Mon. Not. R. astr. Soc. (1991) **248,** 332–352

The evolution of groups of galaxies in the Press–Schechter formalism

Richard G. Bower
Department of Physics, University of Durham, Durham DH1 3LE

Accepted 1990 September 7. Received 1990 August 8; in original form 1990 May 31

SUMMARY

The Press–Schechter theory provides a simple analytical description for the evolution of gravitational structure in a hierarchical universe. In this paper, we extend the original theory in order to focus our attention on the subset of regions that will eventually collapse to form clusters of a set mass. The first part of the paper is concerned with obtaining the *conditional multiplicity function* of groups at an epoch with redshift z, given that they are bound into an object of a particular mass at the present epoch. This is combined with the present distribution of group masses in order to obtain the *joint multiplicity function*. The difficulty in obtaining these functions lies in determining the cross-corelation between a set of small volumes and the larger volume that contains them. We show that this correlation has a simple form. The formulae we derive are checked for self-consistency, and are compared with the N-body simulations of Efstathiou *et al.* The numerical results are found to be in extremely good agreement.

These multiplicity functions are applied to study the evolutionary histories of groups and clusters. We obtain a number of results relating to the infall of galaxies in small groups into large clusters, and to the density of the more massive groups at early times. In particular, we illustrate our analysis by making an in-depth comparison of the theoretical evolution with the Butcher–Oemler effect observed in rich clusters at moderate redshifts. We find that the growth of the infall rate over these look-back times is not (by itself) sufficient to explain the rapidly increasing fraction of blue galaxies in these clusters, but that a good quantitative fit to the available data can be obtained if allowance is made for the increased star burst activity that is seen in the spectra of the distant blue galaxies.

1 INTRODUCTION

The gravitational structures that are seen in the Universe today are widely accepted to have grown out of small perturbations in the mean density of the early Universe (e.g. Peebles 1980). The fluctuations may result from quantum mechanical effects at very early times. In most scenarios, for example the standard Cold Dark Matter model (*cf.* Davis *et al.* 1985, and references therein) an appreciable fraction of the power in these fluctuations has a small-scale length. The larger structures, such as galaxies, groups of galaxies and rich clusters, then grow by the agglomeration of smaller units. The merging process acts on all scales so that all the condensations are growing with time. At late times, more and more of the mass becomes bound into larger and larger condensations. Following Peebles, the process is termed *hierarchical clustering.*

It is of great interest to study the size distribution of the condensed masses. In particular, we wish to determine the size distribution of the masses combining to form a cluster of given size at the present-day. In a strict hierarchical visualization, the merging lumps would all be of similar size. As we will show, the merging in fact occurs between lumps covering the whole range of available sizes. Our solution also allows us to discuss the distribution in present-day mass of condensations of a particular size chosen at some previous epoch.

The evolution of a system of self-gravitating particles is best studied using a full N-body computer code. This method, however, has two disadvantages. First, the simulation is limited both by the resolution of the smallest particle size, and by the statistical variations in the properties and histories of the small numbers of massive groups. Secondly, while we would be able to obtain a solution to a particular well-defined problem, without an analytic framework we would gain neither an understanding of the physical processes involved, nor the ability to adapt our solution to a change in the cosmological parameters. Fortunately, N-body simulations started from scale-free initial conditions by

Efstathiou *et al.* (1988, EFWD) showed that the analytic theory developed by Press & Schechter (1974, PS) provides a good quantitative statistical description of the evolution of the distribution of mass amongst groups and clusters of various sizes (this distribution is referred to as the *multiplicity function*). However, the work of PS only predicts the mass distribution in the Universe as a whole. In the present work, we will extend the theory to follow the multiplicity function of the progenitor components of a present-epoch group of given size.

While this paper is principally concerned with providing a mathematical description for the evolution of the clustering hierarchy, the results are of considerable utility to a number of problems in cosmology and extra-galactic astronomy. We may apply the results to gain insight into the formation of clusters. For example, various astronomical problems lead to questions such as: "Is there a particular 'epoch of cluster formation'?"; or "What fraction of the mass of growing clusters is gained from the accretion of large groups as opposed to individual galaxies?" In addition, it is of considerable interest to review the spherical collapse model for galaxy clusters of Gunn & Gott (1972) in the light of our improved understanding of the hierarchical growth. The study is also relevant to problems concerning the evolution of galaxies in different environments. For example, the results we present might be combined with studies of the effects of the environment on galaxy evolution in order to compare the average star formation histories of galaxies in the field with those in rich clusters. At an earlier time, the modulation of the growth of galaxy-sized condensations by larger scale inhomogeneities may have resulted in the differing properties of galaxies in high- and low-density regions of the Universe. Such *natural biasing* in the formation of galaxies has been suggested by White *et al.* (1987) in order to reconcile the observed mass density of the Universe with the critical mass density (i.e. $\Omega = 1$) suggested by grand unified theories of inflation (e.g. Guth 1981). In the present paper, we chose to limit our attention to one simple problem. We consider whether the increased blue fraction of galaxies in rich clustes at moderate redshifts (the Butcher–Oemler effect, Butcher & Oemler 1978) might be explained in terms of an increased infall rate of field galaxies into rich clusters at these redshifts. A detailed consideration of the other issues outlined above will be published subsequently.

An outline of the paper is as follows. In Section 2, we outline the principles of the Press–Schechter formalism that we will use in the subsequent work. Section 3 describes our basic mathematical results. In turn, we derive the *conditional multiplicity function* describing the mass distribution of objects under the condition that they will collapse to form an object of set size at the present epoch, and the *joint multiplicity function* of objects of mass M at a specified epoch that collapse to form part of a group of total mass M' at the present. The behaviour and the component parts of each of these functions are discussed in some detail. In Section 4, we demonstrate the validity of this work in two ways. First, we show that the joint multiplicity function is consistent with the universal (i.e. Press–Schechter) multiplicity function both at the previous epoch and at the present. Secondly, we compare our results with the numerical simulations of EFWD. Extremely good quantitative agreement is found. In Section 5, we apply our results to study the hierarchical growth of groups and clusters. Using the results of the previous sections, it is possible to obtain expressions for the infall rate into proto-clusters as a function of present-day cluster mass. We also consider the density of massive groups formed at early times. In Section 6, the results for the redshift evolution of the infall rate are compared with the observed increase in the fraction of blue galaxies in rich clusters at moderate redshifts (the Butcher–Oemler effect). Our basic mathematical results are concisely summarized in Section 7.

2 THE BASIC PRINCIPLES OF THE PS THEORY

For a detailed mathematical description of the PS theory, we refer the reader to the original paper and EFWD as the workings are a special case of those presented in Section 3. Here we briefly outline the basic principles involved.

At some early epoch, the Universe is imagined to be well-described by an isotropic random Gaussian field of small density perturbations. The phases of fluctuations are random so that the field is entirely defined by its power spectrum (i.e. by the amplitudes of its Fourier components). The initial density variations are small (i.e. $\delta \equiv \Delta\rho/\bar{\rho} \ll 1$) and their evolution, in this regime, is described by the *linear* form of Vlasov's equation. The density perturbations grow with time until they become non-linear (i.e. $\delta \sim 1$), at which point their evolution can no longer be followed by simple analytic techniques. PS circumvent this difficulty by assuming the following.

