)\delta_0^3(M) , \tag{7c}$$

$$T_{\rm vir}(M) = 2.31 \times 10^5\,\delta_0(M)M_{12}^{2/3}h_{50}^{2/3}\ {\rm K} . \tag{7d}$$

Here $\delta_0(M) = (1 + z_i)\delta_i(M)$ is the fluctuation amplitude evolved by linear theory to the present day, $M_{12}$ is mass in units of $10^{12}\,M_\odot$, $z_{\rm vir}$ is the collapse redshift, $r_{\rm vir}$ is the physical virial radius, $D_{\rm vir}(M)$ is the local density contrast of the shell relative to the present cosmic background density $\rho_0 \equiv 3H_0^2/8\pi G$, and $T_{\rm vir}$ is the virial temperature. The quantity $\beta(M) \equiv 1 - 3d \log \delta/d \log M$ enters into the measure of the local shell density because the layering of shells depends on the slope of the initial overdensity profile $\delta(M)$.

This simplified behavior of mass shells is actually in very good agreement with what happens in detailed self-similar collapse models (Bertschinger 1983). What also emerges from the self-similar models is that the postcollapse profiles of a shock-heated $\gamma = 5/3$ gas are extremely similar to the postcollapse profiles of collisionless mass shells. This allows the simplifying assumption that baryons and dark matter will collapse in similar ways, until cooling removes pressure support from the gas. This lack of segregation in the adiabatic regime is also seen

in three-dimensional hydrodynamic calculations (Evrard and Davis 1983; Evrard 1988).

Perhaps a significant leap of intuition is required to assume that the results above obtained for a given spherically symmetric perturbation are applicable to arbitrary three-dimensional structures with mean overdensity $\delta(M)$. That is, if we draw a sphere around a random point in the universe containing mass $M$ which happens to be overdense by an amount $\delta(M)$, the mass on that shell will behave like the spherically symmetric top-hat model above. This is unlikely to be true in the general case, since the material within a shell around a randomly chosen point is rarely close to spherically symmetric. In particular, even if the volume is sufficiently overdense to collapse by a certain epoch, tidal forces may shear out the mass within the region before it has a chance to collapse on itself. However, the counterargument is that galaxies do not form at random points in space; they tend to form around peaks in the initial density field. The behavior of the mass distribution around a peak may follow more closely the evolution above. The fact that equipotential contours are more spherical than equidensity contours may justify the spherical approximation for peaks with surrounding mass distributions of a moderate ellipticity. This, coupled with the fact that "object finding" algorithms operating in the non-linear regime tend to find matter associated with peaks in the initial density field (Frenk *et al.* 1988), might lead us reasonably to expect the top-hat predictions to apply to properties of objects collapsed from initial peaks.

We can then forge ahead and apply the above prescription to a spectrum of fluctuations and replace $\delta_0(M) = \nu\sigma_0(M)$, where $\sigma_0(M)$ is the rms present-day fluctuation amplitude on a (top-hat–filtered) mass scale $M$ (Peebles 1980). Two normalizations of the CDM spectrum listed in Table 1 are considered, one favored by the $N$-body work of Davis *et al.* (1985) (the DEFW normalization), the other arising from statistical work of Bardeen *et al.* (1986, hereafter BBKS). The bias parameter $b$ measures the enhancement of galaxy fluctuations relative to the underlying mass on a scale of $16h_{50}^{-1}$ Mpc. In this sense, the DEFW normalization is "more biased" than BBKS by about 50%.

Figure 1 shows the virial properties of $\nu\sigma$ cold dark matter perturbations with both normalizations. Note the scalings with $\nu$. With the BBKS normalization, 3 $\sigma$ objects with galactic halo masses ($M_{12} \simeq 1$) would collapse at a redshift $z = 5$, have radii $r \simeq 50$ kpc, a virial temperature $T = 2 \times 10^6$ K corresponding to a rotation speed $v_{\rm rot} = 230$ km s$^{-1}$, and a density contrast of about $2 \times 10^4$. These numbers are consistent with properties of observed galactic halos and would imply dissipation and collapse by a factor $\sim 5$ for the baryonic component in order to agree with optical radii of $\sim 10$ kpc and density contrasts of $\sim 10^6$ (corresponding to $\sim 1$ proton cm$^{-3}$). The CDM fluctuations on galactic scales follow $\delta_0(M) \propto \nu M^{-1/6}$, which, combined with equation (7d), yields $T \sim v^2 \propto \nu M^{1/2}$. This leads to

TABLE 1

NORMALIZATION PARAMETERS

| Model | $\sigma_0$(16 Mpc)[a] | $b$ |
|---|---|---|
| BBKS ....... | 0.59 | 1.7 |
| DEFW ...... | 0.40 | 2.5 |

[a] $\sigma_0(x) \equiv \langle(\delta\rho/\rho)^2\rangle^{1/2}$ on scale $x$ at present epoch.
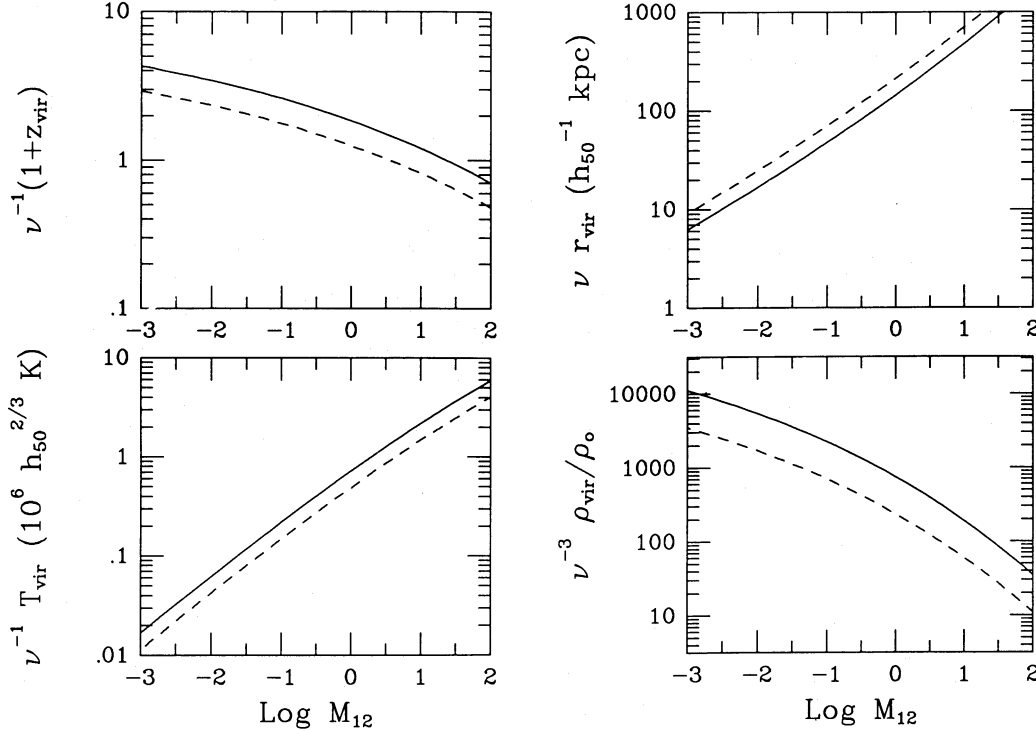
FIG. 1.—Properties of cold dark matter perturbations evolved to virial equilibrium using the spherical top-hat prescription (eq. [7]), with $\delta_0(M) = v\sigma_0(M)$. The solid and dashed lines show $1\,\sigma$ objects with the BBKS and DEFW normalizations, respectively. Scalings with $v$ are shown on the axis labels.

the observed form of the luminosity-velocity relations $L \propto M \propto v^{-2}v^4$ under the assumption of a constant $M/L$ ratio. However, the narrow scatter in luminosities at a given velocity for each Hubble type implies a narrow band in $v$, the width in $\log v$ being roughly $\Delta \log v \sim 0.5\Delta \log L \simeq 0.1$ if there is 0.5 mag scatter in luminosity at a given velocity (Kaiser 1988).

### b) Number Density of Collapsed Halos

The Press-Schechter (1974) formula yields the spatial number density of halos of mass $M$ collapsed at redshift $z$:

$$n(M, z)d \log M = -\sqrt{\frac{2}{\pi}}\frac{\rho_0}{M}\frac{d \log \sigma}{d \log M}\,v_z(M)$$
$$\times \exp\left[-\frac{v_z^2(M)}{2}\right]d \log M\,, \quad (8)$$

where $v_z(M) = 1.68(1 + z)/\sigma_0(M)$ is the normalized fluctuation amplitude reaching virial equilibrium at redshift $z$. The basis of the above formula is a striaghtforward counting argument. Given an initial Gaussian distribution of overdensities with variance $\sigma(M)$, one examines how the the fraction of collapsed mass in the universe changes in going from $M$ to $M + \delta M$. The *Ansatz* that the change in this fraction is due solely to objects of mass $M$ produces the above formula.

There has been much lingering theoretical uncertainty about the applicability of the Press-Schechter formula. One particular problem is whether the differentiation *Ansatz* properly accounts for the erasure of substructure in hierarchical clustering—the "cloud-in-cloud" problem. Another problem is that a factor of 2 "fudge" was introduced into the original

derivation to account for the accretion of underdense regions onto overdense objects. It is not entirely clear that this factor is correct or, indeed, that any accretion factor is even warranted. Luckily, these difficult questions can be circumvented by using *empirical* evidence that the formula works. A recent study by Efstathiou *et al.* (1988) of self-similar clustering in $N$-body experiments shows that the Press-Schechter formula predicts the number densities of collapsed objects extremely well for a wide range of initial power spectra. We therefore assume that equation (8) correctly predicts the mass function of halos at a given redshift.

