GALAXY FORMATION

Joseph Silk
Astronomy Department
University of California
Berkeley, California 94720 U.S.A.

ABSTRACT. Theories of galaxy formation via hierarchical clustering of cold dark matter and by fragmentation of gaseous pancakes or shells are reviewed and compared. Dissipative processes are crucial to all theories of galaxy formation, and are discussed in terms of a simple model involving multiple cloud interactions. Stellar energy input is found to play an important role in protogalaxies, both supporting the gas against collapse, thereby prolonging the duration of the active star formation phase, and driving winds from the less massive galaxies. The significance of such processes is explored both for chemical and dynamical evolution and for biasing galaxy formation towards density peaks in the primordial density fluctuation spectrum.

1. INTRODUCTION

What determines the masses and the binding energies of galaxies? These issues are the principal focus of galaxy formation theory. Once these are resolved, more detailed models can be constructed which may (or may not!) approximate the structural, morphological, dynamical, and chemical properties of galaxies.

There are two extreme viewpoints concerning the origin of galactic binding energy. It may derive from the cosmological initial conditions. Primordial curvature fluctuations, laid down at the inflationary epoch some 10^{-35} s after the big bang, develop adiabatically, and generate growing density fluctuations within the particle horizon on galactic and subgalactic scales. Small scales go nonlinear first, at redshift $z \sim 30$, and cluster hierarchically, merging together to form larger and larger systems. Gravitational clustering of weakly interacting dark matter determines the halo scale, and galaxies develop by baryonic infall into dark potential wells (Peebles 1984; Blumenthal *et al.* 1984). This scheme reduces to a resurrection of galaxy formation from primordial isothermal fluctuations (Peebles and Dicke 1968). The binding energy of a galaxy is effectively imprinted in the initial conditions.

In an alternative scenario, all primordial fluctuations on galactic and subgalactic scales are suppressed (Silk 1968; Bond *et al.*). Only fluctuations survive

J. Kormendy and G. R. Knapp (eds.), Dark Matter in the Universe, 335–359.

on galaxy cluster scales, collapsing anisotropically to form thin sheets or pancakes which subsequently fragment (Zel'dovich 1970). Dissipative processes are largely responsible for the binding energies of the galaxies that form (Larson 1975; Carlberg 1984). Galaxy halos are presumed to develop by secondary infall, if weakly interacting dark matter dominates the mass density of the universe (Gunn 1977). In a variation on this scheme, rare seed fluctuations of subgalactic mass explode, sweeping up a shell of matter that fragments on scales of mass greatly exceeding that of the initial seeds (Ostriker and Cowie 1981; Ikeuchi 1981; Carr and Ikeuchi 1985). Both pancake and explosive amplification schemes form galaxies by much the same process of dissipative fragmentation.

Although the physical processes that dominate hierarchical clustering and pancake fragmentation, namely gravitational clustering and N–body dynamics, as opposed to dissipative hydrodynamics and gaseous shocks, seem very different, one ends up forming the luminous cores of galaxies in much the same manner. Dissipation dominates in luminous cores, as is evident from the simple observation that galaxies have a much higher surface brightness than galaxy clusters. Put more quantitatively, the self–similar clustering hierarchy that describes the large–scale galaxy distribution is broken on the scale of galaxies. There is a unique scale where dissipation sets in, determined by the comparison of free–fall collapse and cooling time–scales Rees and Ostriker 1977; Silk 1977). The former is proportional to $n^{-1/2}$ and the latter is approximately proportional to $T^{-3/2}/(nZ)$, where n is mean gas density, Z is metallicity, and T is gas temperature, over the range 10^5–10^7 K and Z $\gtrsim$ 0.01 $Z_\odot$. Equating these two time–scales yields a critical mass of order 100 $\alpha^5 \alpha_g^{-2}$, for a primordial abundance gas, or more generally $\sim 10^{68}(Z/0.01\ Z_\odot)$ baryons, above which no cooling occurs. Here $\alpha = 1/137$ and $\alpha_g = Gm_p^2/\hbar c = 6 \times 10^{-39}$. By way of comparison, a similar argument utilizing pressure support yields the characteristic mass of a star as $\sim \alpha_g^{-3/2} \sim 10^{57}$ baryons.

These considerations might lead one to think that one could actually predict the existence of stars and of galaxies from fundamental physical principles. While this argument may be valid for stars, it is almost certainly specious when applied to galaxies, for the following reason. Solving the galaxy formation problem requires specification of initial conditions, which are then evolved forward in time according to the equations of cosmology, N–body dynamics and hydrodynamics. We have no a priori knowledge of initial conditions, although it is perhaps not inconceivable that quantum gravity may eventually provide them. For the present, however, one has to work within the framework of preconceptions or prejudices about the initial conditions. One can then test the resulting theoretical models of galaxies against the real universe, to see whether one's starting point has any possible validity.

To commence, I contrast the hierarchical clustering and fragmentation theories (§II). I then develop a simple model of dissipative galaxy formation, common to either theory, involving multiple cloud interactions (§III). The role of stripping by winds in chemical and dynamical evolution is described in §IV, and general implications of dissipative models are discussed in §V.

II. CLUSTERING OR FRAGMENTATION?

Galaxy formation scenarios are not sufficiently well defined that one can speak of any two, or even more, principal theories. There are innumerable variations, but as previously mentioned, one can at least discern the two main themes of hierarchical clustering and gaseous fragmentation, around one or other of which most theories are developed. To be specific, I shall choose the cold dark matter and pancake fragmentation theories, and discuss how they cope when confronted with observational and theoretical realities.

One of the more contentious issues concerns the epoch of galaxy formation. A possibly fatal blow against the pancake theory in a massive neutrino–dominated universe arises because the galaxy correlations on large scales can only be fit if pancaking occurred recently, at redshift $z \lesssim 1$, in which case galaxies are uncomfortably young (White *et al.* 1983). Very few, if any, galaxies can have formed at $z \gtrsim 3$, despite the fact that most luminous quasars are found at high redshift. Moreover, galaxies observed at $z \sim 1$, inevitably very luminous ellipticals, appear to contain old stellar populations that must have formed much earlier, certainly at $z > 3$. The one ray of hope for pancake theories is that one can readily imagine variations on these theories that form rare objects early, with the bulk of galaxies forming late. The formation of most galaxies as recently as $z \sim 1$ cannot be excluded. Such variations presently seem rather ad hoc, involving speculations about isothermal seeds, strings, or an admixture of hot and cold dark matter. Even without such drastic measures, there are always a few, albeit very few, pancakes that collapse directly to form galaxies even in a massive neutrino–dominated universe. Hence it is premature to dismiss massive neutrinos, at least until the laboratory mass measurements are completed. Moreover, other more exotic possibilities exist for warm dark matter which would allow resurrection of a pancake theory. I shall subsequently refer to such theories as galaxy formation by generic pancake fragmentation. Included in this category for the purposes of the present discussion are the explosive amplification schemes for forming galaxies by fragmentation of swept–up shells.

There is also increasing evidence for the existence of relatively young galaxies. The prototype of such objects is the, hitherto unique, case of I Zw18, a gas–rich HII region–like galaxy with no evidence of an underlying old population and a metallicity of 1/30 that of the sun. Its current star formation burst is almost certainly its first (Kunth and Sargent 1985). Studies of deep counts show evidence for a relatively blue component in the distribution of distant galaxies (Kron *et al.* 1985), and distant clusters contain many galaxies that appear to have the characteristics of ellipticals which have undergone recent star formation (Dressler 1986).

Then, of course, there are the cluster cooling flows. The central elliptical or cD galaxies at the centers of these flows appear to be undergoing considerable star formation (a few $M_\odot\ yr^{-1}$) if a conventional solar neighborhood IMF is adopted, but the amount of gas apparently disappearing from the flow, presumably to form stars, is an order of magnitude larger (Fabian *et al.* 1984). These stars must have a lower supernova rate per unit star formation rate, and possibly also may be deficient in massive stars relative to the local IMF, in order to be consistent with observational constraints.

A generic pancake theory fares nicely with regard to forming galaxies in the very recent past: it provides an ample gas supply to explain even such possibly primitive objects as the intergalactic cloud in Leo and I Zw18. Hierarchical clustering theories, on the other hand, tend to use up the available gas supply during collapse on subgalactic scales: it is not at all obvious how much gas can avoid premature fragmentation into stars.