(i) The evolution of the density perturbations can be traced using linear theory until the clump 'turns around' and breaks away from the universal expansion. For a spherical tophat perturbation (in a Universe of closure density) this occurs at a density contrast $\delta_c \sim 1.68$ (Gunn & Gott 1972).

(ii) At this density contrast, the region collapses rapidly and independently of its surroundings. The internal structure of the clump is lost by violent relaxation (Lynden-Bell 1967). To the rest of the Universe, the region behaves as if it were a single body of large mass.

At the expense of our knowledge of the internal structure of the collapsing clumps, we can therefore continue to apply the linear equation to the 'gas' of particles to calculate their evolving mass distribution. Although this results in a considerable simplification of the equations describing the evolution of the density field, we must still solve the problem of counting the non-linear objects of a given mass. This requires that we are able to relate the probability distribution of the density fluctuations (averaged over a volume V) to the probability that a given region is sufficiently overdense to collapse on the scale V but is not absorbed in the collapse of a larger volume. This probability can then be used to determine the fraction of the mass of the Universe that will be contained in clumps of mass $\bar{\rho}V$. PS suggest a plausible transformation between the probability distribution function of the Gaussian field (smoothed on scale V), $P(\delta_V)$ and the *multiplicity function* of the condensed groups,

$$\rho(M, z) = -2\rho_0 \frac{\partial \mathcal{P}[\delta_V > \delta_c(z)]}{\partial M} dM, \qquad (1)$$

where ρ_0 is the mean density of the Universe, and $\delta_c(z)$ is the

overdensity threshold for collapse at the epoch corresponding to the redshift z. [Note that we use the symbol $\mathscr{P}$ to distinguish the (complementary) cumulative probability distribution of δ_V from its probability density distribution, $P(\delta_V)\,d\delta_V$.] The derivation of the above relation is, however, flawed by the fact that the factor of 2 must be inserted 'by hand' in order that the total mass of the Universe remains constant. This weak point of the theory has been criticized (e.g. Bond *et al.* 1990; Peacock & Heavens 1990), however, the final solution gains credibility from comparison with the N-body simulations of EFWD. We will adopt the Press–Schechter transformation on an empirical basis in the work presented here.

In addition to the approximations made in setting up the PS formalism, two further assumptions will be made in order to simplify the mathematical treatment of this problem.

(i) The Universe is taken to be flat (i.e. $\Omega=1$). This ensures that the density fluctuations grow linearly with the expansion factor. Note, however, that the growth is still approximately linear in an open universe (e.g. Shandarin, Doroshkevich & Zel'dovich 1983) and that it is therefore simple to adapt the Press–Schechter formalism to work in this regime also. The effect is solely to stretch the redshift axis of the evolution.

(ii) The initial power spectrum may be taken to be a power law (i.e. $|\delta_k|^2 \propto k^n$). Since the variance of the density field can be written explicitly as a function of the smoothing scale, this simplifies the mathematical form of the formulae. The reader should note, however, that the formulae presented in this paper may readily be reworked leaving the variance of the density field as a general function of the smoothing scale.

With these assumptions there is only one characteristic scale in the Universe (the expansion factor). The evolution of the mass distribution is therefore self-similar, having the same functional form at each epoch and only altering by a magnification factor, M_*, that determines the 'typical' size of the condensed groups. The actual form that is derived for the *universal multiplicity function* is

$$\rho(M,z)\,dM=\frac{\rho_0}{\sqrt{2\pi}}\left(\frac{n+3}{3}\right)\left(\frac{M}{M_*(z)}\right)^{(n+3)/6} \times \exp\left[-\frac{1}{2}\left(\frac{M}{M_*(z)}\right)^{(n+3)/3}\right]\frac{dM}{M}, \tag{2}$$

where $M_*(z)\propto(1+z)^{-6/(n+3)}$.

The original PS theory deals with the evolution of the multiplicity function of the Universe as a whole. No attempt is made to determine how these masses are distributed in space. EFWD started to extend the basic theory in order to examine how the multiplicity function was altered in regions of large-scale density enhancement. The solution to this problem allows the evolution in a region that condenses to become a rich cluster at the present epoch to be compared with the evolution elsewhere in the Universe. The analytical work of EFWD showed that the most massive groups at early times were incorporated into the most massive groups at the present – a result in qualitative agreement with their N-body simulations. However, this method of analysis is only valid when the correlations between the fluctuations on different scales are negligible. Therefore, the sub-components are required to be very much less massive than the final cluster.

In the present work, we show that these cross-correlations have a simple dependence, and we develop a more sophisticated treatment of the problem that is applicable over the full range of sub-component masses. We are able to derive a *conditional multiplicity function* for the sub-components (at an epoch corresponding to redshift z) of a present-epoch group of given mass. This function is demonstrated to have the required property that it recovers the universal multiplicity function (at epoch z) when combined appropriately with the universal multiplicity distribution of the groups at the final epoch. We also demonstrate that it gives an extremely good *quantitative* fit to the numerical results obtained by EFWD from N-body simulations.

3 THE EVOLUTION OF GROUPS – MATHEMATICL RESULTS

3.1 Conditional probabilities in random Gaussian fields

In the Universe at an early time, we represent the probability distribution of the small variations in the otherwise smooth distribution of matter as a three-dimensional Gaussian random field. In other words, the probability distribution of the density, measured as an average over a randomly selected volume V, is

$$P(\delta_V)=\frac{1}{\sqrt{2\pi\sigma_V^2}}\exp\left(-\frac{\delta_V^2}{2\sigma_V^2}\right),$$

where δ_V is the *overdensity* of the volume $=(\rho_V-\bar{\rho})/\bar{\rho}$, and σ_V is the rms variation of δ_V. Even if the random field is not truly Gaussian, the Central Limit Theorem usually justifies the use of this approximation. The reader should note that the volume V contains a mass $M\approx\rho_0 V$.

In this section, our aim is to determine the probability distribution of the density fluctuations in a volume V that is contained within a larger volume V' in which the average overdensity has a predetermined value. Aside from the requirement that it is entirely contained within V', V is chosen at random. Bayes' formula allows us to express this conditional probability distribution in terms of the *joint* probability distribution of δ_V and $\delta_{V'}$,

$$P(\delta_V|\delta_{V'})=\frac{P(\delta_V,\delta_{V'})}{P(\delta_{V'})}, \tag{3}$$

where $P(\delta_V)$ is the probability that we measure the density inside the volume V (chosen without reference to V') to be δ_V. (For strict mathematical correctness, we should require the overdensity to lie between δ_V and $\delta_V+\varepsilon$, where ε is an arbitrarily small number. We use the looser terminology for the sake of clarity.) $P(\delta_V,\delta_{V'})$ is the probability that we measure the overdensities of V' (chosen at random) *and* V (chosen at random from within V') to be $\delta_{V'}$ and δ_V, respectively; $P(\delta_V|\delta_{V'})$ is the probability that we measure the overdensity inside V to be δ_V given that we already know the overdensity of the large volume V' (that contains V) to be $\delta_{V'}$.