### c) Cooling

The ability of the gas in a halo to cool on an interesting time scale is the principal physical condition for forming a galaxy. Other time scales of interest are the dynamical time of the halo and the age of the universe. These characteristic time scales can be determined from the estimates of the physical properties of collapsed shells given above. For example, the cooling time is given by

$$\tau_{\rm cool}(M) = 43\Omega_b^{-1}\beta(M)\delta_0^{-2}(M)M_{12}^{2/3}\tilde{\Lambda}^{-1}(T_{\rm vir})h_{50}^{4/3}\,{\rm Gyr}\,, \quad (9)$$

where $\tilde{\Lambda}(T) = \Lambda(T)/(\mu m_p) \simeq 10^{24}\Lambda(T)$, with $\Lambda(T)$ the total cooling rate in cgs units, $\mu$ the mean molecular weight and $m_p$ the proton mass.

We define the dynamical time as the virial crossing time of a shell:

$$\tau_{\rm dyn}(M) \equiv \frac{2r_{\rm vir}}{v_{\rm vir}} = 9.21\delta_0^{-3/2}(M)h_{50}^{-1}\,{\rm Gyr}\,. \quad (10)$$

The classic criteria for galaxy formation are the ratio of the cooling to the dynamical time,

$$\frac{\tau_{\rm cool}(M)}{\tau_{\rm dyn}(M)} = 4.7\Omega_b^{-1}\beta(M)\delta_0^{-1/2}(M)M_{12}^{2/3}\tilde{\Lambda}^{-1}(T_{\rm vir})h_{50}^{1/3} , \quad (11)$$

and the ratio of the cooling time to the Hubble time,

$$\frac{\tau_{\rm cool}(M)}{\tau_{\rm H}} = 3.2\Omega_b^{-1}\beta(M)\delta_0^{-2}(M)M_{12}^{2/3}\tilde{\Lambda}^{-1}(T_{\rm vir})h_{50}^{-1/3} . \quad (12)$$

Lines where the above critical ratios are unity define regions in the n-T diagram shown in Figure 2. Cooling from a primordial gas is assumed; the cooling curve $\Lambda(T)$ is taken from Fall and Rees (1985). Compton cooling is also included. Note that the abscissa is inverted and shows the overdensity $D \equiv n/n_0$, with $n_0 = 5 \times 10^{-7}(\Omega_b/0.1)h_{50}^2$ cm$^{-3}$ the present mean background particle number density. This density is directly related to the collapse epoch shown on the right-hand axis. The top axis shows the one-dimensional velocity dispersion $v_{1d}^2 = (kT/\mu m_p)$ with $\mu = 0.6$. The heavy solid line denotes values where $\tau_{\rm cool}/\tau_{\rm dyn} = 1$, the heavy dashed line $\tau_{\rm cool}/\tau_{\rm H} = 1$. The diagonal dotted lines are lines of constant mass $M \propto (T^3/\rho)^{1/2}$. Evolved 1 $\sigma$, 2 $\sigma$, and 3 $\sigma$ CDM perturbations with the BBKS normalization are shown as long dashed lines. Perturbations that lie in region I have too small an amplitude to have collapsed by the present. Those in region II have collapsed and can cool in a dynamical time. The collapsed systems in region III have a cooling time intermediate between a dynamical time and a Hubble time, while those in region IV cannot have cooled by today. Dividing up this diagram among the simple Hubble types Dw, E, and Sp is the subject addressed in the next section.

### IV. MASS FUNCTIONS FOR DIFFERENT MORPHOLOGICAL TYPES

The n-T plane presented in the last section is now used as a springboard for determining mass functions for the morphological types Dw, E, and Sp. The basic idea is to divide the plane up among the morphologies according to some well-defined physical criteria expressed as functions of n and T. The areas demarcated in the n-T plane can then be mapped onto the v-M plane, where a probabilistic argument, based on the fact that the initial values of v are normally distributed, is used to determine the fraction of collapsed objects falling within each morphological class. These fractions, when combined with the Press-Schechter formula for the number density of collapsed halos at a given redshift, produce mass functions for the individual Hubble types.

#### a) Physical Criteria for the Morphologies

The long-established idea (Larson 1974) for the origin of dwarf galaxies—that they represent objects with binding energy too small to retain gas after an initial burst of star formation activity—is used to differentiate normal galaxies from Dw's. In particular, we use the estimate of the critical temperature by Dekel and Silk (1986), $T_c = 10^{5.4}$ K ($v_{1d} = 60$ km s$^{-1}$), below which dwarf galaxies would form. Two models distinguishing spheroids (E's) from disks (Sp's) will be investigated. The element common to both models is that normal galaxies have virial temperature greater than the dwarf value, $T > T_c$.
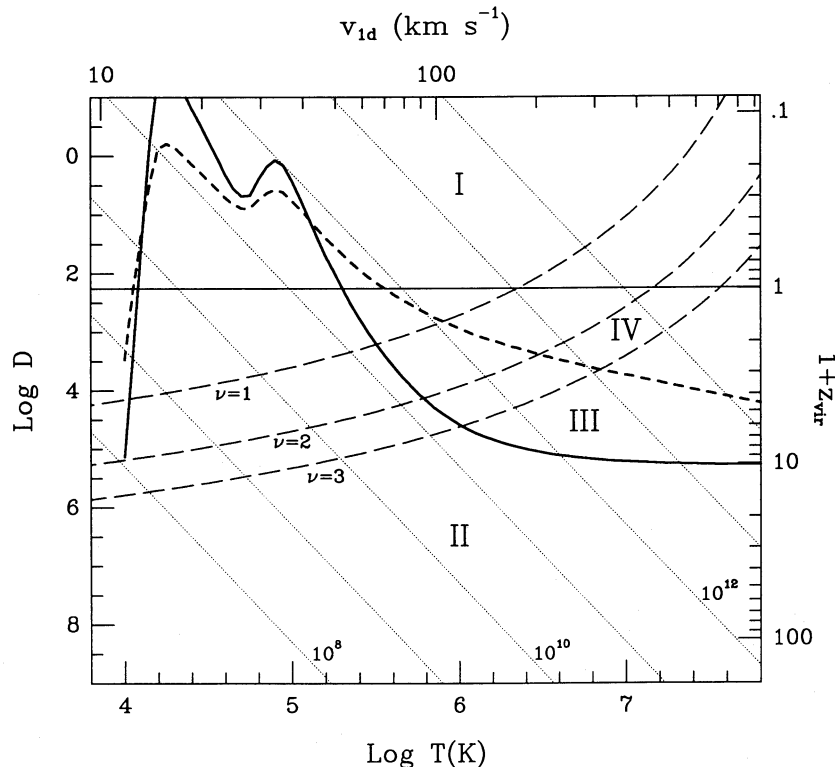


FIG. 2.—The n-T plane containing the essentials of galaxy formation physics. Diagonal dotted lines are lines of constant mass; the long-dashed lines are CDM perturbations with the BBKS normalization and values of v indicated. Perturbations below the solid horizontal line have collapsed by today. Those below the heavy solid and dotted curves have $\tau_{\rm cool} < \tau_{\rm dyn}$ and $\tau_{\rm cool} < \tau_{\rm H}$, respectively. The classic interpretation of this diagram is that galaxies form in region II, where cooling is rapid; the transition from galaxies to groups occurs in the intermediate cooling region III; and groups and clusters fall in region IV, where cooling is inefficient.

TABLE 2

MODEL CONVENTIONS

| Model | Label | Dw | E | Sp |
|---|---|---|---|---|
| $\tau$-ratio ............. | $\tau_r$ | $T < T_c$ | $\tau_{cool} < \tau_{dyn}$ | $\tau_{dyn} < \tau_{cool} < \tau_H$ |
| $v$-threshold ....... | $v_t^I$ | $T < T_c$ | $v > v_E$, $\tau_{cool} < \tau_{dyn}$ | $v_{Sp} < v < v_E$, $\tau_{cool} < \tau_{dyn}$ |
|  | $v_t^{II}$ | $T < T_c$ | $v > v_E$, $\tau_{cool} < \tau_H$ | $v_{Sp} < v < v_E$, $\tau_{cool} < \tau_H$ |

The first model attempts to encapsulate the physical ideas of Larson outlined in § I with an additional assumption linking star formation efficiency with rate of gas cooling. In this model, ellipticals are taken to be objects which can cool rapidly, $\tau_{cool} < \tau_{dyn}$, whereas spirals are identified with objects having $\tau_{dyn} < \tau_{cool} < \tau_H$. The implicit assumption is that rapid cooling will lead to fragmentation and star formation on roughly a dynamical time, whereas slow cooling implies slow star formation and hence significant amounts of gas left over to quasi-statically settle onto a disk after the underlying dark halo has had a few dynamical times to become stable. For brevity, let us call this collection of ideas the $\tau$-ratio model of galactic morphology.

An alternative, less physically intuitive differentiation between ellipticals and spirals arises from the empirical $L \propto v^4$ relations as noted by Blumenthal et al. (1984). With constant $L/M$ assumed, the spherical collapse model predicts $v^2 L \propto v^4$. The observed narrow scatter in the $L$-$v$ the relation and the trend of decreasing velocity from spheroid to disk systems at a given luminosity thus provokes a model where morphology is governed by the normalized perturbation height $v$. That is, ellipticals are systems with $v > v_E$, and spirals result from perturbations with $v_{Sp} < v < v_E$. Note that this model assumes that gas dissipation and star formation are kind enough to produce a stellar component with velocity dispersion or rotation speed following (to within a constant factor) the character-

istic velocity of the parent halo. The values of $v_E$ and $v_{Sp}$ are left as free parameters which will be adjusted to reproduce the correct global abundance of spheroid and disk galaxies. Again, some cooling criterion must apply to form galaxies, and the two possibilities—cooling within a dynamical time and cooling within a Hubble time—are considered separately. This collection of criteria will be referred to as the $v$-threshold model, with two variations arising from the two cooling conditions.