This difficulty arises again and again when other properties of galaxies are confronted with theoretical expectations. For example, the Hubble–de Vaucouleurs profiles of galaxy spheroids can be beautifully explained by gravitational collapse from cold but irregular initial conditions of some thousands of mass points (van Albada 1982). The extreme densities of elliptical galaxy cores (Fall 1979a) require some additional gaseous dissipation, however, as is also required to account for metallicity gradients in the old stellar population. That gas–rich systems are necessary throughout the build–up of structure from small to large galaxies, if mergers are important, is further indicated by the correlation between metallicity and luminosity (Mould 1984), which demonstrates that continuing star formation and gas recycling must have occurred.

Morphological correlations between galaxies and their environment (Dressler 1980) are most simply explained by appeal to tidal interactions, mergers or infall of gas–rich protogalactic clouds, although formation of galaxies biased towards rare, and hence clustered, fluctuations in cold dark matter offers some hope of a similar explanation. Globular clusters also are best explained if they formed at the same time as their parent galaxy. This is difficult to arrange in a cold dark matter theory, unless some mechanism such as heat input by low mass stars is postulated to prevent the protoglobular clusters from premature collapse (Silk and Norman 1981). However conditions in a generic pancake fragmentation theory seem favorable for the onset of thermal instability during gaseous protogalactic collapse, and this leads rather naturally to the observed mass–scales and distribution of globulars (Fall and Rees 1985). Gaseous collapse in a preexisting halo of cold dark matter would presumably also lead to a favorable environment for thermal instability, provided there was a suitable reservoir of gas.

Dark halos are a property of galaxies that find a relatively natural explanation in the cold dark matter theory, with dark matter defining the potential wells within which gas dissipates and fragments into stars (White and Rees 1978).

In the pancake model, halo formation is a secondary process, and it is by no means clear whether hot dark matter can avoid overdilution of phase space density to form the halos that are found around a number of galaxies, both large and small (Tremaine and Gunn 1979; Bond *et al.* 1983; Melott 1983). The rotation of spirals is a major success of hierarchical clustering models with dark halos. Tidal torquing between neighboring, asymmetric fluctuations of comparable mass induces an initial specific angular momentum in a newly formed protogalaxy that, after collapse through a dark halo, leads to rotationally supported disks with flat rotation curves and maximum rotational velocities that are in good accord with observations (Fall and Efstathiou 1980; White 1986). One possible flaw in this scenario concerns luminous ellipticals, which, despite having higher surface brightness than spirals, are flattened but not rotationally supported (Davies *et al.* 1983). Evidently, dissipation via collapse through dark halos cannot account for ellipticals. Possible formation mechanisms include loss of halos prior to dissipative infall (Fall 1979b), an option now considered unlikely in view of x–ray observations which indicate the presence of dark halos, or allowance for very different initial conditions, such as a much larger initial density as would happen if ellipticals formed early or preferentially in denser regions. Secondary infall of dark halos, as would happen in a pancake theory, seems more relevant for ellipticals, since it avoids excessive spin–up in dissipative collapse. Alternatively, mergers of gas–rich protogalaxies could plausibly lead to a satisfactory model. Adequate two–component models have yet to be constructed. To simultaneously obtain the $\sim r^{-3}$ profile of a flattened, anisotropy–supported luminous core coexisting with the $\sim r^{-2}$ halo profile will require models incorporating both dissipative and violent relaxation, over a greater dynamic range than hitherto available.

There are some properties that either class of theories seems capable of explaining. These include the galaxy luminosity and multiplicity functions, whose general features can be reasonably well understood in terms of processes involving dissipation (fixing L_*) and merging (yielding a characteristic power–law tail at low luminosity and exponential decline at high luminosity (Schaeffer and Silk 1985). The correlations between luminosity and either velocity dispersion (Faber and Jackson 1976) or maximum rotational velocity for various Hubble types (Tully and Fisher 1977; Rubin and Burstein 1985) can be interpreted in a rather simple framework involving dissipation and star formation (see below), and possibly also could arise in a purely dynamical model involving multiple mergers (Farouki *et al.* 1983).

The galaxy correlation function, however, has been one area where N–body simulations have come down rather strongly on the side of hierarchical clustering of cold dark matter. Both the correlation function slope and amplitude and the peculiar velocity field between galaxies can be explained if galaxy formation is biased towards the rarer, more tightly clumped fluctuations (Davis *et al.*, 1985). By contrast, a neutrino pancake model yields excessive large–scale velocities, unless galaxy formation again occurs at an unacceptable recent epoch (Kaiser 1983). Biasing also allows $\Omega = 1$ to be reconciled with the observed velocity field, since the dark matter distribution defines the typical rms fluctuations, and is relatively uniform over supercluster scales, thereby not affecting local determinations of Ω. A simple physical mechanism for biasing (to be described below) utilizes both lack of dissipation and supernova–driven winds to inhibit formation of galaxies from low–σ fluctuations in a cold dark matter background.

While generic pancake fragmentation models are not as satisfactory for explaining galaxy correlations, nor do they have much to contribute towards allowing $\Omega = 1$, they do have an important advantage over cold dark matter in two other aspects. Primary halo formation can be somewhat of an embarrassment for binary galaxies and small groups of galaxies on account of dynamical friction (Barnes 1985). Excessive merging is avoided if galaxies orbit in a common dark halo, and both the lack of correlations found in binary samples (White *et al.* 1983) and the range in dwarf galaxy properties (Aaronson 1986) also support this idea. Communal halos are most easily produced in a warm or hot dark matter model for large–scale structure (e.g. Cowsik 1985).

Another observation best explained by pancake theories is that the large–scale structure of the universe appears rather inhomogeneous on scales of 10–100 Mpc. The microwave background anisotropy measurements, uniform to an upper limit of $\delta T/T \lesssim 3 \times 10^{-5}$ on scales from $5'$ up to $30°$, sample density fluctuations on larger scales, and cannot yet distinguish between rival dark matter–dominated theories (Vittorio and Silk 1985). However the fact remains that there may exist large voids (Kirshner *et al.* 1981), chains of clusters and superclusters of galaxies (Einasto 1986), enhanced cluster–cluster clustering (Bahcall and Soneira 1983), and very large–scale motions such as that of the local supercluster relative to the cosmic microwave background radiation (Aaronson *et al.* 1985). Confirmation of the reality of any of these phenomena would provide a severe blow to the biased cold dark matter theory.

I have summarized the pros and cons for the two rival theoretical approaches to galaxy formation in Table 1. A simple algorithm for assigning numerical significance to their qualitative goodness of fit to astronomical reality suggests that there is at present little to choose between them. Both theories have weak points, both have strong features, and the ultimate theory may well be a hybrid combination of both approaches. Strong motivation to choose one over the other may have to come from considerations outside the domain of astronomy, for example, by laboratory identification of a suitable dark matter candidate.

Table 1

Comparison of Clustering and Fragmentation Scenarios

Property	Clustering of Cold Dark Matter	Fragmentation of Gaseous Pancakes
Formation epoch	XX	?
Profiles of spheroids	X	XX
Metallicity correlations	X	XX
Environment correlations	X	XX
Globular clusters	X	XX
Halos	XX	?
Rotation of Sp's	XX	X
Anisotropy of E's	X	XX
Luminosity function	XX	XX
(L,σ,v_r) correlations	X	X
$\xi(r)$, $v(r)$	XX	X
$\Omega = 1$ and dwarf galaxies	XX	?
Binaries and groups	?	XX
$\delta T/T$	XX	X
Large scale structure	X	XX
SCORE	21	20

III. LUMINOUS CORES AND DISSIPATION

Despite the divergence of viewpoints about the cosmological initial conditions, reflected in the hitherto unsuccessful searches for protogalaxies and for cosmic background radiation anisotropy, there is reasonable unanimity that considerable dissipation occurred during the formation of the mostly baryon–dominated luminous regions of galaxies. The very fact that a galaxy stands out in surface brightness above the background clustering testifies to the different, non–gravitational, physics that must have played a role in its formation. This difference is readily quantified with the galaxy correlation function: the surface brightness of an elliptical within its half–light radius lies well above the extrapolation of $\xi(r)$ from larger scales.