Abstracting from this problem, we consider a pair of Gaussian distributed random variables(x, y) that have a joint probability distribution that is also Gaussian. This is equivalent to stating that the joint probability distribution may be written as

$$P(x, y) = \frac{1}{2\pi}\frac{1}{|\sigma_{ij}|^2}\exp\left[-\frac{1}{2}(\Delta x, \Delta y)\,\sigma_{ij}^{-1}(\Delta x, \Delta y)^T\right], \tag{4}$$

where $\Delta x = x - \bar{x}$, $\Delta y = y - \bar{y}$ and σ_{ij} is the covariance matrix. Using Bayes' formula (equation 3), the conditional probability $P(x\,|\,y)$ can be rewritten as the product of a normalization factor and the exponential term,

$$\exp\left[-\frac{1}{2}\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2\sigma_y^2-\sigma_{xy}^4}\left(\frac{\Delta x}{\sigma_x}-\frac{\Delta y}{\sigma_y}\frac{\sigma_{xy}^2}{\sigma_x\sigma_y}\right)^2\right]. \tag{5}$$

This is a Gaussian distribution with mean

$$\mu_{x|y} = \left(\frac{\sigma_{xy}^2}{\sigma_x\sigma_y}\right)\left(\frac{\sigma_x}{\sigma_y}\right)\Delta y,$$

and variance

$$\sigma_{x|y}^2 = \sigma_x^2 - \left(\frac{\sigma_{xy}^4}{\sigma_y^2\sigma_y^2}\right)\left(\frac{\sigma_x^2}{\sigma_y^2}\right)\sigma_y^2.$$

If the joint probability distribution of δ_V and $\delta_{V'}$ is Gaussian, the above theorem (Adler 1981) can be applied to calculate the conditional probability $P(\delta_V|\,\delta_{V'})$. As a consequence of the Central Limit Theorem, $\delta_{V'}$ and the overdensity of a *particular* volume V have a bivariate Gaussian distribution if the Fourier components of the random density field have independent phases (Bardeen *et al.* 1986). The density fluctuations that arise from quantum effects in an inflationary universe have just this property. The joint distribution of $\delta_{V'}$ and the overdensity of a *randomly chosen* volume V is a weighted sum of the Gaussian distributions for each particular V. Since the means and variances of the individual distributions are similar, the combined distribution is also approximately Gaussian to within a few per cent. Therefore, once we have established the variance and covariance terms

$$\sigma_V^2 = \langle\delta_V^2\rangle, \qquad \sigma_{V'}^2 = \langle\delta_{V'}^2\rangle$$

and

$$\sigma_{VV'}^2 = \langle\delta_V, \delta_{V'}\rangle$$

the conditional probability distribution of a randomly chosen δ_V, given $\delta_{V'}$, is accurately determined.

The variances of δ_V and $\delta_{V'}$ are defined by the power spectrum that we assume for the density fluctuations. In Appendix A, we show that the power spectrum also determines the co-variance, $\sigma_{VV'}^2$. We find that

$$\sigma_{VV'}^2 = \sigma_{V'}^2 \tag{6}$$

The simplicity of this result is striking. However, it arises directly if the mean overdensity of the small volumes V contained in V' is assumed to equal the overdensity of region V' as a whole:

$$\sigma_{VV'}^2 = \langle\langle\delta_V\rangle_V\delta_{V'}\rangle_{V'} = \sigma_{V'}^2,$$

3.2 The conditional probability distribution of density fluctuations

We now apply the conditional probability derived in Section 3.1 to the study of the modulation of density fluctuations on scale V by a field on scale V'. Rewriting equation 5, including the correct normalization factor and using equation 6 we obtain

$$P(\delta_V|\,\delta_{V'})d\delta_V = \frac{1}{[2\pi(\sigma_V^2-\sigma_{V'}^2)]^{1/2}}\exp\left[-\frac{(\delta_V-\delta_{V'})^2}{2(\sigma_V^2-\sigma_{V'}^2)}\right]d\delta_V. \tag{7}$$

It can be seen that this probability distribution has the anticipated mean (i.e. $\langle\delta_V\rangle = \delta_{V'}$), and that in the case that $V' \gg V$ (i.e. $\sigma_{V'} \ll \sigma_V$), it reproduces the 'k-split' approximation used by EFWD (i.e. the distribution of δ_V has the same dispersion everywhere, but the mean level is shifted up or down by the 'background' overdensity, $\delta_{V'}$).

3.3 The conditional multiplicity function

Following EFWD, we proceed to determine $\tilde{F}(M, \delta_c; M', \delta')$, the fraction of regions of mass M, contained within a larger scale region of mass M' and overdensity δ', that are more overdense than the critical density, δ_c (at which non-linear gravitational effects become dominant). Note that we are now working in terms of mass, but that this is directly proportional to the volume of the region, $M \approx \rho_0 V$ since the overdensities are small. We have

$$\tilde{F}(M, \delta_c|\,M', \delta') = \int_{\delta_c}^{\infty} P(\delta_V = \delta\,|\,\delta_{V'} = \delta')\,d\delta \tag{8}$$

$$= \int_{\delta_c}^{\infty}\frac{1}{\sqrt{2\pi(\sigma_M^2-\sigma_{M'}^2)}}\exp\left[\frac{-(\delta-\delta')^2}{2(\sigma_M^2-\sigma_{M'}^2)}\right]d\delta$$

$$= \frac{1}{\sqrt{\pi}}\int_{\frac{(\delta_c-\delta')}{\sqrt{2(\sigma_M^2-\sigma_{M'}^2)}}}^{\infty} e^{-y^2}\,dy$$

These regions rapidly collapse (if they have not already done so) to form small dense objects. A large proportion of them will, however, be consumed in the collapse of a larger region that surrounds them. In Section 4, we will associate these objects with individual isolated galaxies, groups of galaxies or clusters of galaxies depending on their mass. For convenience, we refer to them collectively as *groups*. The spherical collapse models of Gunn & Gott (1972) suggest that the appropriate numerical value for δ_c is 1.68.

As it stands, equation 8 gives the fraction of mass in collapsed groups of size greater than M only at the epoch, z_i, at which the power spectrum has been defined. However, as the fluctuations are small, they evolve linearly (i.e. the evolution of δ_k does not depend on its amplitude). The power spectrum therefore retains its initial shape, and changes only in its normalization. In a flat universe (i.e. $\Omega = 1$), the amplitude grows as $1/(1+z)$. Therefore we may apply equation 8 to calculate $\tilde{F}$ at any arbitrary epoch z by determining the fraction of regions which have overdensity greater than $\delta_c(z) = \delta_c(1+z)/(1+z_i)$ at epoch z_i.