Table 2 summarizes the models and labeling conventions followed in the subsequent sections. Their schematic representations on the $n$-$T$ plane are shown in Figure 3.

### b) Formalism

We start with the Press-Schechter estimate for number densities $n(M, z)$ of halos as a function of mass at a given redshift $z$ (eq. [8]). The aim is now to partition this population of halos among the different morphologies. This is done by projecting the halos onto the $n$-$T$ diagram and counting the fraction of collapsed haloes which fall in the regions of Figure 3 outlining the different morphological types. The idea is to transform the lines demarcating the morphological classes from the $n$-$T$ plane to the $v$-$M$ plane, and use the fact that the initial distribution of $v$ is normal. The mapping of Figure 3 onto the $v$-$M$ plane is shown in Figure 4.

The number density of galaxies with mass $M$ of a given type $X$ is then taken to be the number density of halos times the
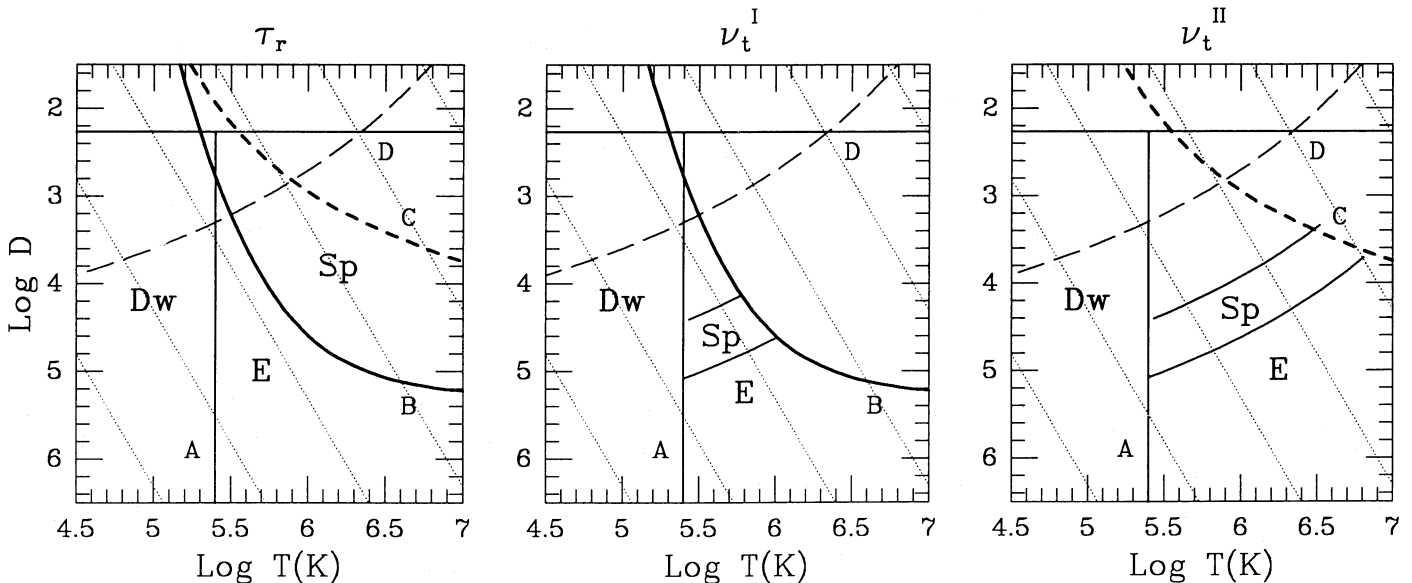


FIG. 3.—Division of the $n$-$T$ plane among the basic Hubble types in each of the models: (a) $\tau$-ratio; (b) $v$-threshold with $\tau_{cool} < \tau_{dyn}$; (c) $v$-threshold with $\tau_{cool} < \tau_H$. The labeled lines are (A) critical dwarf temperature $T_c$; (B) $\tau_{cool} = \tau_{dyn}$; (C) $\tau_{cool} = \tau_H$; (D) $z_{vir} = 0$. The long-dashed line is the locus of 1 $\sigma$ BBKS perturbations. Dotted lines are lines of constant mass. The regions labeled Dw, E, and Sp denote the areas partitioned to dwarf, spheroidal, and disk galaxies, respectively.
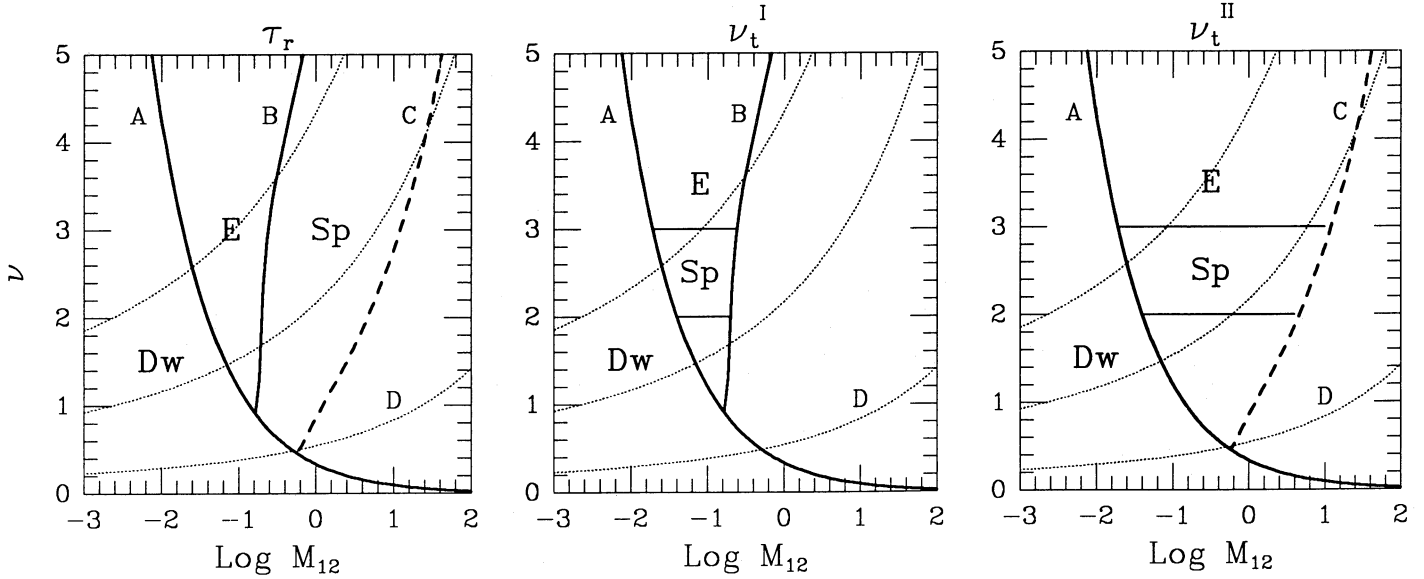
FIG. 4.—Mapping of Fig. 3 onto the $v$-$M$ plane. Labeling is the same as in Fig. 3. The upper, middle, and lower dotted lines show values of $v$ with $z_{vir} = 7$, 3, and 0, respectively. The fact that the spectrum of initial perturbations at each mass $M$ is normal in $v$ allows calculation of mass functions for the individual morphologies.

fraction of collapsed perturbations which had values of $v$ appropriate for that particular galaxy type:

$$n_X(M, z) = n(M, z)f_X(M) . \qquad (13)$$

For example, the criterion for forming a dwarf galaxy, $T < T_c$, translates for a particular mass $M$ to an upper bound on overdensity, $\delta_0(M) < \delta_{Dw}(M) = (T_c/2.3 \times 10^5)M_{12}^{-2/3}$, which in turn determines a critical range of $v(M) < v_{Dw}(M) = \delta_{Dw}(M)/\sigma_0(M)$. The fraction of collapsed halos satisfying this critierion will be

$$f_{Dw}(M) = 1 - \max\left\{1, \frac{\mathrm{erfc}\,[v_{Dw}(M)]}{\mathrm{erfc}\,[v_z(M)]}\right\}, \qquad (14)$$

where

$$\mathrm{erfc}\,(x) = (2\pi)^{-1/2}\int_x^\infty dy \exp\,(-y^2/2)$$

is the complementary error function and $v_z(M) = 1.68$ $(1 + z)/\sigma_0(M)$ is the normalized perturbation height just collapsed at redshift $z$.

In a similar way, ellipticals in both the $\tau$-ratio and the $v$-threshold models are objects with overdensities bounded from below $v(M) > v_E(M)$. The fraction of halos which should be identified with elliptical galaxies is given by

$$f_E(M) = \max\left\{1, \frac{\mathrm{erfc}\,[v_E(M)]}{\mathrm{erfc}\,[v_z(M)]}\right\}. \qquad (15)$$

Finally, the spiral fraction is bounded both from above and below, $v_{Sp}(M) < v(M) < v_E(M)$, although the range differs between the $\tau$-ratio and the $v$-threshold models. The fraction of halos with embedded disk galaxies is thus

$$f_{Sp}(M) = \max\left\{1, \frac{\mathrm{erfc}\,[v_{Sp}(M)] - \mathrm{erfc}\,[v_E(M)]}{\mathrm{erfc}\,[v_z(M)]}\right\}. \qquad (16)$$

Note that there is no dynamics in this model, except what is included in the Press-Schechter relation for the population of

halos. The procedure here is to simply paint the distribution of halos *at a given epoch* with labels corresponding to different Hubble types in a manner consistent with that expected from the initial Gaussian distribution of overdensities. Presumably, the galaxy population seen today is a suitable convolution of galaxies formed at all previous epochs. Without knowing *a priori* the proper form for the time convolution operator to apply at previous epochs (another way of saying that we do not know what happens to galaxies within merged halos), we cannot predict in detail what the galaxy population will look like at the present day. Still, the "snapshot" approach used here is not totally without merit, since there are several important aspects of the models which are insensitive to redshift and hence would be valid for *any* time convolution model.