The characteristic properties of a galaxy must have been frozen at, or soon after, the end of its gas–rich protogalactic evolutionary phase. During the dissipative gas–rich phase, the luminous mass, binding energy, half–luminous mass radius, density and metallicity would all have rapidly evolved. However, once stars formed, and the system was predominantly stellar, any ensuing evolution

https://doi.org/10.1017/S0074180900150466 Published online by Cambridge University Press

would have occurred on a very long time–scale. Indeed, apart from the occasional merger, one can now be confident that many Hubble times must elapse before any significant dynamical or chemical evolution could occur.

This observation provides a potential test of theories of galaxy formation.

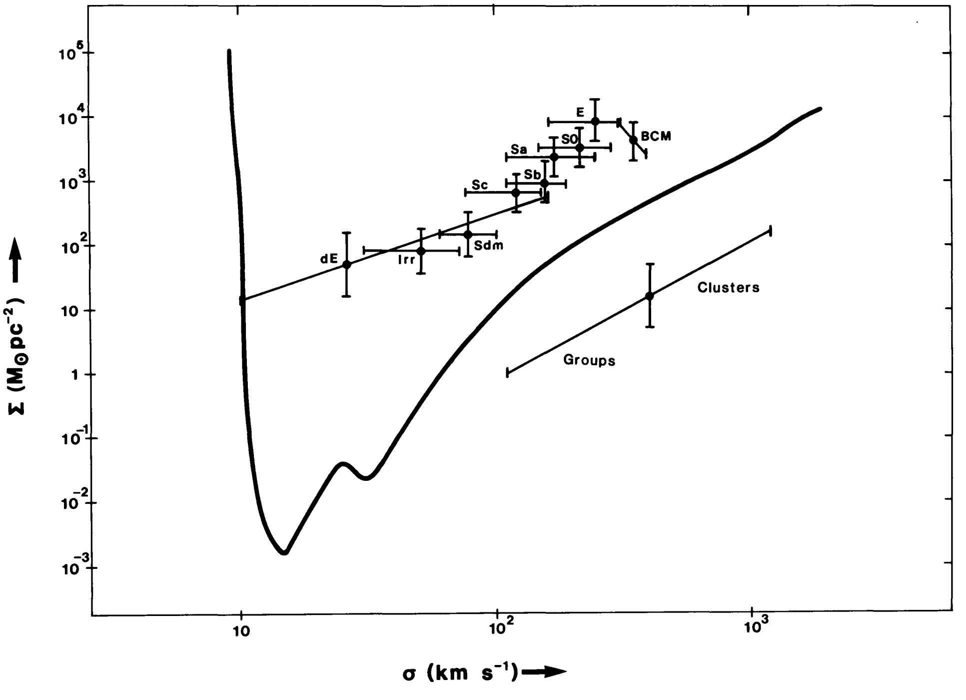

Figure 1. Surface density Σ ($M_\odot$ pc^{-2}) versus velocity dispersion σ (km s^{-1}) for different morphological types of galaxies and for galaxy groups and clusters. Plotted data is based on statistical correlations (see text for references) between luminosity L and σ (for ellipticals and brightest cluster members), and L and maximum rotational velocity (for spirals, S0's, irregulars), L and half–light radius (dwarf ellipticals), and number counts within the central megaparsec radius cores for groups and clusters (from the correlation found by Bahcall 1981). The velocity dispersion is the line–of–sight central velocity dispersion of a presumed isotropic velocity dispersion for the other systems. All correlations are converted to a (Σ,σ) relation by assuming that the systems predominantly consist of stars out to the half–light radii, with appropriate mass–to–light ratios taken from Faber and Gallagher (1979). The cooling curve shown is calculated for a collapsing cloud of primordial abundance. From Silk (1985).

It is well known that luminosity (L), one–dimensional velocity dispersion (σ), maximum rotational velocity (V_m), and surface brightness (Σ) all vary with Hubble type. A two–dimensional classification scheme of, say, Σ and σ serves to spread out Hubble types and, as will be seen shortly, has theoretical significance. I shall work with the old stellar populations, whether in disk or spheroid, and work with parameters averaged over the half–light radius. Virial equilibrium yields

$$\sigma^4 \sim G^2 \left(\frac{M}{L}\right)^2 \Sigma L,$$

so that specification of either L(σ), L(V_m), or Σ(R), all well–observed statistical correlations, together with an adopted M/L appropriate to the Hubble type, suffices to yield $\Sigma(\sigma)$ for the various Hubble types. The results (Silk 1983a) are displayed in Figure 1. If, as I have argued, the old, spherically–averaged over R_e, stellar population bears some memory of the galaxy formation process, then galaxy formation theory must be capable of accounting for the location of the various galaxy types in the (Σ,σ) plane. I also indicate the position occupied by groups and clusters of galaxies, again using luminous mass: this reemphasizes the point that galaxies are well separated in binding energy and in density from large–scale structures.

A simple theoretical explanation of the general features of Fig. 1 is readily available. As noticed by Faber (1982) and by Gunn (1982), the condition that a uniform, spherically symmetric collapsing cloud can cool within a collapse time translates into a critical density at any given temperature, or equivalently, virial velocity dispersion which separates the regime occupied by galaxies from that occupied by clusters. Rees and Ostriker (1977) and Silk (1977) had originally applied this condition to deduce a critical mass of $\sim 10^{11}$-10^{12} $M_\odot$, which cannot be exceeded by any dissipative cloud.

This discussion was generalized by Silk (1984) to the following situation. Consider an ensemble of clouds interacting dynamically in the potential well of the protogalaxy. A necessary, although not sufficient, condition for dissipation of bulk orbital energy and star formation to occur is that the shocks induced in cloud–cloud collisions be radiative. The post–shock temperature is $kT = (3/16) m_p V_s^2$, where the shock velocity is due to virial motions in the galaxy potential well, namely $V_s = 6^{1/2}\sigma$. The pre–shock cooling time is $t_c = \rho^{-1} f(T)$, where f(T) is the ratio of thermal energy density to cooling rate per gm and ρ is the mean cloud density. For the shock to be radiative, the cloud column density in a one–dimensional shock must exceed $\Sigma \equiv \rho v_s t_c$, where Σ_{cool} is a function only of σ. The formation of a dense layer in a radiative shock does not guarantee instability; in fact, three–dimensional compression of the cooled layer is necessary to ensure gravitational instability and fragmentation. But cooling is surely necessary to initiate this process.

Σ_{cool} is displayed in figure 1 for a primordial mixture of abundances. H_2 cooling has been neglected, and the steep rise of Σ_{cool} below $\sigma \sim 10$ km s^{-1} is due to suppression of Lyman alpha cooling. At higher σ, hydrogen and helium cooling by excitations and recombination is important, and eventually bremsstrahlung dominates. Evidently, Σ_{cool} cleanly separates galaxies, where dissipation occurred during formation, from groups and clusters, whose binding energy was evidently acquired with little or no dissipation. The theoretical justification for this conclusion is that once stars formed, dissipation effectively ceased.

How plausible is it that colliding clouds interact dissipatively and trigger star formation? While it is not straightforward to demonstrate this on theoretical grounds, for we are far from a complete theory of star formation, there are strong indirect arguments that support this outcome. Massive stars form contagiously in the spiral arms in giant molecular cloud complexes. The formation of these complexes from smaller clouds appears to be stimulated by coagulation in the spiral density wave (Kwan and Valdes 1983; Tomisaka 1984), and prolific star formation occurs at cloud masses above 10^5-10^6 $M_\odot$. It is likely that formation of OB stars is self-reinforcing, and smaller clouds, present throughout the disk, of mass 10^3-10^4 $M_\odot$ form too few OB stars for efficient formation of massive stars to occur (Silk 1986). Tidal interactions between nearby galaxies are similarly capable of inducing vigorous bursts of massive star formation (de Jong 1985). Again, enhanced collisions between molecular clouds seem the most likely mechanism (Icke 1985). Gas responds inelastically to any tidal perturbation with a greatly amplified response. The motivation for choosing cloud masses is also rather indirect. HI–H_2 cloud complexes in our own galaxy and in M31 have masses of 10^6 or even 10^7 $M_\odot$. Clumpiness seen in "young" galaxies (Lequeux and Viallefond 1980) and in at least one intergalactic cloud (Terzian *et al.* 1986) suggests that primordial building blocks may have masses of $\sim 10^7$ $M_\odot$. Then there are the cosmological arguments: in a cold dark matter–dominated universe, while the dark matter clumps on all scales, baryonic cooling via H_2 formation and infall only occurs for clouds of mass in excess of $\sim 10^6$ $M_\odot$ (Bond and Szalay 1983). Even pancake fragmentation results in formation of baryonic fragments with mass $\lesssim 10^8$ $M_\odot$ (Sunyaev and Zel'dovich 1972).