Continuing to follow the derivation of the original Press–Schechter theory, we related $\tilde{F}$ to the distribution of the total mass, M', among collapsed groups of given size by the ansatz (*cf.* equation 1)

$$\tilde{f}(M, \delta_c(z)|M', \delta')\,dM = -2\,\frac{\partial \tilde{F}(M, \delta_c(z)|M', \delta')}{\partial M}\,dM, \qquad (9)$$

where the multiplicity fraction, $\tilde{f}$ is defined as the fraction of the total mass of M' that is contained in collapsed groups with masses between M and $M+dM$. As the integral in equation 8 has been expressed in terms of the dummy variable y, the partial differentiation may be performed explicitly:

$$\tilde{f}(M, \delta_c(z)|M', \delta')\,dM = -\frac{2}{\sqrt{2\pi}}\frac{\sigma_M}{(\sigma_M^2-\sigma_{M'}^2)^{3/2}}(\delta_c-\delta') \qquad (10)$$
$$\times \exp\left[\frac{-(\delta_c-\delta')^2}{2(\sigma_M^2-\sigma_{M'}^2)}\right]\frac{d\sigma_M}{dM}\,dM.$$

As described in Section 2.2, we assume that the initial spectrum of the density fluctuations is a power law with spectral index n (i.e. $|\delta_k|^2 \propto k^n$). From Appendix A, we obtain the variance of the fluctuations as

$$\sigma_M = AM^{-(n+3)/6},$$

where the normalization constant A is determined by the amplitude of the density fluctuations at some initia epoch z_i, by the spectral index and by the form of the window function. Inserting this into equation 10 we obtain

$$\tilde{f}(M, \delta_c(z)|M', \delta')\,dM = \frac{1}{\sqrt{2\pi}}\frac{(n+3)}{3}\frac{\sigma^2(M)}{[\sigma^2(M)-\sigma^2(M')]^{3/2}} \qquad (11)$$
$$\times(\delta_c-\delta')\exp\left\{-\frac{(\delta_c-\delta')^2}{2[\sigma^2(M)-\sigma^2(M')]}\right\}\frac{dM}{M},$$

where we have changed the notation used for the variances in order to emphasize that most of the functional dependence of $\tilde{f}$ on the scale M is contained in the variance $\sigma^2(M)$.

Equation (11) can be simplified and rewritten to make the scaling of the masses with the amplitude of the initial power spectrum more explicit. We define

$$\Delta = \left[\frac{\sigma^2(M)-\sigma^2(M')}{\sigma^2(M)}\right]^{1/2} = \left[1-\left(\frac{M}{M'}\right)^{(n+3)/3}\right]^{1/2},$$

and write δ_0 for the critical overdensity that a structure must have at the initial epoch, z_i, in order to have just collapsed at the present day [i.e. $\delta_0 = \delta_c(z_i)/(1+z_i) = 1.68/(1+z_i)$]. A region which just collapses at epoch z must have an overdensity at z_i of $\delta_c(z) = (1+z)\,\delta_0$. (We are here making the assumption that the Universe is flat.) In this notation,

$$\tilde{f}(M, \delta_c|M', \delta')\,dM = \frac{1}{\sqrt{2\pi}}\frac{(n+3)}{3}\frac{1}{\Delta^3}M^{(n+3)/6}\frac{(1+z-\delta'/\delta_0)\,\delta_0}{A}$$
$$\times\exp\left[-\frac{M^{(n+3)/3}\delta_0^2(1+z-\delta'/\delta_0)^2}{2\Delta^2A^2}\right]\frac{dM}{M}.$$

Note that $M_*(0) \equiv (\delta_0/A)^{-6/(n+3)}$ determines the typical mass (averaging over the whole Universe) that has just collapsed at the present epoch, and that $M_*(z-\delta'/\delta_0) = [(1+z-\delta'/\delta_0)\,\delta_0/A]^{-6/(n+3)}$ determines the typical size scale that collapses *in this region* at the epoch z (the constant A is determined by the normalization of the power spectrum). There is an obvious restriction that $1+z-\delta'/\delta_0 > 0$, or else the larger region M' would have collapsed *before* the epoch z, so that its division into sub-units at this epoch is impossible. The conditional multiplicity fraction can be simplified by using these variables to scale the solution

$$\tilde{f}(M, \delta_c|M', \delta')\,dM = \frac{1}{\sqrt{2\pi}}\frac{(n+3)}{3}\frac{1}{\Delta^3} \qquad (12)$$
$$\times\left[\frac{M}{M_*(z-\delta'/\delta_0)}\right]^{(n+3)/6}$$
$$\times\exp\left\{-\frac{1}{2\Delta^2}\left[\frac{M}{M_*(z-\delta'/\delta_0)}\right]^{(n+3)/3}\right\}$$
$$\times\frac{dM}{M}.$$

In the case that $M' \gg M$, $\Delta \approx 1$ (i.e. the correlation between the scales is negligible) and we recover the 'k-split' approximation used by EFWD. In the limit $M' \to \infty$ (i.e. the large-scale region grows to include the entire Universe), by definition $\delta' \to 0$, and the conditional multiplicity fraction reduces to the universal multiplicity fraction, equation (2) (divided by ρ_0).

As is explained in the following section, we will only apply this formula to present-day regions which are just sufficiently overdense to collapse on the scale M', but are not dense enough to collapse on a slightly larger scale. This requires that $\delta' = \delta_0$. Adopting this value, we have plotted, in Figs 1–3, $\tilde{f}(M)$ for a selection of parameters covering the range of interest. In each diagram, we show four curves corresponding to the spectral indices 1, 0, −1 and −2. Three of the set of four diagrams at each epoch show the effect of altering the mass of the present-day group, M', between 32, 1.0 and 0.13 $M_*(0)$. In the remaining figure, we show the Press–Schechter *universal* multiplicity function at this epoch. In order to compare the evolution of the groups, the diagrams are repeated at redshifts of 0.5, 1.0 and 2.0. In order to be consistent with EFWD, we have plotted the multiplicity fraction as a function of multiplicity scaled to the present-day peak in the universal multiplicity function, i.e. as a function of $m = \log_2[M/M_*(0)]$.

Careful study of these diagrams allows us to qualitatively understand the behaviour of $\tilde{f}$. First, the fact that the region M' is overdense has the effect of accelerating the evolution of the groups inside its boundary. Thus the peak in the multiplicity function (i.e. the typical size of a group) for the $M' = 32$ case is always shifted to higher masses than the peak in the universal distribution. The degree of 'acceleration' is greater for flatter (i.e. more negative) power spectra. The factor M_* in the exponential of equation (12) is always advanced above the M_* of the universal distribution (equation 2) in any region which is collapsing at the present