### c) Global Population of Normal Galaxies

Figure 5 shows the (comoving) number density of normal (E + Sp) galaxies as a function of redshift for each of the models and the two different CDM normalizations. The characteristic shape of the number density can be understood qualitatively by considering the initial Gaussian distribution of overdensities. At early times, only perturbations in the extreme tail of the Gaussian are able to collapse, and so few objects are found. The logarithmic number density increases rapidly with time as the objects just collapsing ride up the shoulder of the Gaussian. After 1 $\sigma$ objects have collapsed, the population begins to decline as a result of merging or accretion of surrounding material. This characteristic behavior is also exhibited by halos identified in $N$-body experiments (Frenk *et al.* 1988).

What we call the present number density of normal galaxies depends on our assumptions regarding the fate of galaxies within merging dark halos. Even though colliding galactic halos merge on roughly a dynamical time scale (White 1978; Barnes 1984), the dissipated stellar components may possess sufficient binding energy to withstand the impact and avoid
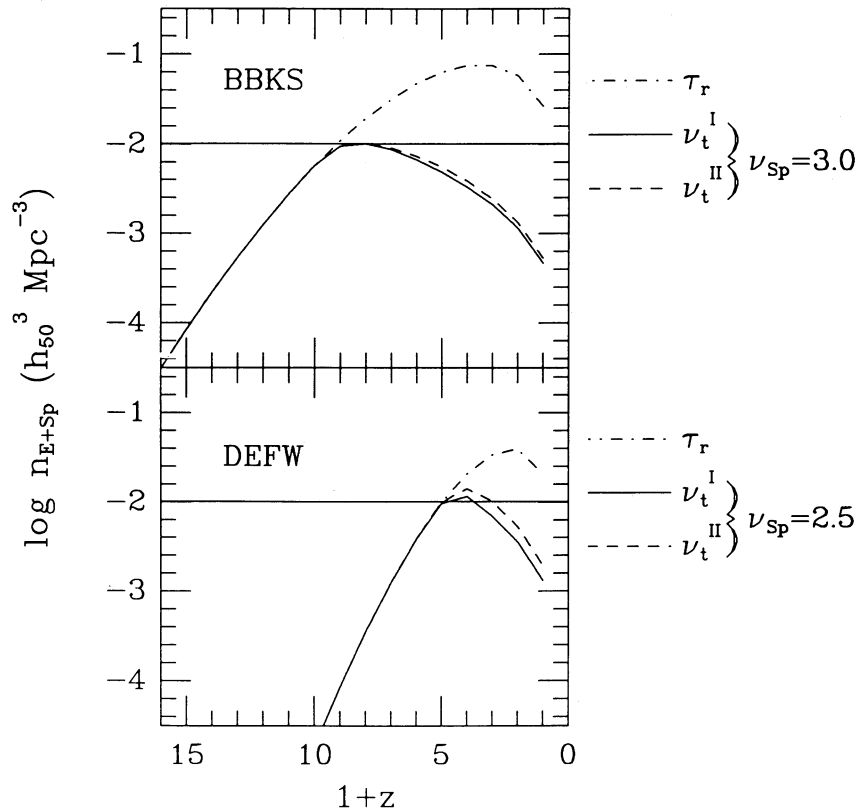
FIG. 5.—Comoving number density of halos which would produce normal galaxies (E + Sp) as a function of redshift. The unbiased $\tau$-ratio model overproduces halos, while the maximum value in the $\nu$-threshold model has been tuned to the observed value indicated by the horizontal solid line. This sets the level as $\nu_{Sp}$ as indicated.

merging for a Hubble time (White and Rees 1978). Recent $N$-body simulations by Carlberg (1988) have shown that dissipated systems formed within hierarchically merging halos can remain distinct long after the halo components have merged. Observationally, this must be true if we are to understand the existence of compact groups such as those compiled by Hickson (1982). This would then motivate us to take the present galaxy density to equal the maximum density found at any redshift.

With this prescription, the $\tau$-ratio model predicts a normal galaxy density factors of 3 and 10 larger than observed in the DEFW and BBKS normalizations, respectively. The peak in the number density occurs at very low redshifts—$z = 1$ (DEFW) and $z = 2$ (BBKS). The number density in the $\nu$-threshold models can be adjusted by varying the threshold parameter $\nu_{Sp}$. Matching the peak density to observations then requires $\nu_{Sp} = 2.5$ with the DEFW normalization and $\nu_{Sp} = 3.0$ with BBKS. The "epoch of galaxy formation," as defined, for example, by the range in redshift for which the number density is greater than half the peak value, would then be $z \sim 2$–$4$ (DEFW) and $z \sim 4$–$9$ (BBKS).

Predictions for the global ratio of disk to spheroidal systems are shown in Figure 6. The $\tau$-ratio model is in serious conflict with observations: ellipticals are nearly always more abundant than spirals; for example, at $z = 2$, E's outnumber Sp's by a factor of 2. This is one of the major shortcomings of this model. The situation would be made even worse if the cooling rate were increased by metal enrichment, since this effectively increases the dividing mass scale between E's and Sp's without

much affecting the high-mass (and high-temperature) upper endpoint of the spirals, resulting in an even larger abundance of spheroids. In the $\nu$-threshold models, one again has a degree of freedom, the value of $\nu_E$, to vary in order to match the observed global abundance. This is achieved by using $\nu_E = 3.35$ (DEFW) and $\nu_E = 3.72$ (BBKS).

The predicted mass functions for spiral and elliptical galaxies are shown in Figure 7 for snapshots taken at redshifts $z = 7$, 3, and 0. The $\tau$-ratio model predicts that spirals are embedded in halos typically a factor of 5 more massive than those of ellipticals. This goes against the observational fact that the most luminous systems are preferentially elliptical, although the model predictions could be fixed if the $L/M$ values of E galaxies were typically larger by about a factor of 10 than those of spirals.