One would expect an aggregation of many thousands of clouds to relax dynamically, much as van Albada (1982) found in his N–body simulations. Only in the core, where collisions occur within a dynamical time, would dissipative effects dominate and enhance the density. Clouds should survive in the outer regions, and coalesce or disrupt in the core, where star formation will be enhanced. This argument suggests that the mean surface density of a galaxy, if clouds are presumed to form stars mostly in the collision–dominated core, will be of order the mean surface density of the clouds, since the ratio of cloud–cloud collision to crossing times is of order the ratio of cloud to galaxy (averaged over cloud orbit) surface densities (Silk and Norman 1980). This takes us one step back in time from present luminous cores to protogalactic clouds, but a crucial question is, what determined the longevity of, and gas density in, these clouds? For if they were to have collapsed and fragmented into stars on a free–fall time scale ($\sim 10^6$ yr), practically all the gaseous dissipation would have terminated before a massive galaxy could have formed over the much longer time–scale of 10^8–10^9 yr. And moreover, knowledge of the cloud surface density is crucial to understanding the location of the old components of galaxies in the (Σ,σ) diagram. The key, it will now be argued, lies in understanding star formation in primordial clouds.

IV. PRIMORDIAL STAR FORMATION AND PROTOGALACTIC EVOLUTION

The best lever on the IMF of primordial stars, that is to say, the first generation of stars in a cloud of primordial abundance, is by inference from study of extreme population II stars. Heavy element abundance ratios in these stars are consistent with a precursor population of ordinary stars of intermediate to high mass (5–100 $M_\odot$). The fact that there is at least one star known (Bessell and Norris 1984) with $[Z] \approx -4.5$ suggests that the primordial IMF extended to below 1 $M_\odot$, for even a single generation of star formation should last long enough, by analogy with our own interstellar medium, for there to be some mixing of heavy elements at the 10^{-4} $Z_\odot$ level. Certainly from the theoretical viewpoint insofar as star formation is concerned, 10^{-4} $Z_\odot$ and zero metallicity should be indistinguishable: fragmentation and opacity effects only change appreciably at higher values of Z.

The number of surviving primordial or "population III" stars is very small. Hence if they did possess a normal, that is to say, solar neighborhood IMF, as is consistent with most other well–studied galaxies, very few must have formed. Indeed, the observations suggest that $dN/dZ \approx$ constant for $Z \lesssim 0.01$ $Z_\odot$ (Beers *et al.* 1985). This means that provided the star formation efficiency was initially very low, and recycling of gas ceased once Z rose to ~ 0.01 $Z_\odot$, one could readily account for the abundances of heavy elements in population II with a precursor population of stars spanning the 0.1 to 100 $M_\odot$ mass range. Alternative possibilities are that the yield of heavy elements was greatly enhanced in metal–poor stars (Jura 1985), that there was considerable and continuing infall of primordial material, as envisaged in some galactic disk models, or of course, that the population III IMF consisted exclusively of massive stars (Truran and Cameron 1971). There is no observational motivation to believe that population III might have primarily consisted of very massive objects or of Jupiter–mass bodies.

Theory, of course, is eminently qualified to give any answers whatsoever to the issue of the mass range of population III. Indeed, one can find suggestions in the literature that range from 10^{-3} $M_\odot$ up to 10^6 $M_\odot$. However the most naive theoretical approach concludes that the mass range of primordial stars should be similar to that of stars forming at present, with some possibility of enhancement of massive OB star formation. This arises from consideration of the role of cooling by molecular hydrogen during the fragmentation of clouds of primordial abundance. The H_2 initially forms via H^- or H^+ production from the residual post–recombination ionized fraction, followed by reaction with H to form H_2 at high density, when three–body formation of H_2 dominates. Consideration of the minimum Jeans mass attained during opacity–limited fragmentation of a spherically symmetric collapsing cloud suggests that, just as with conventional interstellar matter opacities, fragmentation proceeds to below 0.1 $M_\odot$ (Palla *et al.* 1983). If H_2 is destroyed in shocks, as might happen in cloud–cloud collisions in the potential well of a moderately massive protogalaxy, then with only Lyman alpha cooling, the fragmentation analysis favors predominantly massive (10–100 $M_\odot$) star formation (Silk 1985).

Hence it seems that star formation in a protogalaxy must result in formation of some massive stars, HII regions, and supernovae. This means that dynamical feedback of energy input from massive stars will inevitably accompany chemical enrichment by their ejecta. Now in a closed system, with no mass loss or infall, the metallicity asymptotically approaches the net yield, or about twice the solar metallicity. The existence of metal–poor galaxies, combined with the fact that many of these, the dE's, have very low surface brightness compared to luminous E's, strongly suggests that considerable mass loss has occurred. The existence of a correlation between Z and σ also supports the idea that chemical and dynamical evolution have been coupled together during the protogalactic phase.

A simple means of understanding this coupling arises if the stellar energy input from massive stars via winds and supernovae is sufficient to regulate the protogalactic gas density. In a quasispherical protogalaxy, the mean gas density would be similar to that in our present, relatively gas–poor, galactic disk, and the supernova rate per unit mass of gas which is forming stars may even be enhanced. If supernovae are capable of regulating the interstellar medium and star formation in our own galaxy, they should be equally capable of managing this task in a protogalaxy. Collapse induces more star formation, for example by enhancing the rate of cloud collisions, which in turn yield more supernovae that help support the gas, on the average, against further collapse. All of this can happen even when the mean cooling time t_c for the system as a whole exceeds the dynamical time–scale t_d, since stars can form within individual clouds, but one would expect the rate of dissipation and accompanying star formation to increase dramatically once $t_c < t_d$. Figure 1 demonstrates that collapse inevitably brings a protogalaxy into this regime where one might expect bursts of star formation to be initiated.

A very simplified sketch of how self-regulation of the protogalactic gas density occurs may be constructed as follows. Let λ denote the energy input per gm of gas that forms stars associated with supernovae: for a standard IMF, $\lambda \approx 10^{51}$ ergs/100 $M_\odot \approx (700 \text{ km s}^{-1})^2$. Then in order for the supernova energy input, neglecting radiative losses, to be balanced by cooling, one must have

$$\lambda \dot{M}_* = M_{gas} \tfrac{3}{2} \sigma^2 t_c^{-1} \tag{1}$$

https://doi.org/10.1017/S0074180900150466 Published online by Cambridge University Press

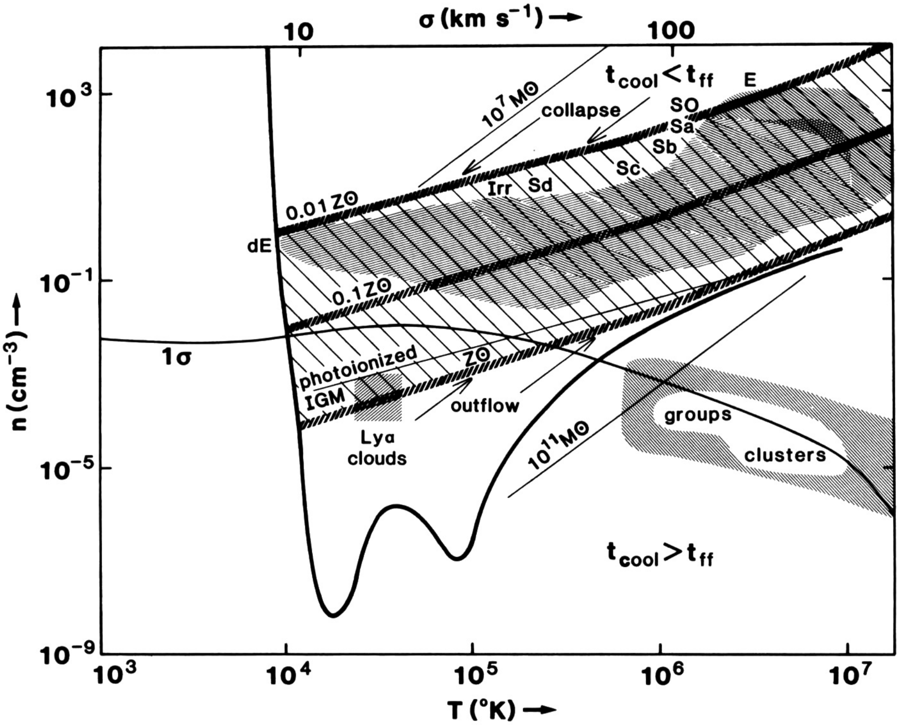

Figure 2. Constraints on protogalactic evolution. In this plot of baryon density versus virial temperature (or equivalently, velocity dispersion), the cooling curve for a collapsing protogalactic cloud or primordial abundance demarcates the region where cooling occurs rapidly or slowly with respect to a dynamical collapse time-scale. This curve reduces to the line marked "photoionized IGM" if the intergalactic medium is highly ionized, as is the case after quasars turn on at z = 2 or 3. The hatched area indicates the density range over which supernova-driven energy input can support a gaseous protogalaxy against collapse for several values of initial metallicity at onset of starburst. The various data points for galaxies and groups are taken from Figure 1; the region occupied by the Lyman alpha clouds seen in absorption against high redshift quasars is also shown. From Silk (1985).