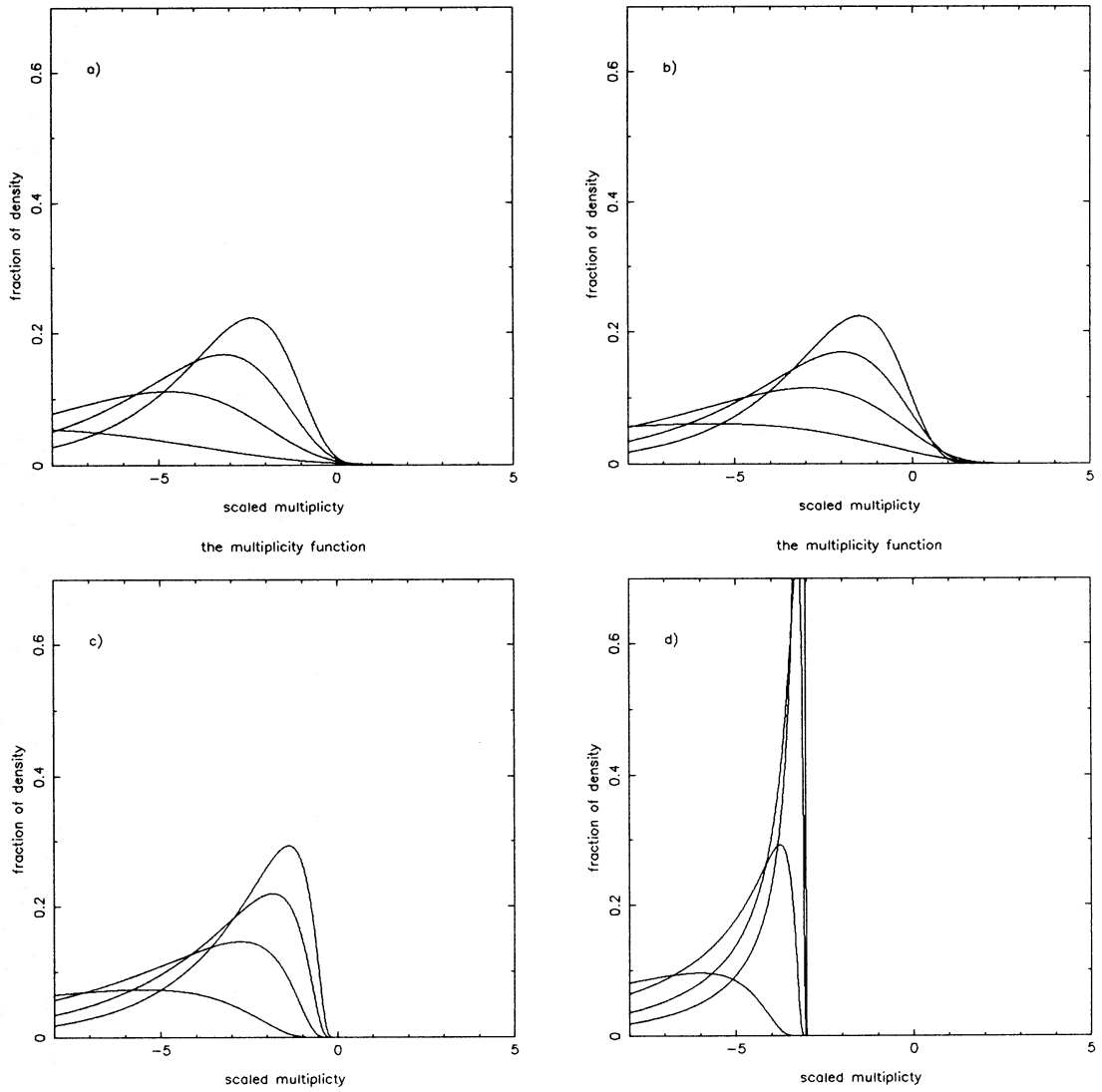

Figure 1. The conditional multiplicity fraction (the fraction of the present-day group mass, M', bound into groups of mass M at redshift z) plotted for $z = 2.0$. The group mass M is presented as multiplicity scaled to the typical mass of a present-day group, i.e. $m = \log_2[M/M_*(0)]$. The four curves in each plot correspond to spectral indices, n, of 1, 0, −1 and −2 (most to least peaked respectively). Plot (a) shows the universal multiplicity function (equation 2) scaled to $\rho_0 = 1$; (b) shows the conditional multiplicity fraction, $\tilde{f}$, in a region which collapses to form a single group of scaled multiplicity $m' = 5.0$ at the present epoch; (c) and (d) show $\tilde{f}$ for $m' = 0.0$ and −3.0, respectively.

day, irrespective of its mass. However, there is clearly an upper limit to the masses of the groups that can be contained within M' (i.e. M' itself). This cut-off is introduced by the factor Δ in the exponent that has arisen from the cross-correlation between the mass scales. The sharpness of the truncation changes with the spectral index, and the epoch at which the groups are 'observed'. It is this factor that distinguishes our formula from the k-split approximation derived by EFWD; for steep power spectra, e.g. $n = +1$, the cut-off is very sudden, and the k-split is a good approximation over almost the full range of masses below M' provided no attempt is made to renormalize the distribution. In the formula presented here (we will show later that it appears to be the exact solution), the normalization of the multiplicity fraction (i.e. the requirement that $\int_0^\infty \tilde{f}\, dM = 1$) is maintained by the $1/\Delta^3$ factor. It has the effect of piling up the 'extra' groups, which would have masses larger than M' in the absence of the cut-off, just below M'. If the power spectrum is steep, and the peak of the distribution in larger regions has advanced beyond M', a very sharp spike is produced. For flat initial power spectra (e.g. $n = -2$), the distributions are always very broad, a wide range of group sizes contributing significantly to the final mass. The effect of the cut-off at M' is felt in the multiplicity fraction of group masses significantly smaller than M'.

We may now understand the time evolution of the multiplicity fraction displayed in Figs 1–3. At large redshifts, there is little tendency for groups inside M' to have masses that approch M'. The distribution is therefore very similar to the universal distribution at this epoch, but advanced to a somewhat larger mean mass (*cf.* Fig. 1b). At later times, groups in the high-mass tail of the multiplicity function show some tendency to grow to masses approaching M'. This causes a distortion in the shape of the multiplicity fraction (*cf.* Figs 1c