In contrast, the mass functions for E's and Sp's in the $\nu$-threshold models span roughly the same range in mass, with a slight tendency for the most massive system to be elliptical. The width of the distribution is controlled from below by the critical dwarf temperature $T_c$ and from above by the imposed cooling condition. This width remains constant in time with the more stringent cooling condition $\tau_{cool} < \tau_{dyn}$ (model $\nu_t^I$) because this condition occurs essentially along a line of constant mass $M \simeq 3 \times 10^{11}\ M_\odot$. The observed width of the spiral and elliptical luminosity functions, about 2 orders of magnitude, is nearly reproduced with this cooling criterion in the BBKS normalization, while the DEFW biasing produces extremely narrow mass distributions. With more generous cooling within a Hubble time, $\tau_{cool} < \tau_H$, the mass functions
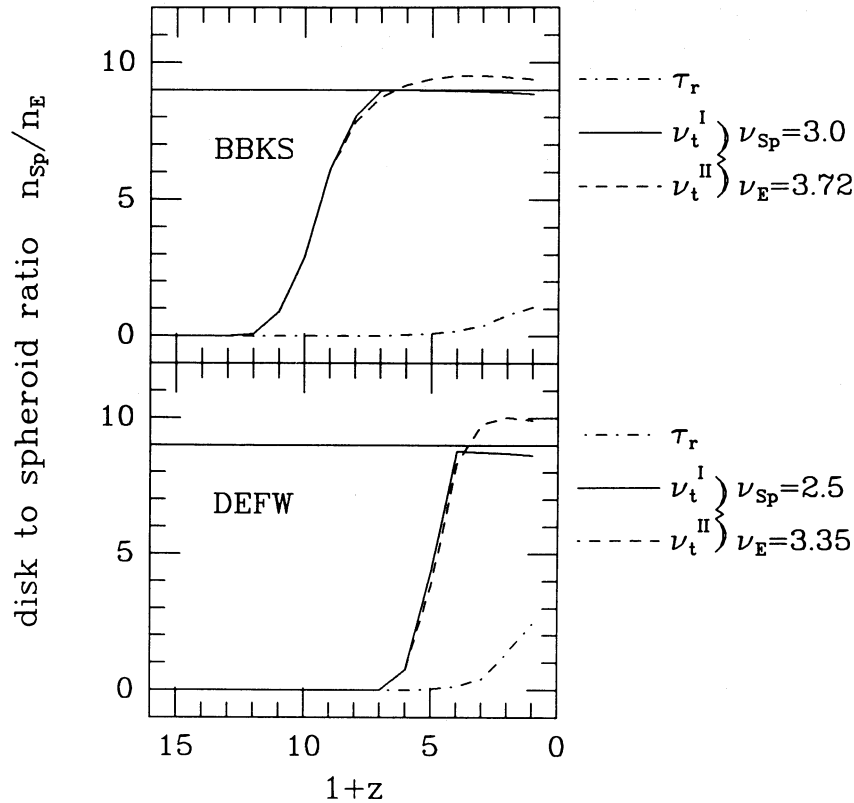
FIG. 6.—Ratio of the number densities of disk to spheroid systems as a function of redshift. The $\tau$-ratio model produces too few spirals, while the $\nu$-threshold model is again adjusted to the observed value (*horizontal line*), setting the level of $\nu_E$.
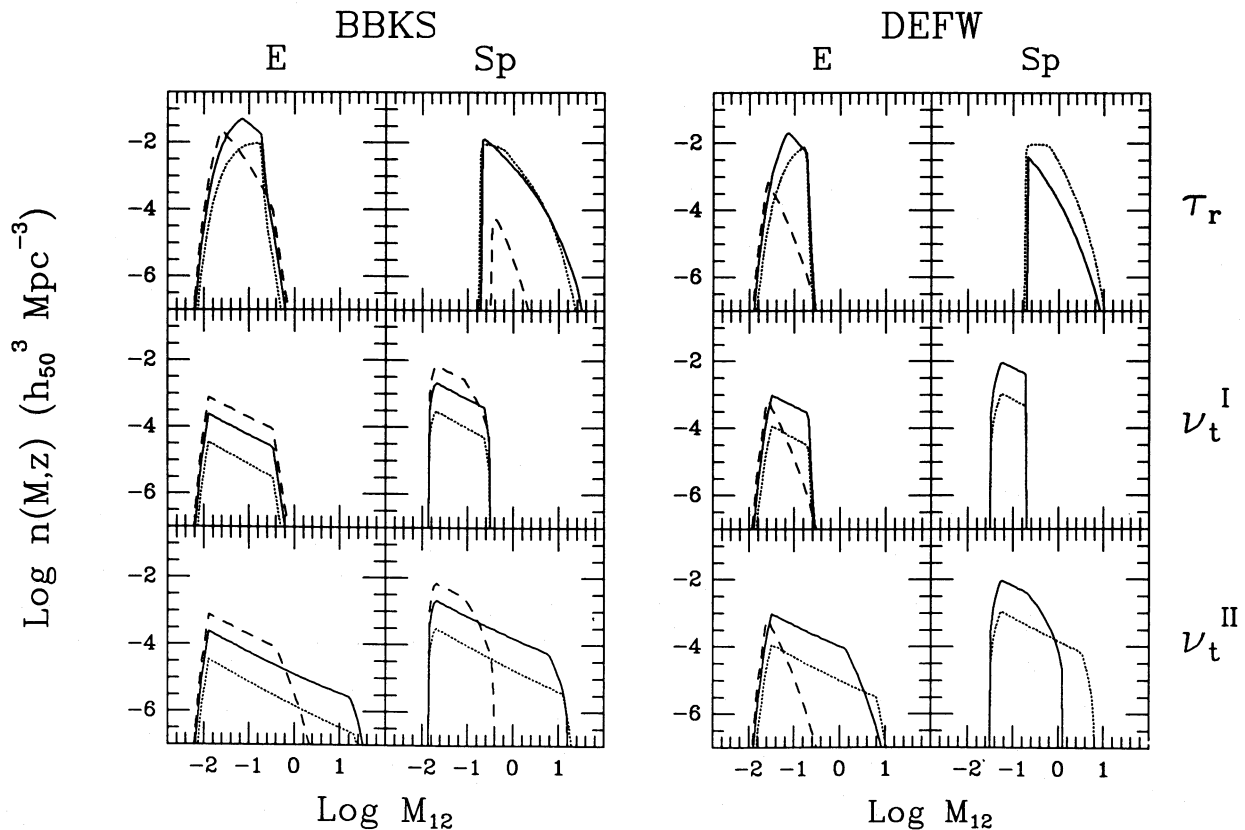


FIG. 7.—Mass functions of the halos which would produce E and Sp galaxies viewed as snapshots at three redshifts: $z = 7$ (*dashed lines*); $z = 3$ (*solid lines*); and $z = 0$ (*dotted line*). The top panels are for the $\tau$-ratio model, the middle for $\nu$-threshold with $\tau_{cool} < \tau_{dyn}$, and the lower panels show the $\nu$-threshold model with $\tau_{cool} < \tau_H$.

widen by an order of magnitude or more. The maximum possible increase is required for the DEFW disk systems; their cooling and formation would have to continue to the present epoch. On the other hand, the most massive BBKS systems must be prevented from forming stars efficiently in order not to produce a luminosity function significantly wider than 2 decades in $L$.

To summarize, the $\tau$-ratio model is unable to produce even rough agreement with the observed mass functions and abundances of disk and spheroidal systems. Ellipticals are predicted to be more abundant and substantially less massive than disk galaxies. The $\nu$-threshold model has two free parameters which are set by the observed number density and relative global abundance of E's and Sp's. With the added assumption of similar $L/M$ ratios for spirals and ellipticals, this model can produce mass functions in good qualitative agreement with observations.

### d) Dwarfs

The predicted mass function for dwarfs is the same in both models and is shown in Figure 8. Although observational knowledge of the dwarf population is hindered by their intrinsic low luminosity and surface brightness, a limited comparison with dwarf properties is possible. One issue is the faint-end slope $\alpha$ given in

$$n(M, z)d \ln M \propto M^\alpha \exp(-M)dM .$$

A value $\alpha \simeq -2$ is predicted by the models, steeper than the
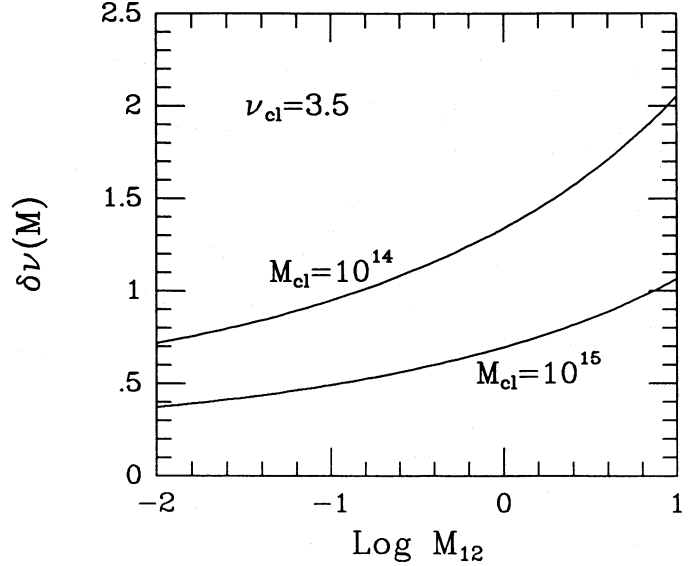


FIG. 9.—Difference in the values of $\nu$ required to produce a given overdensity $\delta$ between perturbations in the field and those embedded within 3.5 $\sigma$ clusters on mass scales $10^{14}$ and $10^{15}$ $M_\odot$.

observed value $\alpha \simeq -1.25$. The steep predicted value is the same as that seen by Frenk *et al.* (1988). There are two ways to explain this discrepancy. One is to postulate that faint galaxy counts have been significantly underestimated, as suggested by Binggeli (1987). Another, perhaps more attractive alternative is that dwarf $L/M$ values are strongly mass-dependent (Dekel and Silk 1986; Davis and Efstathiou 1988). The faint end of the luminosity function is then dependent on the model for $L/M$. For example, a power-law model

$$L/M \propto M^q \tag{17}$$

can reconcile an observed faint-end slope $\alpha_{obs}$ with a theoretical prediction $\alpha_{th}$ if

$$q = \frac{\alpha_{th} - \alpha_{obs}}{\alpha_{obs} + 1} . \tag{18}$$

The numbers quoted above would imply $q \simeq 3$, or $L \propto M^4$, a fairly strong dependence. The models of Dekel and Silk predict $L \propto M^2$. Note that one qualitative agreement of the model with observations is the existence of an overlap in mass of normal and dwarf galaxies.

### V. "NATURAL" MORPHOLOGICAL BIASING IN CLUSTERS

The critical values $\nu_{Sp}$ and $\nu_E$ were tuned in the $\nu$-threshold models to agree with the observed global proportion of E and Sp galaxies. In attempting to identify a *physical* mechanism responsible for morphology in this model, recall that the critical values of $\nu$ correspond to critical values of the overdensity

$$\delta_0(M) = \nu\sigma_0(M) . \tag{19}$$

The overdensity $\delta_0(M)$ determines the virial properties of collapsed structures (eq. [7]), and hence, presumably, the physics of the star formation within them. To test the significance of the values of $\delta_E(M)$ and $\delta_{Sp}(M)$ defined by using the critical value $\nu_E$ and $\nu_{Sp}$ in equation (19), we can bias the galactic population with long-wavelength density perturbations
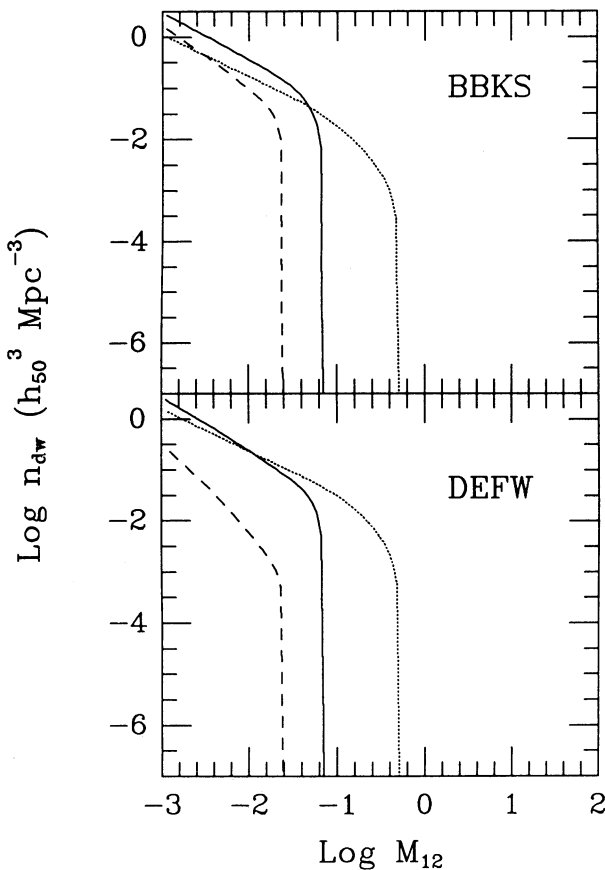


FIG. 8.—Mass function of halos which would produce dwarfs at three different redshifts. Line styles are for the same redshifts as in Fig. 7.