where $\dot{M}_*$ is the average star formation rate, M_{gas} is the gas mass, and the mean cooling time–scale $t_c = \frac{3}{2}kTn^{-1}f^{-1}$, with

$$f \sim AZT^{-1/2} \quad \text{for} \quad Z \gtrsim 0.01 Z_\odot, \quad 10^5 \lesssim T \lesssim 10^7 K.$$

Integrating (1) subject to the condition that $\dot{M}_* = -\dot{M}_{gas}$ yields, after one dynamical time has elapsed, the mean initial gas density

$$n \approx 0.01 \sigma_{100}^{2.4} (Z_\odot/Z)^2 cm^{-3}, \tag{2}$$

where $\sigma_{100} \equiv \sigma/100$ km s^{-1}. At this stage, relatively little of the initial gas reservoir has been converted into stars. In fact, from (1), it is apparent that the gas supply is only exhausted if no mass loss occurs after a time of order

$$(\lambda/\sigma^2)t_c \sim 100 \sigma_{100}^{-2} t_d. \tag{3}$$

This can mean that the protogalactic phase takes ~ 10 crossing times or $\sim 10^9$ yr in a typical massive elliptical galaxy.

Self–regulation becomes effective after roughly one dynamical time has elapsed. Condition (2) then yields the mean star density attained during the self–regulation phase, with some uncertainty depending on how much enrichment has occurred during the initial star burst. The higher the enrichment, the greater is the cooling rate and therefore the initial gas density that can be supported against collapse. Comparison of the predicted density range with observations of the density of the old stellar components of galaxies is shown in figure 2. The specific binding energy of galaxies can evidently be understood by the simple hypothesis of self–regulation of the gas density by supernova energy input. One can make this comparison a little more precise by assuming that the enrichment acquired after one dynamical time is incorporated into the ambient gas and determines its subsequent cooling efficiency. Now at t_d, a mass fraction

$$Z \equiv y\sigma^2/\lambda \equiv 0.01 \lambda_{100}^2 \tag{4}$$

of heavy elements has been produced. Combination of (2) and (4) yields a dependence for the mass of a galaxy approximately as σ^4. Provided the normalization is adjusted to make some allowance for dark matter, and an IMF emphasizing massive stars is used, this accords well with the Faber–Jackson relation.

There is a further noteworthy implication of this model. If the protogalaxy potential well is sufficiently shallow, the supernova remnants can drive a wind and expel most of the gas. The condition for this to occur is that the expanding remnants overlap before decelerating below the protogalaxy velocity dispersion (Bregman 1978). In fact, a wind is driven if $\sigma \lesssim$ 60 km s^{-1}, with remarkably little sensitivity to other parameters: for example, this critical velocity dispersion varies only as $n^{1/22}$ and as $\lambda^{3/11}$ (Dekel and Silk 1985).

The mass loss from low σ galaxies has two consequences. The metallicities will be determined by (4). Indeed, $Z \propto \sigma^2$ fits the metallicity over a wide range of elliptical galaxies, from dE's to luminous E's (Figure 3). A one–dimensional velocity dispersion of 60 km s^{-1} corresponds to $M_B \approx$ -18, and a plot of the nitrogen–to–sulfur abundance ratio reveals an intriguing change of slope at this magnitude: below $M_B \approx$ -18, it is flat; above this magnitude, it is linear. Wyse and Silk (1983) argue this may be indirect evidence for mass loss and failure to recycle stellar debris in low luminosity galaxies. A second consequence is that the drastic mass loss initiated at t_d will leave behind a remnant galaxy of greatly reduced binding energy. If the remnants are to be identified with dE's, then it is necessary to introduce sufficient dark matter to prevent overexpansion (Dekel and Silk 1985).

In fact, there is a natural prescription for producing dE's with the observed range of surface brightness in a cold dark matter–dominated universe. Cold dark matter clusters as $M_H \propto R^{6/(5+n)}$ where n is the index of the initial power spectrum prior to galaxy formation and M_H is the halo mass on the newly bound scale R. Now on small scales, comparable in mass to dwarf galaxies, n $\approx$ -3, and so

$$M_H \propto R^3. \tag{5}$$

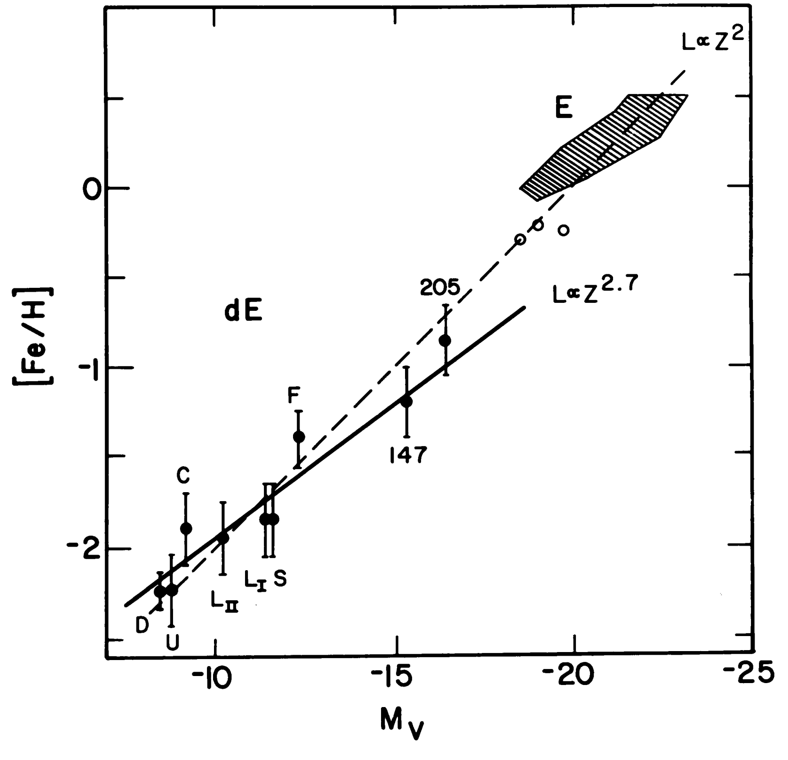

Figure 3. Metallicity versus luminosity for the local dwarf ellipticals and for "normal" ellipticals. Predictions are shown for the wind–driven mass loss model discussed in the text. From Dekel and Silk (1986).