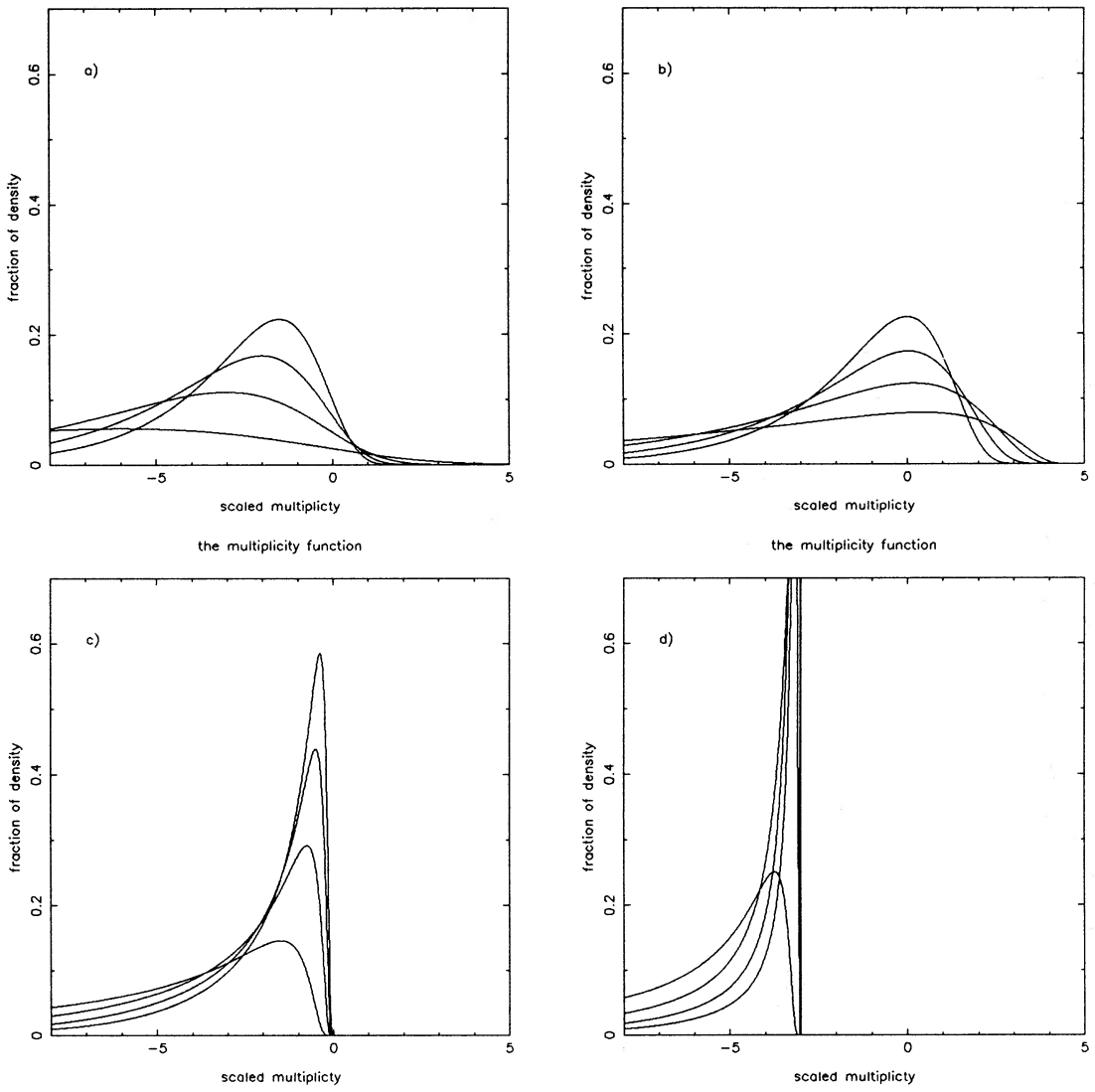

Figure 2. The multiplicity functions shown in Fig. 1 reproduced at $z = 1.0$.

and 2b) which grows as the fraction of mass that would otherwise be bound into groups larger than M' increases (e.g. Fig. 2c). At late times, almost all the mass is bound into groups that are only slightly less massive than M' (e.g. Fig. 3c). Note that while the qualitative explanation of the evolution of $\tilde{f}$ is the same for all spectral indices and present-day masses (M'), the time-scale is strongly dependent on n and M'.

3.4 The joint multiplicity function

The conditional multiplicity fraction derived in the previous section is not the ultimate goal of this paper. In order to study the evolutionary history of groups and clusters of galaxies, we also require the *joint multiplicity function*, i.e. the mass density of groups with masses in the range M to $M+dM$ at epoch z that are incorporated into groups with masses M' to $M'+dM'$ at the present epoch.

Before proceeding, it is helpful to review the reasoning behind the PS ansatz (equation 1) in greater detail. We consider a volume V chosen at random. The probability that is volume has overdensity greater than the threshold for collapse, δ_c, is

$$\mathscr{P}(\delta_V > \delta_c) = \int_{\delta_c}^{\infty} \frac{1}{\sqrt{2\pi\sigma_V^2}} \exp(-\delta'^2/2\sigma_V^2)\, d\delta'.$$

However, we wish only to count objects which collapse on the scale of V but do *not* collapse on a slightly larger scale, $V+dV$. The probability that this is true of region V is

$$p(V)\, dV = \mathscr{P}(\delta_V > \delta_c)\mathscr{P}(\delta_{V+dV} < \delta_c | \delta_V > \delta_c),$$

or, if we introduce the short-hand notation $\mathscr{P}(V) = \mathscr{P}(\delta_V > \delta_c)$ and $\mathscr{P}(\overline{V+dV}|V)$ for the conditional probability that $V+dV$ has *not* collapsed, $\mathscr{P}(\delta_{V+dV} < \delta_c | \delta_V > \delta_c)$,

$$p(V)\, dV = \mathscr{P}(V)\mathscr{P}(\overline{V+dV}|V) = \mathscr{P}(V)(1 - \mathscr{P}[V+dV|V]), \tag{13}$$

with the obvious extension to the notation. Applying Bayes' formula to write the conditional probability, $\mathscr{P}(V+dV|V)$, in terms of the joint probability that both V and $V+dV$ are

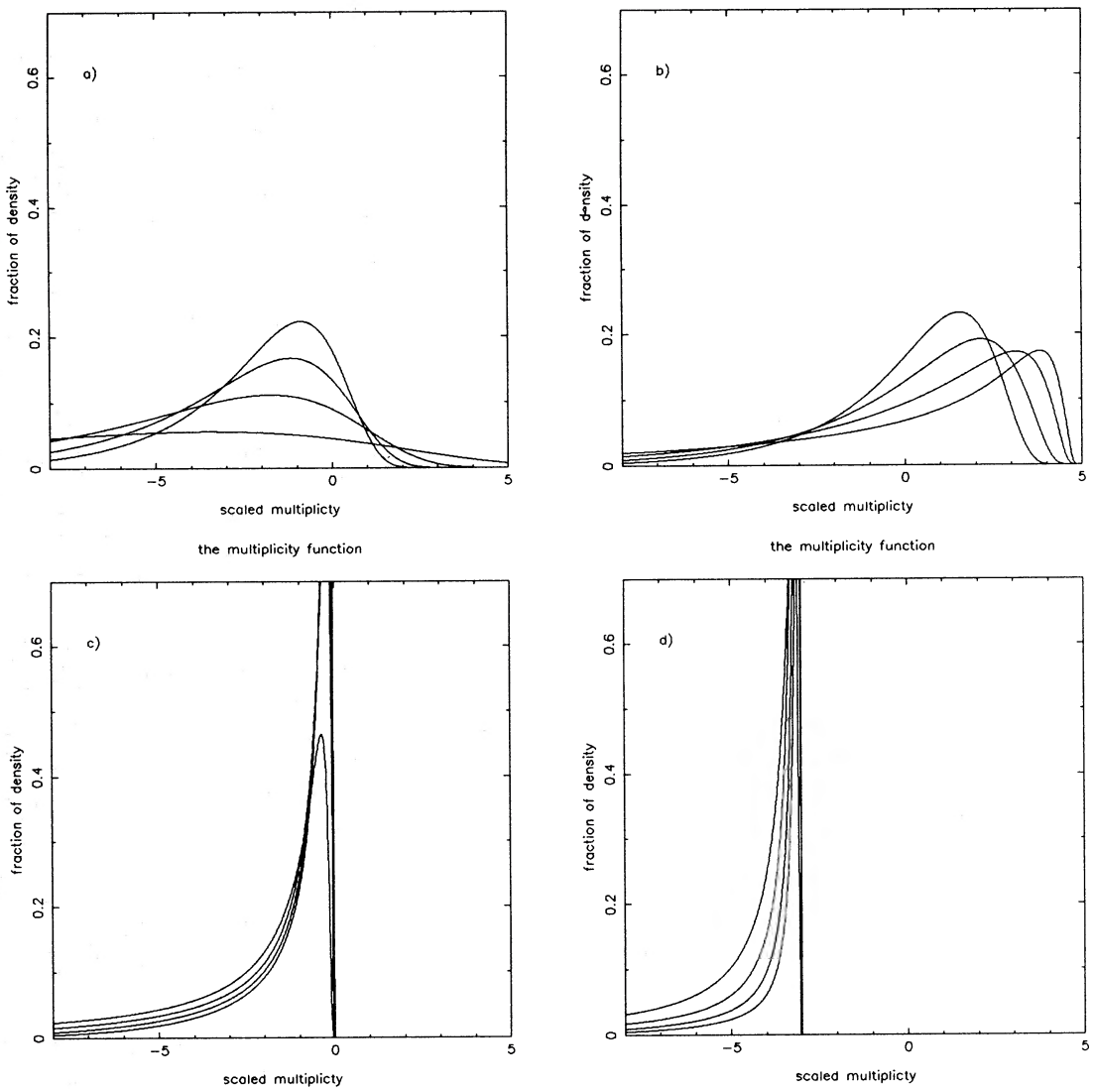

Figure 3. The multiplicity functions shown in Fig. 1 reproduced at $z = 0.5$.