(Kaiser 1984) and ask whether the observed cluster population abundances are reproduced with the same critical values of $\delta_E(M)$ and $\delta_{Sp}(M)$ inferred globally.

We consider a population of galactic-scale perturbations $\delta(M)$ which have been enhanced with a statistically independent background overdensity $\delta_{bias}$. Let us further associate the biasing density with that expected from protoclusters of mass $M_{cl}$ and height $\nu_{cl}$, allowing us to write $\delta_{bias} = \nu_{cl} \sigma(M_{cl})$. Ignoring coupling of the perturbations on the two different scales, the galactic-scale perturbations will be normally distributed in the variable

$$\nu' = \frac{\delta(M)}{\sigma(M)} - \nu_{cl} \frac{\sigma(M_{cl})}{\sigma(M)} \;. \qquad (20)$$

The degree of biasing characterized by the second term on the right-hand side, $\delta\nu(M) = \nu_{cl}[\sigma(M_{cl})/\sigma(M)]$, is shown in Figure 9 for 3.5 $\sigma$ protoclusters of mass $10^{14}\ M_\odot$ and $10^{15}\ M_\odot$, corresponding roughly to the core and total masses of a typical rich cluster.

The consequences of cluster biasing on the disk-to-spheroid ratio are shown for the $\nu_t^I$ model in Figure 10. (Results for the $\nu_t^{II}$ model are similar, and the $\tau$-ratio model shows no effect.) Because the general effect of biasing is to reduce the value of $\nu$ required for a given overdensity $\delta$, the elliptical fraction in clusters can be significantly enhanced over the global value. The observed enhancement, from 10% E's in the field to 20% in clusters, is reproducible by the stronger biasing on the core mass scale. This is in accord with the idea that morphology is more closely linked to local conditions than to the larger scale

cluster environment.

The shape of the elliptical and spiral mass functions are little changed by the cluster bias, as shown in Figure 11. This again is in agreement with the comparisons made by Binggeli (1987) between LFs in the field, Virgo, and Coma for the different morphological types. His conclusion was that the underlying LFs for the Hubble types were independent of position, but that the ratio among the types was modulated by the local environment. The $\nu$-threshold model naturally accounts for this behavior in terms of varying degrees of biasing expected in different environments. Indeed, "antibiasing" in a void would serve to enhance the spiral fraction.

## VI. SUMMARY AND DISCUSSION

A procedure for calculating mass functions for different morphological types was presented and used to test existing ideas for the origin of the major Hubble types. The procedure yields a snapshot of the galaxy distribution at a given redshift by combining three ingredients: (1) the Press-Schechter estimate of the halo mass function; (2) simplified dynamics based on the spherical top-hat prescription; and (3) appropriate fractions based on physical criteria and a Gaussian distribution of overdensities. A characteristic temperature $T_c = 10^{5.4}$ K divides dwarfs from normal galaxies. This division produces a low-mass cutoff in the mass functions of normal disk and spheroidal systems. The steep faint-end slope found for the mass function of all galaxies can be reconciled with observations if dwarf luminosity is a strongly increasing function of mass, $L \sim M^4$.

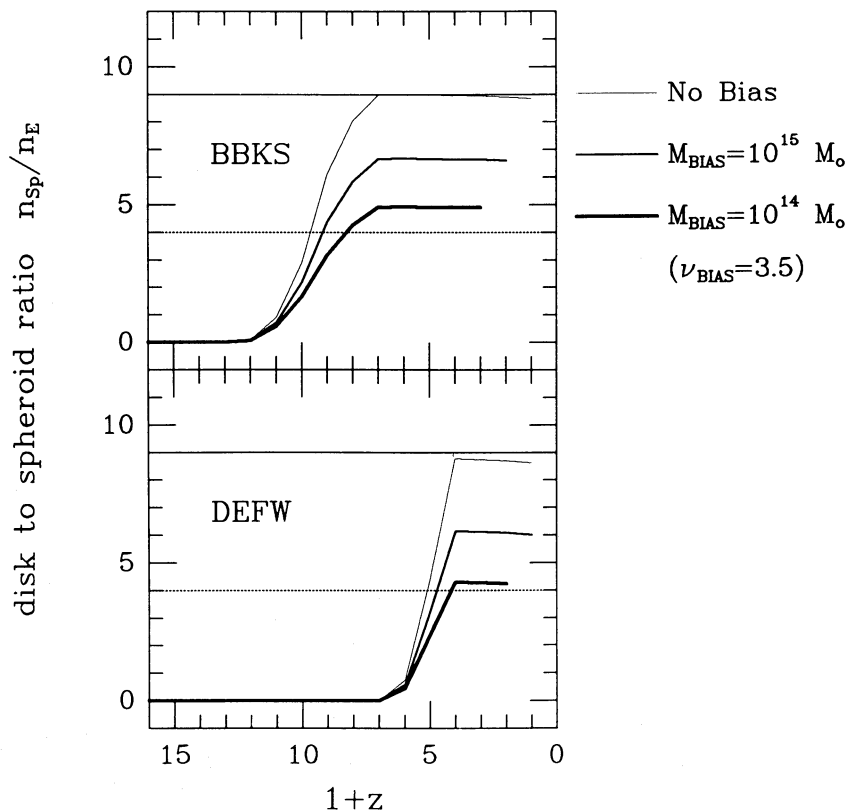Of the two models investigated to differentiate disk from



FIG. 10.—Effects of biasing on the disk-to-spheroid ratio. With the critical overdensities $\delta_E$ and $\delta_{Sp}$ defined by the global population results held fixed, the cluster fraction of E galaxies can be enhanced by a factor $\sim 2$ as observed.
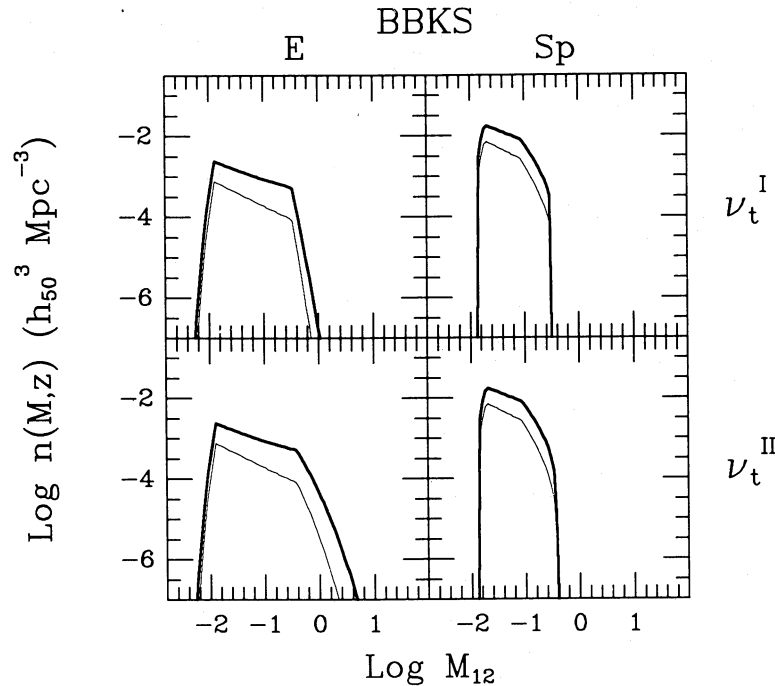
FIG. 11.—Mass functions at $z = 7$ for the unbiased population (*thin line*) and the population biased in $10^{15} M_\odot$ clusters (*heavy line*). The shapes of the mass functions vary little in going from the field to clusters in agreement with Binggeli's (1987) observations.

spheroids, the physically motivated $\tau$-ratio model fails to reproduce the low observed global fraction of E's. It also predicts that disk systems form in halos typically 5 times more massive than the halos of ellipticals. Further, the observed narrow luminosity-velocity relations are not satisfied by the galactic halos in this model. Although the latter two of these maladies could be cured by suitably contrived $L/M$'s (Kaiser 1988), the model appears in very serious trouble with the overproduction of ellipticals. One could attempt to shrink the number of ellipticals produced by simply reducing the area on the $n$-$T$ plane populated by ellipticals, either by pushing up the critical temperature $T_c$ or by pushing down the critical cooling ratio to $\tau_{cool}/\tau_{dyn} = Q$, where $Q < 1$. However, reducing the total number of ellipticals in this way would be done at the expense of creating an extremely narrow E mass function only about a factor of 2 wide in mass. This again would make the model hard to reconcile with observations.

The empirically motivated $v$-threshold model, which was designed to reproduce the $L$-$v$ relations, had two free parameters which were tuned to produce a peak galactic number density and global abundance ratio equal to those presently observed. The resulting threshold for bright galaxy formation, $v \simeq 2.5$–3, is in good agreement with that required to reconcile an $\Omega = 1$ CDM universe with observations determined by statistics of peaks (Bardeen *et al.* 1986) and $N$-body simulations (Davis *et al.* 1985). The model can produce mass functions for disk and spheroids which span the same range of about two decades in mass, in good qualitative agreement with observations. It also yields a "natural" morphological bias, in the sense that the elliptical fraction in high-density regions is enhanced.