Also, supernova heating is inefficient in dwarf galaxy potential wells, since remnants undergo strong radiative losses below an expansion velocity of about 100 km s^{-1}. Hence allowing for this inefficiency, the cumulative energy input from supernova, taken to be proportional to the stellar mass, is

$$L_* \propto M_{gas}\sigma \propto M_H\sigma, \tag{6}$$

for a universal gas–to–dark matter ratio. Finally, after mass loss, the stars are bound by the halo, and

$$\sigma^2 \sim M_H/R. \tag{7}$$

Combining (5), (6) and (7) yields

$$L_* \propto R^4 \propto \sigma^4 \quad \text{and} \quad M_H/L_* \propto L_*^{-1/4}. \tag{8}$$

In other words, the halos that naturally arise in the cold dark matter theory yield a run of surface brightness with luminosity similar to that observed for dE's (Figure 4), and predict a systematic increase of mass–to–light ratio for

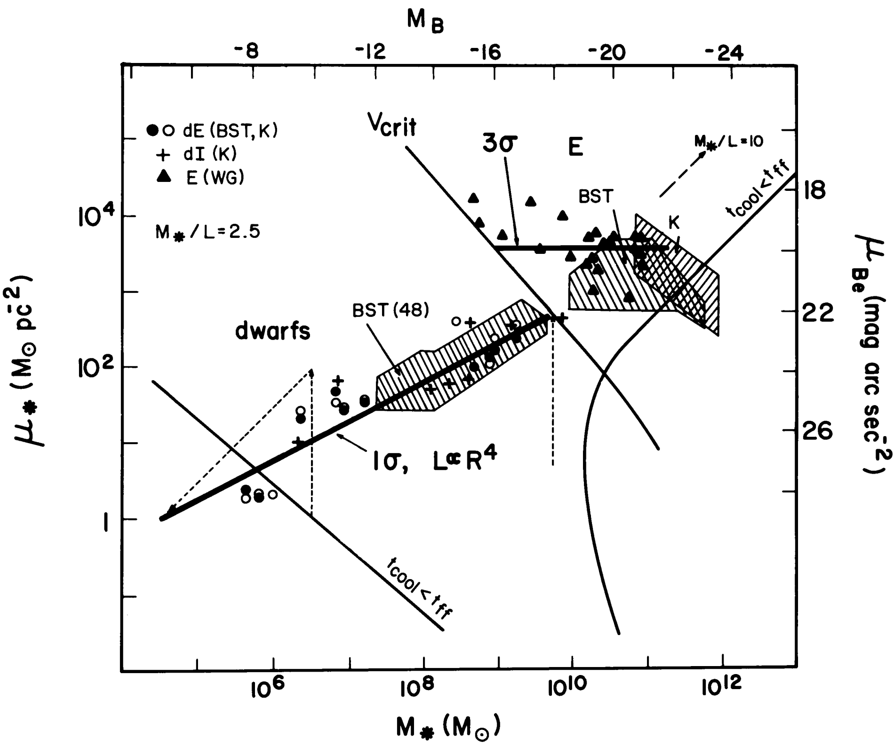

Figure 4. Surface–brightness within the effective radius plotted versus luminosity, for a compiled sample of dwarf ellipticals (in the Local Group and in Virgo), dwarf irregulars, and ellipticals. Predictions are shown for wind–stripping in a cold dark matter–dominated universe. From Dekel and Silk (1986).

https://doi.org/10.1017/S0074180900150466 Published online by Cambridge University Press

the lowest luminosity dE's. The data fit these correlations when the cold dark matter theory is crudely normalized, although the low mass–to–luminosity ratio of the Fornax dwarf galaxy (Cohen 1983) remains somewhat puzzling.

V. TO BIAS OR NOT TO BIAS

Dwarf stripping solves an outstanding problem of the cold dark matter scenario for galaxy formation. If $\Omega = 1$, and this is the principal motivation for hypothesizing cold dark matter, one has to start with the universal mass–to–luminosity ratio

$$M/L_B \approx 1500\Omega h.$$

Then in a theory in which clustering occurs on all scales, one has to produce M/L_B values lower by a factor of 10 in galaxy clusters, groups and dark halos. This is accomplished, in principle, by arguing that the efficiency at which baryonic matter forms galaxies is enhanced by a factor of 10 or so in clusters and on smaller scales. If $\Omega_b/\Omega = 0.1$, then this enhancement explains the origin of the observed ratio of luminous to dark matter of 1/10 in galaxy clusters, groups and luminous galaxies (Faber 1982; Gunn 1982).

An additional bonus is that the clustered fluctuations are predominantly the rare peaks in an initially gaussian fluctuation spectrum. If only these form luminous galaxies, then the required amplitude of the average, less clustered, fluctuations can be lowered by about a factor of ν^2 (if one forms the observed luminous galaxies from $\sim \nu\sigma$ fluctuations) (Bardeen 1985; Kaiser 1985). Consequently, if $\nu \approx 3$, the dynamically measured value of Ω, on the scale of clusters or even superclusters, is lowered to about 0.1, while the true Ω, which samples matter that is uniform over these scales, is unity. Biasing not only solves these problems, but also yields acceptable correlation functions and peculiar velocities for the large–scale galaxy distribution (Davis *et al.* 1985). What remains unclear is *why* galaxy formation only occurs efficiently in the rare peaks of the primordial fluctuation spectrum.

Dwarf stripping by winds provides a physical biasing mechanism. Inspection of figure 2 reveals that 1σ fluctuations are caught between two limits. In order to be able to be in the dissipative regime ($t_c < t_d$), they must have velocity dispersion below ~ 100 km s^{-1}; however wind–stripping becomes effective in this range of velocity dispersion. Only the rare fluctuations find a niche in the (n,σ) plane, where $t_c < t_d$ and $\sigma > 100$ km s^{-1}, allowing the formation of luminous galaxies that have undergone efficient star formation by retention of the bulk of their initial gas supply. This result is displayed in Figure 4, which plots surface brightness (or mass density) versus luminosity (or mass). The unstripped ellipticals form a distinct sequence from the stripped systems, which have expanded as argued previously to form low surface brightness dE's. Wind–stripping can explain the bifurcation in the sequence of elliptical galaxies, in the context of a cold dark matter scenario for the initial fluctuations and halo formation. A consequence of this interpretation is that the dE's, which formed from the 1σ fluctuations, are the only true mass tracers in an $\Omega = 1$ universe.

6. CONCLUSIONS

The cold dark matter theory of galaxy formation by hierarchical clustering appears increasingly attractive in its explanations of many characteristics of galaxies and of galaxy clustering. Nevertheless, it also has serious flaws that generic pancake fragmentation theories are especially adept at explaining, despite their own weakness in other areas. This suggests to me that the ultimate theory of galaxy formation is going to contain elements of both approaches. Indeed, one lesson to be learned from the discussion presented here of the (n,σ) or (Σ,σ) diagram is that it is relatively easy to mask the initial conditions set by cosmology. Physical processes, especially involving star formation, are sufficiently ill understood that they may well play a dominant role in determining the principal characteristics of galaxies.

Cosmology is more intimately connected to the non–dissipative dark matter component, but here there is considerably more freedom for speculation because of inevitable uncertainty about its spatial distribution. Eventually, the dissipation diagram could become the Hertzsprung–Russell diagram of extragalactic astronomy, with evolution tracks linking cosmological initial conditions to fossilized properties of galaxies that were frozen at the end of the epoch of galaxy formation. For now, however, galaxy formation modellers would be well advised to concentrate more on accounting for galaxy characteristics by utilizing models involving dynamical relaxation and dissipation, and less on fidelity towards specific theories based on cold, warm, hot, or even no dark matter.

REFERENCES

Aaronson, M. 1986, these proceedings.
Aaronson, M, Mould, J. and Schechter, J. 1985, private communication.
Bahcall, N. 1981, *Ap.J.* **247**, 787.
Bahcall, N. A. and Soneira, R. M. 1983, *Ap.J.* **270**, 20.
Bardeen, J. M. 1985, in *Inner Space—Outer Space*, eds. E. Kolb and M. Turner (Chicago:University of Chicago Press) (in press).
Barnes, J. 1985, Ph.D. thesis, U.C. Berkeley (unpublished).
Beers, T. C., Preston, G. W. and Shectman, S. A. 1985, preprint.
Bessell, M. S. and Norris, J. 1984, *Ap.J.* **285**, 622.
Blumenthal, G., Faber, S., Primack, J. and Rees, M. 1984, *Nature* **301**, 584.
Bond, J. R., Efstathiou, G. and Silk, J. 1980, *Phys. Rev. Lett.* **45**.
Bond, J. R., Szalay, A. S. and White, S. D. M. 1983, *Nature* **301**, 584.
Bond, J. R. and Szalay, A. S. 1983, *Ap.J.* **274**, 443.