above the critical density, we have:

$$p(V)\,dV = \mathscr{P}(V)\left(1 - \frac{\mathscr{P}(V+dV, V)}{\mathscr{P}(V)}\right)$$

$$= \mathscr{P}(V) - \mathscr{P}(V+dV, V)$$

The next step is to note that if $V+dV$ collapses, then so must V. We can therefore set $\mathscr{P}(V+dV, V) = \mathscr{P}(V+dV)$, giving

$$p(V)\,dV = \mathscr{P}(V) - \mathscr{P}(V+dV) = -\frac{\partial \mathscr{P}(V)}{\partial V}\,dV.$$

The total number of possible collapsed volumes of size V is proportional to $1/V$, but the mass of each group is proportional to V. The mass density (or multiplicity function) of groups of mass $M(\propto \rho_0 V)$ is therefore taken to be proportional to

$$-\rho_0 \frac{\partial \mathscr{P}(M)}{\partial M}\,dM.$$

In order that all the mass of the Universe is bound into collapsed objects (even if their size is very small), the constant of proportionality must be 2.0 $[= 1/\mathscr{P}(0)]$. This brings us to equation (1).

Following the same technique, we can write the *joint multiplicity function* as the product of probabilities

$$\rho(M, z, M')\,dM\,dM' = \alpha\rho_0 \mathscr{P}(M') \times \mathscr{P}(\overline{M'+dM'}|M') \mathscr{P}(M|\delta_{M'} = \delta_0) \times \mathscr{P}(\overline{M+dM}|M, \delta_{m'} = \delta_0),$$

where we make use of the short-hand notation adopted in equation (13). α is a numerical constant of order unity. In addition, we have made use of the differentiability of the fluctuations in the density field (when smoothed on scale M) to note that the requirement that the region M' have mean overdensity greater than δ_0, but that the immediately surrounding region $M'+dM'$ have $\delta_{M'+dM'} < \delta_0$, forces $\delta_{M'} = \delta_0$. Expanding the probabilities as done previously, and using the definition of the conditional multiplicity fraction, $\tilde{f}$

(equation 9), gives

$$\rho(M, z, M')\, dM\, dM' = -2\rho_0 \tilde{f}(M, \delta_c | M', \delta_0)\, dM \times \frac{\partial P(\delta_{M'} > \delta_0)}{\partial M'}\, dM',$$

where the numerical factor of 2 (corresponding to $\alpha = 4$) has been chosen in accordance with the original hypothesis of PS. The expression is then simplified following the original derivation of PS to give

$$\rho(M, z, M')\, dM\, dM' = 2\rho_0 \tilde{f}(M, \delta_c | M', \delta_0) \tag{14}$$

$$dM \frac{1}{\sqrt{2\pi}} \frac{(n+3)}{6} \frac{\delta_0}{\sigma(M')} \times \exp\left[\frac{-\delta_0^2}{2\sigma^2(M')}\right] \frac{dM'}{M'}$$

$$= \frac{\rho_0}{2\pi}\left(\frac{n+3}{3}\right)^2 \frac{1}{\Delta^3}\left[\frac{M}{M_*(z-1)}\right]^{(n+3)/6} \times \exp\left\{-\frac{1}{2\Delta^2}\left[\frac{M}{M_*(z-1)}\right]^{(n+3)/3}\right\} \times \left[\frac{M'}{M_*(0)}\right]^{(n+3)/6} \times \exp\left\{-\frac{1}{2}\left[\frac{M'}{M_*(0)}\right]^{(n+3)/3}\right\} \frac{dM}{M} \frac{dM'}{M'}$$

4 SELF-CONSISTENCY AND NUMERICAL ACCURACY

4.1 Recovering the universal Press–Schechter distribution

Since we have made no assumptions regarding the relative sizes of M and M', we should expect to be able to recover the universal (PS) multiplicity functions of M and M' from their joint distribution function. To recover the distribution of M' from $\rho(M, z, M')\, dM\, dM'$ is trivial (given the PS factor of 2). The integral of $\rho(M, z, M')\, dM\, dM'$ over M may be split into the integral of $\tilde{f}[M, \delta_c(z) | M', \delta_0]\, dM$ over M, and the function

$$\frac{\rho_0}{\sqrt{2p}}\left(\frac{n+3}{3}\right)\left(\frac{M'}{M_*(0)}\right)^{(n+3)/6} \exp\left[-\frac{1}{2}\left(\frac{M'}{M_*(0)}\right)^{(n+3)/3}\right] \frac{dM'}{M'}$$

(which is exactly the universal multiplicity function at the present-day, equation 2). From the way in which we have defined $\tilde{f}$ as a fraction (equation 9), its integral over M must be unity.

It is not, however, obvious that we can recover the universal distribution of M, as we require a seeminly unlikely conspiracy that allows the suppression factor Δ to exactly compensate for the acceleration of the clustering hierarchy caused by the overdensity of the final region. The dependence of the joint multiplicity function on M' is of such a form that the integration cannot be performed analytically by standard techniques. However, we show in Appendix B, using an unusual analytical technique, that the integral

$$\int_{M'=M}^{\infty} \rho(M, z, M')\, dM\, dM' \tag{15}$$

does indeed exactly recover the universal multiplicity function $\rho(M, z)$. We emphasize that this is an extremely striking result. It assures us of the mathematical self-consistency of the Press–Schechter formalism that we have applied in the derivation of the joint multiplicity function (equation 14).

In addition to the result described above, White & Frenk (1990), have noted that the conditional multiplicity functions obey the relation

$$\int_{M_2} \tilde{f}(M_1, z_1 | M_2, z_2)\, \tilde{f}(M_2, z_2 | M_3, z_3)\, dM_2 = \tilde{f}(M_1, z_1 | M_3, z_3),$$

where $z_1 > z_2 > z_3$. This result is unexpected as it is not required for the self-consistency of the theory.

4.2 Comparison with *N*-body simulations

Comparison with the N-body simulations of EFWD provides a second test of the validity of the joint multiplicity function. These authors plot the fraction of groups of mass M (at five different epochs) which are incorporated into massive groups at the end of the simulation (massive groups are defined so as to contain 18 per cent of the total mass at the final output time). This fraction is readily calculated from the joint multiplicity function,

$$F_{mg} = \frac{\int_{M'_{mg}}^{\infty} \rho(M, z, M')\, dM\, dM'}{\rho(M, z)\, dM}.$$

The parameters z and $M_*(0)$ are defined by the output times and initial perturbation amplitudes of the simulations. Table 1 lists relevant values. The epoch is calculated from the expansion factors given by EFWD ($1 + z = a_f/a$). A formula is also given for M_* at the final epoch time, but note that (i) the definition of M_* used by EFWD differs from ours by a factor of $2^{3/(n+3)}$; (ii) the initial amplitude of the $n = -2$ simulations is defined at an expansion factor $a = 2$. There are no free parameters in our calculation of F_{mg}.