For the $v$-threshold model, the "epoch of galaxy formation," as determined by the era during which the number density of halos capable of producing normal galaxies is above half its

peak value, occurs at redshifts $z \simeq 4$–9 for the BBKS normalization and $z \simeq 2$–4 for the DEFW level of bias. The DEFW model is still consistent with redshift 4 quasars (Efstathiou and Rees 1988), but detection of significant numbers of higher redshift galaxies could pose problems for this normalization. The results presented here are incapable of discriminating between the normalizations, essentially because of the uncertainty in what cooling critiera to apply to normal galaxies. The mass functions in the $v$-threshold model with the DEFW bias and cooling allowed on a Hubble time look rather similar to those in the BBKS bias with cooling on a dynamical time (Fig. 7).

Does the $v$-threshold model conflict with the physically reasonable idea that disks form on time scales long compared with a dynamical time? Perhaps not, as long as the star formation rate is not directly linked to the rate at which gas can cool. Breaking this link is not implausible. Several lines of argument point to enhanced high-mass star formation in the first generations of stars (Carr, Bond, and Arnett 1984). These early generations could reheat much of the cooled galactic gas through UV emission and supernovae, and also could pollute the first low-mass (Population II) stars with metals, possibly to a level $Z \simeq 0.25 Z_\odot$ (Cayrel 1986). The newly reheated and enriched envelope could then cool quasi-statically onto a disk once the halo potential had settled down. This heating process would move the baryonic component of galaxies to positions on the $n$-$T$ plane different from that expected from shock-heating via gravitational collapse, so the simple line where $\tau_{cool} = \tau_{dyn}$ in Figure 2 need no longer apply. In this way, spiral disks would still be able to form over many halo dynamical times even though their halos initially landed in regions of the $n$-$T$ plane where $\tau_{cool} < \tau_{dyn}$. This line of argument would effectively invalidate the $\tau$-ratio model, since it is predicated on a direct link between the cooling and star formation rates.

What physical mechanism underlies the $v$-threshold model?

It is interesting to note that lines of constant $v$ correspond closely to lines of constant pressure,

$$P = nT \sim \delta^4(M)M^{2/3} \sim v^4 , \qquad (21)$$

since the CDM spectrum has $\delta(M) \propto vM^{-1/6}$ on galactic scales. Thus, the critical values of $v_E$ and $v_{Sp}$ define critical virial pressures $P_E$ and $P_{Sp}$. But why should pressure regulate star formation? Because we lack fundamental theories of hierarchical fragmentation and star formation (Scalo 1988), an unambiguous answer is impossible to give at this time. However, Figure 12 gives an indication of the possible importance of pressure through two interesting coincidences. Here the virial pressure for 1 $\sigma$, 2 $\sigma$, and 3 $\sigma$ collapsing objects is shown as a function of mass using the BBKS normalization. The horizontal line equals the observed thermal pressure in the cool component of the interstellar medium (ISM), $P \simeq 4000$ K cm$^{-3}$ (Jenkins, Jura, and Loewenstein 1983; Bohlin, Savage, and Drake 1978; Cowie 1987). The first coincidence is that the peak in virial pressure for objects collapsing in the CDM model occurs at galactic mass scales. If a pressure-sensitive trigger is involved in setting off efficient star formation, then that trigger is more likely to be reached by collapsing objects with $M \sim 10^{11}$–$10^{12}$ $M_\odot$ than by higher or lower mass objects. The second coincidence is that the cool ISM pressure, which presumably is in balance with the pressure generated by supernova heat input (McKee and Ostriker 1977), is remarkably similar to the virial pressure in a collapsing 2 $\sigma$ object on galactic mass scales (3 $\sigma$ for the DEFW bias). Perhaps this empirical evidence is telling us that *concentrated star formation* (= bright galaxy formation) occurs efficiently only in environments pressurized at or above a critical level?

Unfortunately, this argument may be misleading, since constant $v$ does not *uniquely* imply constant pressure. For example, mass surface density scales as the square root of the pressure:

$$\Sigma = M/R^2 \sim \delta^2(M)M^{1/3} \sim v^2 . \qquad (22)$$

However, since $P \propto v^4$, the difference in virial pressure between a 1 $\sigma$ and a 3 $\sigma$ collapsed object is nearly 2 orders of magnitude. This may be sufficient to explain why 3 $\sigma$ objects form galaxies while 1 $\sigma$ objects do not. Still, the more subtle distinction between disks and spheroids is only a factor of $(v_E/v_{Sp})^4 \simeq 3$ in pressure. Whether or not star formation is sufficiently sensitive to support this finer morphological division remains to be seen.
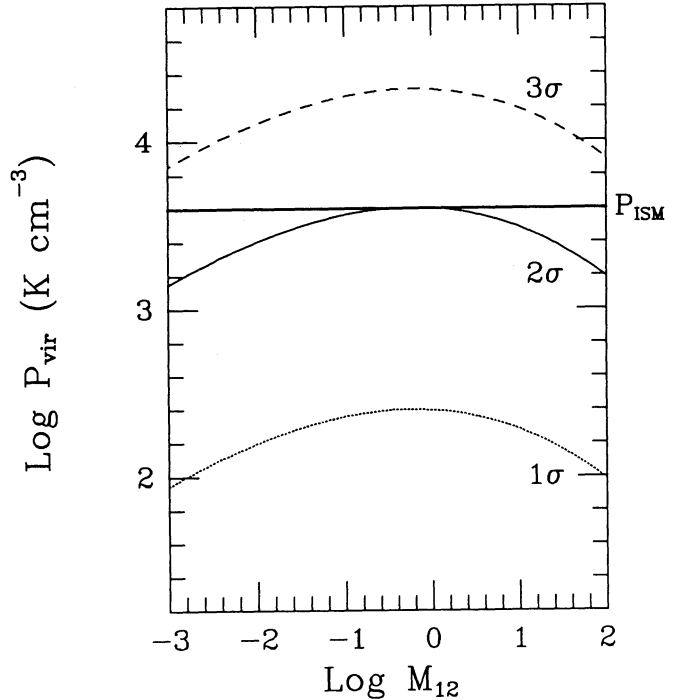


FIG. 12.—Virial pressure as a function of mass for 1 $\sigma$, 2 $\sigma$, and 3 $\sigma$ perturbations with the BBKS normalization. The horizontal line indicates the observed pressure in the ISM. Is the correspondence with the 2 $\sigma$ BBKS virial pressure merely coincidental?

With the answers buried deep within multiphase models of the ISM with any number of complex nonlinear processes at work (turbulence, magnetic fields, cloud coagulation and fragmentation, shock-heating, molecular and grain cooling, and so on), a true physical foundation for this model is almost certain to be slow in coming.

## APPENDIX

### DETAILS OF THE TOP-HAT DYNAMICAL MODEL

Consider at some initial redshift $z_i$ a spherically symmetric perturbation characterized by $\delta_i(r)$, where $\delta_i$ is the *mean* overdensity interior to a radius $r$. To avoid shell crossing, we assume that $\delta_i(r)$ is monotonic decreasing in $r$. Make the transformation to a Lagrangian coordinate $M$ using the unperturbed mass shell coordinates:

$$r_i(M) = \left[\frac{3M}{4\pi\bar{\rho}(z_i)}\right]^{1/3} = 1.50(1 + z_i)^{-1}M_{12}^{1/3}h_{50}^{-2/3} \text{ Mpc} , \qquad (A1)$$

where $M_{12} = M/10^{12} M_\odot$. Given $\delta_i(r) \to \delta_i(M)$, the turnaround redshift of a given mass shell is (Peebles 1980, eq. [19.53])

$$\frac{1 + z_{turn}(M)}{1 + z_i} = \frac{\delta_i(M)}{1.06} . \qquad (A2)$$

Assuming that, after turnaround, shells collapse by a factor of 2 and then are halted either by shock heating or violent relaxation (Gunn 1977; Bertschinger 1983), we obtain the collapse redshift and virial radius of a shell $M$:

$$1 + z_{\rm vir}(M) = 2^{-2/3}[1 + z_{\rm turn}(M)] = \frac{\delta_0(M)}{1.68} , \qquad (A3)$$

$$r_{\rm vir}(M) = 0.45\delta_0^{-1}(M)M_{12}^{1/3}h_{50}^{2/3} \text{ Mpc} , \qquad (A4)$$

where $\delta_0(M) = \delta_i(M)(1 + z_i)$ is the initial overdensity propagated to the present using linear theory.

The lining up of shells according to $r_{\rm vir}(M)$ allows calculation of the *local* mass density

$$\rho_{\rm vir}(M) = \frac{dM}{4\pi/3[r_{\rm vir}^3(M + dM) - r_{\rm vir}^3(M)]} = \left[4\pi r_{\rm vir}^2 \frac{dr_{\rm vir}(M)}{dM}\right]^{-1} .$$

Define $\beta(M)$ as a measure of the profile slope:

$$\frac{dr_{\rm vir}(M)}{dM} = \frac{r_{\rm vir}}{3M}\left(1 - \frac{3M}{\delta}\frac{d\delta}{dM}\right) \equiv \frac{r_{\rm vir}}{3M}\beta(M) . \qquad (A5)$$

Plugging this into the above relation with equation (A4) gives

$$\rho_{\rm vir}(M) = 37.3\rho_0 \, \delta_0^3(M)\beta^{-1}(M) , \qquad (A6)$$

where $\rho_0$ is the present cosmological background density.