Bregman, J. 1978, *Ap.J.* **224**, 768.
Carlberg, R. 1984, *Ap.J.* **286**, 403.
Carr, B. J. and Ikeuchi, S. 1985, *M.N.R.A.S.* **213**, 497.
Cohen, J. G. 1983, *Ap.J. Letters* **270**, L41.
Cowsik, R. 1985, preprint.
Davies, R. L., Efstathiou, G., Fall, S. D. M., Illingworth, G. and Schechter, P. L. 1983, *Ap.J.* **266**, 41.
Davis, M., Efstathiou, G., Frenk, C. S. and White, S. D. M. 1985, *Ap.J.* **292**, 371.
de Jong, T. 1985, *Proc. Erice School on Spectral Evolution of Galaxies* (in press).
Dekel, A. and Silk, J. 1986 (in press).
Dressler, A. 1980, *Ap.J.* **236**, 351.
Dressler, A. 1986, in *Spectral Evolution of Galaxies*, ed. C. Chiosi and A. Renzini (D. Reidel:Dordrecht).
Einasto, J. 1986, these proceedings.
Faber, S. M. 1982, in *Astrophysical Cosmology*, ed. H. A. Brück, G. V. Coyne and M. S. Longair (Vatican:Pontificia Academia Scientrarum), p. 219.
Faber, S. and Gallagher, J. 1979, *Ann. Rev. Astr. Ap.* **17**, 135.
Faber, S. M. and Jackson, R. F. 1976, *Ap.J.* **204**, 668.
Fabian, A. C., Nulsen, P. E. J. and Canizares, C. R. 1984, *Nature* **311**, 733.
Fall, S. D. M. 1979a, *Rev. Mod. Phys.* **51**, 21.
Fall, S. D. M. 1979b, *Nature* **281**, 200.
Fall, S. D. M. and Efstathiou, G. 1980, *M.N.R.A.S.* **193**, 189.
Fall, S. D. M. and Rees, M. J. 1985, *Ap.J.* (in press).
Farouki, R. T., Shapiro, S. L. and Duncan, M. J. 1983, *Ap.J.* **265**, 597.
Gunn, J. E. 1977, *Ap.J.* **218**, 592.
Gunn, J. E. 1982, in *Astrophysical Cosmology*, ed. H. A. Brück, G. V. Coyne and M. S. Longair (Vatican:Pontificia Academia Scientiarum), p. 233.
Icke, V. 1985, *Astr. Ap.* **144**, 115.
Ikeuchi, S. 1981, *P.A.S.J.* **33**, 211.
Jura, M. 1985, preprint.
Kaiser, N. 1983, *Ap.J. Letters* **273**, L17.
Kaiser, N. 1985, in *Proc. Inner Space–Outer Space Symposium*, eds. E. Kolb and M. Turner (Chicago:University of Chicago Press) (in press).
Kron, R. G., Koo, D. C., and Windhorst, R. A. 1985, *Astr. Ap.* **146**, 38.
Kirshner, R. P., Oemler, A., Schechter, P. L. and Shectman, A. 1981, *Ap.J.* **248**, L57.
Kunth, D. and Sargent, W. L. W. 1985, preprint.
Kwan, J. and Valdes, F. 1983, *Ap.J.* **271**, 604.
Larson, R. 1975, *M.N.R.A.S.* **173**, 671.
Lequeux, J. and Viallefond, F. 1980, *Astr. Ap.* **91**, 269.
Melott, A. 1983, *M.N.R.A.S.* **202**, 595.
Mould, J. 1984, *P.A.S.P.*, **96**, 773.

Ostriker, J. P. and Cowie, L. 1981, *Ap.J. Letters* **243**, L127.
Palla, F., Stahler, S. and Salpeter, E. E. 1983, *Ap.J.* **271**, 632.
Peebles, P. J. E. and Dicke, R. 1968, *Ap.J.* **154**, 891.
Peebles, P. J. E. 1984, *Ap.J.* **277**, 470.
Rees, M. J. and Ostriker, J. P. 1977, *M.N.R.A.S.* **179**, 541.
Rubin, V. R. and Burstein, D. 1985, *Ap.J.* (in press).
Schaeffer, R. and Silk, J. 1985, *Ap.J.* **292**, 319.
Silk, J. 1968, *Ap.J.* **151**, 459.
Silk, J. 1977, *Ap.J.* **211**, 638.
Silk, J. 1983a, *Nature* **301**, 574.
Silk, J. 1983b, *M.N.R.A.S.* **205**, 705.
Silk, J. 1984, *The Big Bang and Georges Lemaître*, ed. A. Berger (D. Reidel:Dordrecht), p. 279.
Silk, J. 1985, *Ap.J.* (in press).
Silk, J. 1986, in *Luminous Stars and Associations in Galaxies*, I.A.U. Symposium **116**, ed. C. de Loore, A. Willis and P. Lazkarides (Dordrecht:D. Reidel)
Silk, J. and Norman, C. 1981, *Ap.J.* **247**, 59.
Sunyaev, R. A. and Zel'dovich, Ya. B. 1972, *Astr. Ap.* **20**, 189.
Terzian, Y., Schneider, S. E. and Salpeter, E. E. 1986, in *Structure and Evolution of Active Galactic Nuclei*, ed. G. Giuricin, F. Mardirossian, M. Mezzetti and M. Ramella (D. Reidel:Dordrecht)
Tohline, J. 1980, *Ap.J.* **239**, 417.
Tremaine, S. and Gunn, J. E. 1979, *Phys. Rev. Lett.* **42**, 407.
Truran, J. and Cameron, A. G. W. 1971, *Astrophys. Space Sci.* **14**, 179.
Tully, R. B. and Fisher, J. R. 1977, *Astr. Ap.*, **54**, 661.
van Albada, T. 1982, *M.N.R.A.S.* **201**, 939.
Vittorio, N. and Silk, J. 1985, *Ap.J. Letters* **293**, L1.
White, S. D. M. 1986, these proceedings.
White, S. D. M., Davis, M. and Frenk, C. 1983, *Ap.J. Letters* **274**, L1.
White, S. D. M., Huchra, J., Latham, D., and Davis, M. 1983, *M.N.R.A.S.* **203**, 701.
White, S. D. M. and Rees, M. J. 1978, *M.N.R.A.S.* **183**, 341.
Wyse, R. G. and Silk, J. 1985, *Ap.J. Letters* (in press).
Zel'dovich, Ya. B. 1970, *Astr. Ap.* **5**, 84.

DISCUSSION

RUBIN: It is nice to see your theory, which couples dynamical and chemical evolution. But you should be warned that the break in the slope of the [NII]/[SII] data may be an artifact of the observations. For high-luminosity galaxies, the values refer to the disk, excluding the nucleus. For the low-luminosity objects, the values generally come from a single aperture measurement centered on the nucleus. It is not clear how the plot would look with only disk data at all luminosities.

AUDOUZE: I would like to comment briefly about the concept of primary versus secondary elements. It might be more appropriate to speak about elements which are produced by high-mass stars and which look "primary" while those which are mainly produced in low-mass objects seem to be secondary.

SILK: I agree. The distinction I would make is that the wind-stripping model predicts that elements produced by massive stars will not be recycled through successive generations of star formation in galaxies which are fainter than $M_B \approx -18$, while elements produced by stars of all masses are recycled in the more luminous galaxies.

REES: You emphasised that the energy input from stars and supernovae could exert a negative feedback on the rate of star formation, making it impossible to convert all gas into stars on a protogalactic dynamical timescale. There are then in principle two possibilities: either the original gas is all eventually turned into stars, albeit slowly, or the first generation of stars generates enough energy to expel the bulk of the gas (rather then just slowing down further star formation). Your discussion of dwarf galaxies depended on the latter possibility being true. Is this just an assumption, or is there some physical reason why the negative feedback should "overshoot" as you envisage?

SILK: In a gaseous proto-galaxy whose one-dimensional velocity dispersion is less than about 60 km s^{-1}, I showed that a supernova-driven wind would be inevitable. This assumes that at least one supernova is produced for every several hundred solar masses of newly formed stars. If one could avoid making supernovae, then the evolution would proceed much more gently, until the gas supply is exhausted. Likewise, in massive proto-galaxies with deeper potential wells, where a wind is effectively quenched, the assumption of a standard initial mass function guarantees enough supernovae to maintain a negative feedback and prolong the duration of the star forming phase.

SPERGEL: I understand that your wind-stripping scenario predicts the existence of objects with a lot of mass but not much luminosity. In general, is it true that you predict a trend where lower-luminosity objects have higher M/L ratios?

SILK: That's right. In fact, it is necessary for this theory that dwarf ellipticals have high M/L ratios.

SPERGEL: So in some dwarf ellipticals, even though you have gotten rid of a lot of the gas, you might expect to have hot gas left behind. Would you predict that dwarf ellipticals might have x-ray halos?

SILK: I expect that the gas would be long gone. In any case, the potential well wouldn't be deep enough for the gas being blown out to have x-ray emitting temperatures.

REES: I'd like to ask at what redshift the first galaxies formed. You mentioned the Spinrad galaxy at z = 1.8 and quoted his argument that from its colors you can infer that it must have formed well before that redshift. I'd like to ask how strong that argument is, and how it depends on the initial mass function and on other assumptions.

SILK: That particular object does not give very strong constraints, because it has strong emission lines. But there is recent work by Hamilton and Kron in which they look at distant ellipticals not chosen from radio-source surveys. These turn out to have standard red populations. The galaxies are observed at redshifts of 0.8 - 0.9. Hamilton and Kron make the standard synthesis models and find rather large ages. The redshift of formation turns out to be at least 5. But let me emphasize that these are very rare, very luminous ellipticals and that the answer is model-dependent.

BERTSCHINGER: The coincidence in your diagram between the position of the cooling curves and the masses and radii of galaxies may be a red herring, particularly in the heirarchical clustering picture. For example, Simon White told us that Frenk et al., in their n-body simulations, were able to get reasonable maximum velocities of rotation curves in a situation where there was no cooling or dissipation. Furthermore, I have a poster paper where I consider just infall onto peaks in a cold dark matter spectrum, and I'm able to reproduce the Schechter mass function, again without any cooling. Cooling is not important in these cases because you are building up small structures first, so the cooling time is always shorter than the dynamical time.

SHAPIRO: I noticed that the cooling curve that you used to determine the range of virialized mass scales which would cool in less than the Hubble time does not take account of inverse Compton cooling by the background radiation. Since the rate of Compton cooling increases very rapidly with increasing redshift and is independent of gas temperature, your conclusions would be very different at high z if you included it. Have you assumed that all of the mass scales of interest condense out at redshifts $z \ll 10$? If so, you would be self-consistent, although I would question whether your timescale arguments are as general as they might be.

SILK: In both cold dark matter scenarios and fragmentation scenarios, formation of luminous galaxies occurs at a redshift less than 10 when Compton cooling is unimportant. In the former picture, small mass scales can begin to condense out somewhat earlier, but Compton cooling is irrelevant, for the following reason. The dark-matter potential wells define the appropriate mass scales, and once sufficient dissipation and density enhancement of the baryonic component occur, local cooling processes dominate. There are, of course, more exotic possibilities, such as the explosive amplification model of galaxy formation, in which Compton cooling plays an important role in determining the relevant fragmentation scales. My arguments could be generalized to study fragmentation and dissipation at large redshift, although I have not done so.

DRESSLER: In your scorecard comparing the hierarchical and fragmentation models, you gave an edge to the latter for ease of producing a correlation of galaxy morphology and environment. I'd appeal that score on the grounds that once you have a mechanism that preferentially produces spheroids in dense regions, such as biased galaxy formation with cold dark matter, this may provide the primary ingredient in galaxy differentation. Evidence for this is the fact that Sa or SO galaxies are, at least to first order, identical in and out of clusters, which suggests that partitioning the baryons between spheroid and disk may be more important than later environmental effects.

SCHECHTER: Joseph, I was jealous of your colorful viewgraphs and so dazzled by them that I didn't notice that the ellipticals are 5σ away from the cooling curves. Can you tell us how they got there?

SILK: There was a lot of dissipation as the galaxies formed.

SCHECHTER: So the number of σ away from the curves is a measure of how much the space density has evolved.

SILK: Yes, although the amount of evolution depends on the theory - on whether you think it took a 1σ or a 3σ fluctuation to make a galaxy.

PRIMACK: I'd like to comment on that. As we pointed out in our paper in Nature, it is misleading to use these pictures in which you show 1σ and 3σ tracks if you are trying to figure out where a galaxy originally came from. The problem is that there is some evolution in temperature, i.e., in (velocity dispersion)^2, as well as a great deal of evolution in density. If you plotted mass *versus* temperature, then the ellipticals would look like 3σ fluctuations when you normalize things so that typical spirals are 1σ fluctuations.

SILK: I don't think that masses of ellipticals are known well enough to do that.

E. TURNER: In the simple biased galaxy formation models, you have a single free parameter: the number ν of sigma you associate with the threshold for the formation of a visible galaxy. It must be adjusted to give 1) the correct galaxy number density, 2) the right fraction of the total (inflationary) Ω=1 associated with galaxies, 3) the right galaxy covariance function, 4) the correct depth of galaxy potential wells,

https://doi.org/10.1017/S0074180900150466 Published online by Cambridge University Press

etc. Can this be done with a single value for the threshold? If not, is there any obvious second parameter you would call on?

KAISER: To get the required enhancement of clustering of galaxies relative to the matter (requirement 3), you need $\nu^2 \simeq 1.7\ (1+z_f)$, where z_f is the redshift of galaxy formation. Only for a moderately recent formation epoch, $z_f \simeq 3 - 5$, say, can the other requirements be satisfied. The most stringent requirement is that in rich clusters, where galaxy formation was by assumption ~ 5 times more efficient than average, we know that at least ~ 10% of the gas turned into galaxies. This means that the global efficiency is then at least 2%, implying $\nu \simeq 3$. For this choice of parameters, our estimates of number density and potential depth do not seem to be grossly unacceptable.

GUNN: I'd like to comment on the question of how much of the mass in the Universe is in galaxies, and of how this is related to the threshold ν for galaxy formation. Nick Kaiser and I have a small disagreement about this. I think that these questions are not necessarily closely connected. Suppose you assume that the mass is distributed around the high-ν peaks like the correlation function at that epoch. Then you can easily show that a large fraction ~ 70% of the time, if you have a 3ν fluctuation of $10^9\ M_\odot$, it will typically bind $10^{12}\ M_\odot$ to itself. Even though these 3ν fluctuations involve only 0.1% of the mass in the Universe, they each scoop up 1000 times their own mass. So you can gather up nearly all of the material around the 3ν peaks. I've made this argument just at one mass scale. I know that the real world is more complicated than this, but some such phenomenon must exist.

MELOTT: In response to the concerns voiced by Ed Turner and Jim Gunn, I'd like to report some results recently found by Jim Fry and me. In cold dark matter simulations, we calculated two-point and three-point correlations for biased sets. I should stress that this was done with a large smoothing window, corresponding to about $10^{12}\ M_\odot$, and that for the first time the correlations were calculated in a mass-weighted way, rather than including all connected regions with equal weight. We found that a simple 1σ bias was sufficient to boost the correlation amplitude to an acceptable level and to simultaneously make $Q \approx 1$, where Q is the reduced amplitude of the three-point function. An interesting extra benefit of this procedure is that voids are much more prominent than when using a larger bias threshold and a small smoothing window.

FABER: I'd like to ask the people who make n-body models whether large-scale structure makes a good case for the cold dark matter model. This is a point I was trying to raise in my introductory talk. Is the problem with the models that the correlation function doesn't stay positive at large enough radii to give the Kaiser effect, or is it that the predicted correlation function for galaxies stays high on large scales while the observed correlation function goes negative?

WHITE: In the cold dark matter picture the mass correlation function is supposed to go negative at a length scale of ~ 70 Mpc, well outside the range you can actually measure in the n-body calculations. The biasing, or Kaiser effect, will only amplify the underlying correlation. So an inconsistency only arises if it is really true that the cluster-cluster correlation function is positive at 100 Mpc; there isn't a problem with the fact that it is positive at 25 Mpc. I'd like to repeat that I don't think our present simulations can tell you much one way or the other about large-scale structure, because the regions of the Universe that we are simulating aren't large enough.

N. BAHCALL: I'd like to make a comment about large-scale structure to partially answer Sandy's question. There is structure seen not only in the cluster-cluster correlation function at large scales like 50 Mpc, but also in the supercluster correlation function. Here we see a positive correlation among superclusters at scales of $\sim$ 100 - 130 Mpc, and the correlation is even stronger than the cluster-cluster correlation. This is of course somewhat uncertain because of the small size of the supercluster sample, but it does show that there is structure at these very large scales. Now there is a problem with the galaxy-galaxy correlation function going negative while the cluster-cluster function stays positive only if you believe the amplification model. Another point of view has been taken by Szalay and Schramm, who look at it as a scale-invariant clustering process. Looking at clusters of galaxies, the correlation strength appears to continuously increase with mean separation of the system for clusters of all richness classes. Now we have shown that the same correlation function exists for the superclusters, when normalized to the mean separation of the systems. In this case the possible slight difference of the galaxy-galaxy correlation function may be due to gravitational perturbations on top of some infrastructure which dominates the large-scale structure.