The virial velocity and temperature are found from assuming energy conservation for each shell. Although the zero point of the total energy is not well defined, the monotonic decreasing condition on $\delta_0(M)$ guarantees no shell crossing, and so an energy integral exists for each shell

$$\frac{1}{2}v_{\rm vir}^2(M) - \frac{GM}{r_{\rm vir}(M)} = \frac{GM}{r_{\rm turn}(M)} = -\frac{1}{2}\frac{GM}{r_{\rm vir}(M)} , \qquad v_{\rm vir}^2(M) = \frac{GM}{r_{\rm vir}(M)} = 3(56.4 \text{ km s}^{-1})^2 \, \delta_0(M)M_{12}^{2/3}h_{50}^{2/3} . \qquad (A7)$$

The virial temperature can be found by equating specific energies $3nkT_{\rm vir} = \rho_b v_{\rm vir}^2$ and using $n = \rho_b/(\mu m_p)$ with $\mu = 0.6$ to give

$$T_{\rm vir}(M) = 2.31 \times 10^5 \, \delta_0(M)M_{12}^{2/3}h_{50}^{2/3} \text{ K} . \qquad (A8)$$

We are now in a position to define the cooling time on a given shell. The pressure drop due to radiative cooling is controlled by

$$n\frac{d(\frac{3}{2}kT)}{dt} = -n_{\rm H}^2\Lambda(T) ,$$

where $n_{\rm H}$ is the number density of hydrogen ions. For a fully ionized primordial ($X = 0.76$, $Y = 0.24$) H-He mixture, $n_{\rm H}/n \simeq 0.45$. This leads to the definition of the cooling time:

$$\tau_{\rm cool} = \frac{\frac{3}{2}kT}{(0.45)^2\rho_0 \tilde{\Lambda}(T)} ,$$

where $\tilde{\Lambda}(T) = \Lambda(T)/(\mu m_p) \simeq 10^{24}\Lambda(T)$. Using $T = T_{\rm vir}(M)$ and $\rho_b = \Omega_b \rho_{\rm vir}(M)$ gives the cooling time for a collapsed shell:

$$\tau_{\rm cool}(M) = 43.1\Omega_b^{-1} \, \delta_0^{-2}(M)M_{12}^{2/3}\beta(M)\tilde{\Lambda}(T_{\rm vir})^{-1}h_{50}^{-4/3} \text{ Gyr} . \qquad (A9)$$

Define the dynamical time to be the virial crossing time for a shell:

$$\tau_{\rm dyn}(M) \equiv 2\frac{r_{\rm vir}}{v_{\rm vir}} = 9.21 \, \delta_0^{-3/2}(M)h_{50}^{-1} \text{ Gyr} . \qquad (A10)$$

The interesting ratios are those of the cooling to the dynamical time,

$$\frac{\tau_{\rm cool}(M)}{\tau_{\rm dyn}(M)} = 4.68\Omega_b^{-1} \, \delta_0^{-1/2}(M)M_{12}^{2/3}\beta(M)\tilde{\Lambda}(T_{\rm vir})^{-1}h_{50}^{-1/3} , \qquad (A11)$$

and the cooling to present Hubble time,

$$\frac{\tau_{\rm cool}(M)}{\tau_{\rm H}} = 3.24\Omega_b^{-1} \, \delta_0^{-2}(M)M_{12}^{2/3}\beta(M)\tilde{\Lambda}(T_{\rm vir})^{-1}h_{50}^{-1/3} . \qquad (A12)$$

Finally, for large perturbations Compton cooling will also be important. The relevant ratio is

$$\frac{\tau_{\text{comp}}(M)}{\tau_{\text{dyn}}(M)} = 2120\,\delta_0^{-5/2}(M)h_{50}\ , \tag{A13}$$

so the Compton cooling time scale will be shorter than the dynamical time for collapse redshifts $z_{\text{vir}} \gtrsim 10$.

## REFERENCES

Bardeen, J. M., Bond, J. R., Kaiser, N., and Szalay, A. S. 1986, *Ap. J.*, **304**, 15 (BBKS).
Barnes, J. 1984, *M.N.R.A.S.*, **208**, 885.
Bertschinger, E. 1983, *Ap. J. Suppl.*, **58**, 39.
Binggeli, B. 1987, in *Nearly Normal Galaxies*, ed. S. M. Faber (New York: Springer-Verlag), p. 195.
Blumenthal, G. R., Faber, S. M., Primack, J. R., and Rees, M. J. 1984, *Nature*, **311**, 517.
Bohlin, R. C., Savage, B. D., and Drake, J. F. 1978, *Ap. J.*, **224**, 132.
Bond, J. R., and Efstathiou, G. 1984, *Ap. J. (Letters)*, **285**, L45.
Carlberg, R. G. 1988, *Ap. J.*, **324**, 664.
Carr, B. J., Bond, J. R., and Arnett, W. D. 1984, *Ap. J.*, **277**, 445.
Cayrel, R. 1986, *Astr. Ap.*, **168**, 81.
Cowie, L. L. 1987, in *Interstellar Processes*, ed. D. J. Hollenbach and H. A. Thronson (Dordrecht: Reidel), p. 245.
Davis, M., and Efstathiou, G. 1988, in *Proc. Vatican Study Week 1987, Large-Scale Motions in the Universe* (Vatican City: Vatican Press).
Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. M. 1985, *Ap. J.*, **292**, 371 (DEFW).
Dekel, A., and Silk, J. 1986, *Ap. J.*, **303**, 39.
Dressler, A. 1980, *Ap. J.*, **236**, 351.
Efstathiou, G., Frenk, C. S., White, S. D. M., and Davis, M. 1988, *M.N.R.A.S.*, **234**, 715.
Efstathiou, G., and Rees, M. J. 1988, *M.N.R.A.S.*, **230**, 5P.
Eggen, O. J., Lynden-Bell, D., and Sandage, A. R. 1962, *Ap. J.*, **136**, 748.
Evrard, A.E. 1988, in *Large-Scale Structure and Motions in the Universe*, ed. F. Mardirossian (Dordrecht: Reidel), in press.
Evrard, A. E., and Davis, M. 1988, *Nature*, **333**, 335.
Faber, S. M. 1982 in *Astrophysical Cosmology*, ed. H. A. Bruck, G. V. Coyne, and M. S. Longair (Vatican City: Pontificia Academia Scientiarum), p. 191.
Fall, M., and Rees, M. J. 1985, *Ap. J.*, **298**, 18.
Frenk, C. S., White, S. D. M., Davis, M., and Efstathiou, G. 1988, *Ap. J.*, **327**, 507.
Geller, M. J., and Huchra, J. P. 1983, *Ap. J. Suppl.*, **52**, 61.
Gunn, J. E. 1977, *Ap. J.*, **218**, 592.
Gunn, J. E., and Gott, J. R. 1972, *Ap. J.*, **209**, 1.

Hickson, P. 1982, *Ap. J.*, **255**, 382.
Hickson, P., Kindl, E., and Huchra, J. P. 1988, *Ap. J.*, **331**, 64.
Jenkins, E. B., Jura, M., and Loewenstein, M. 1983, *Ap. J.*, **270**, 88.
Kaiser, N. 1964, *Ap. J. (Letters)*, **284**, L9.
———. 1988, in *Large-Scale Structure and Motions in the Universe*, ed. F. Mardirossian (Dordrecht: Reidel), in press.
Krishner, R. P., Oemler, A., Jr., Schechter, P. L., and Shectman, S. A. 1983, *A.J.*, **88**, 1285.
Kormendy, J. 1987, in *Nearly Normal Galaxies*, ed. S. M. Faber (New York: Springer-Verlag), p. 163.
Larson, R. B. 1969, *M.N.R.A.S.*, **145**, 405.
———. 1974, *M.N.R.A.S.*, **169**, 229.
———. 1975, *M.N.R.A.S.*, **173**, 671.
Larson, R. B., Tinsley, B. M., and Caldwell, C. N. 1980, *Ap. J.*, **237**, 692.
Mathews, W. G., and Baker, J. C. 1971, *Ap. J.*, **170**, 241.
McGlynn, T. A. 1984, *Ap. J.*, **281**, 13.
McKee, C. F., and Ostriker, J. P. 1977, *Ap. J.*, **218**, 148.
Oemler, A. 1974, *Ap. J.*, **194**, 1.
Peebles, P. J. E. 1980, *The Large-Scale Structure of the Universe* (Princeton: Princeton University Press).
———. 1982, *Ap. J. (Letters)*, **263**, L1.
Postman, M., and Geller, M. J. 1984, *Ap. J.*, **281**, 95.
Press, W. H., and Schechter, P. 1974, *Ap. J.*, **187**, 425.
Rees, M. J., and Ostriker, J. P. 1977, *Ap. J.*, **179**, 541.
Sandage, A. R. 1961, *The Hubble Atlas of Galaxies* (Washington, DC: Carnegie Institution of Washington).
Sandage, A. R., Binggeli, B., and Tammann, G. A. 1985, *A. J.*, **90**, 1759.
Scalo, J. M. 1988, in *Molecular Clouds in the Milky Way and External Galaxies*, ed. R. Dickman, R. Snell, and J. Young (Dordrecht: Reidel), in press.
Silk, J. I. 1977, *Ap. J.*, **211**, 638.
Van Albada, T. S. 1982, *M.N.R.A.S.*, **201**, 939.
White, S. D. M. 1978, *M.N.R.A.S.*, **184**, 185.
White, S. D. M., and Rees, M. J. 1978, *M.N.R.A.S.*, **183**, 341.
Yahil, A., Sandage, A. R., and Tammann, G. A. 1980, *Ap. J.*, **242**, 448.
Yoshii, Y., and Arimoto, N. 1987, *Astr. Ap.*, **188**, 13.

AUGUST E. EVRARD: Department of Astronomy, University of California, Berkeley, CA    94720