Table 1. Parameters used to form comparison with the N-body simulations of EFWD.

n	$M_*(0)$	z_{f-1}	z_{f-2}	z_{f-3}	z_{f-4}	z_{f-5}
+1	260	0.848	2.415	5.311	10.66	20.55
0	194	0.585	1.512	2.982	5.311	9.00
−1	103	0.359	0.847	1.510	2.411	3.636
−2	82.4	0.166	0.360	0.585	0.848	1.155

The PS predictions are superposed on the plots of EFWD in Figs 4–7. The fits can be seen not only to match the qualitative behaviour of the N-body result, but also, with a few exceptions, to give good numerical agreement. Where the fit is poor, it is not clear whether it is the PS theory or the N-body approach that is inaccurate – e.g. in the $n = -2$ case, the discrepancy can be understood in terms of saturation and transient effects in the simulations. We conclude that the joint multiplicity function, that we have derived using the principles of PS, *provides an accurate description for the evolution of gravitational structure in the expanding universe.*

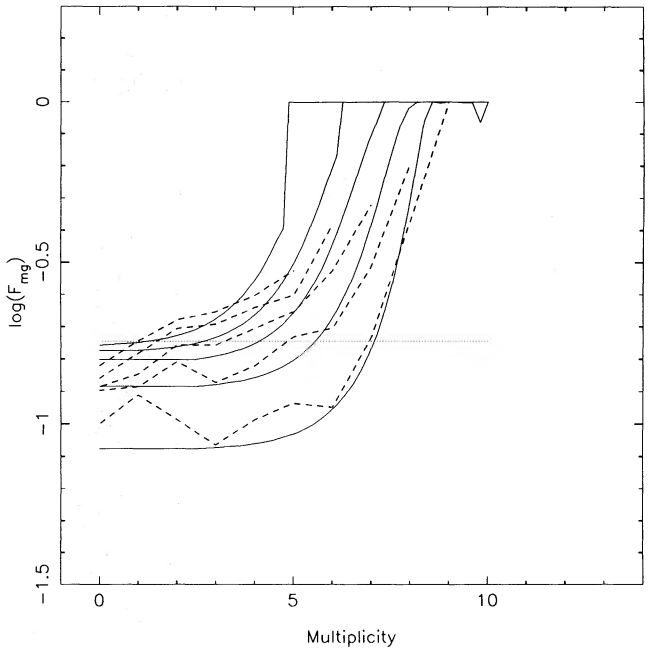

Figure 4. Comparison of the mass fraction (i.e. the fraction of the mass contained today in the most massive groups that is bound into groups more massive than M at a previous epoch) calculated from the joint multiplicity function (smooth solid lines) with that measured in the simulations of EFWD. The last five output times from the simulations are shown as dashed lines. This plot shows the case of $n=1$. The parameters used to define the theoretical curves are given in Table 1.

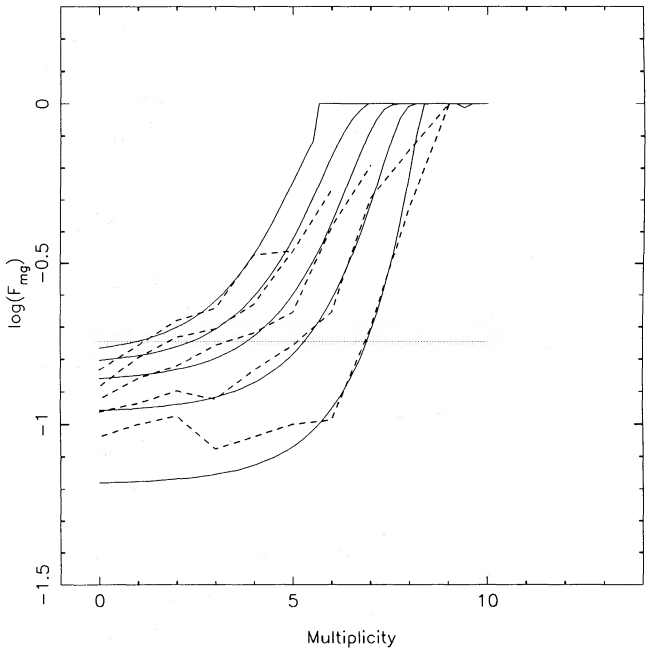

Figure 5. As for Fig. 4, but for the case of $n=0$.

5 APPLICATION TO THE EVOLUTION OF GROUPS OF GALAXIES

In order to illustrate the manner in which a cluster of galaxies builds up from the infall of smaller groups, it is convenient to divide groups into two types above and below a mass-scale M_s. We apply the multiplicity fractions derived previously (equations 12 and 14) to address problems relating to the mass flux of groups through the mass-scale M_s.

5.1 The growth of proto-clusters

At epoch z, the fraction of the final mass of a cluster that is contained in high-mass groups (i.e. $M>M_s$) is related to the conditional multiplicity fraction (equation 12) by

$$F_s(z, M'; M_s) = \int_{M_s}^{\infty} \tilde{f}(M, z|M')\, dM, \tag{16}$$

where $\tilde{f}(M, z|M') = \tilde{f}[M, \delta_c(z)|M', \delta_c(0)]$ in the notation of equation (12),

$$= + \int_{u_s}^{\infty} \frac{2}{\sqrt{2\pi}} \frac{z}{\Delta^3} \exp\left(-\frac{z^2u^2}{2\Delta^2}\right) du$$

in the notation of Appendix B. After some working this can be rewritten in terms of the complementary error function

$$F_s(z, M'; M_s) = \mathrm{erfc}\left\{\frac{z}{\sqrt{2}\Delta_s}\left[\frac{M_s}{M_*(0)}\right]^{(n+3)/6}\right\},$$

where the final cluster mass enters through

$$\Delta_s = \left[1 - \left(\frac{M_s}{M'}\right)^{(n+3)/3}\right]^{1/2}.$$

We apply this formula directly to calculate the epoch at which half of the present-day mass of a cluster is bound into higher mass groups (i.e. $M>M_s$). Setting $F_s(z_s, M'; M_s)=1/2$ gives

$$z_s(M'; M_s) = \frac{0.97}{[M_s/M_*(0)]^{(n+3)/6}}\left[1 - \left(\frac{M_s}{M'}\right)^{(n+3)/3}\right]^{1/2} \tag{17}$$

For the case $M_s = M_*(0)$, z_s is plotted for each of the spectral indices in Fig. 8. Its behaviour may be readily understood as follows. In the limit $M' \to M_s$, $z_s \to 0$, and as $M' \to \infty$, $z_s \to z_\infty = 0.97[M_s/M_*(0)]^{-(n+3)/6}$; however, the transition between the two extremes is sensitive to the spectral index. For $n=+1$, z_s reaches 90 per cent of its final value at $5.6\,M_s$, i.e. other than the small range of groups that have only recently grown through M_s, all large present-day groups assembled half their mass into groups more massive than M_s at very similar epochs. On the other hand, z_s varies more slowly with M' for a flat initial power spectrum ($n \leq -2$). This leads to a somewhat stronger correlation between the present-day mass of a cluster and the history of its evolution.

5.2 Infall into clusters of galaxies

We have not yet exploited the full information contained in the function F_s. The growth of F_s with time gives us information on the rate at which the mass contained in groups smaller than M_s is being accumulated into groups more massive than M_s. It is helpful to distinguish the two ways in

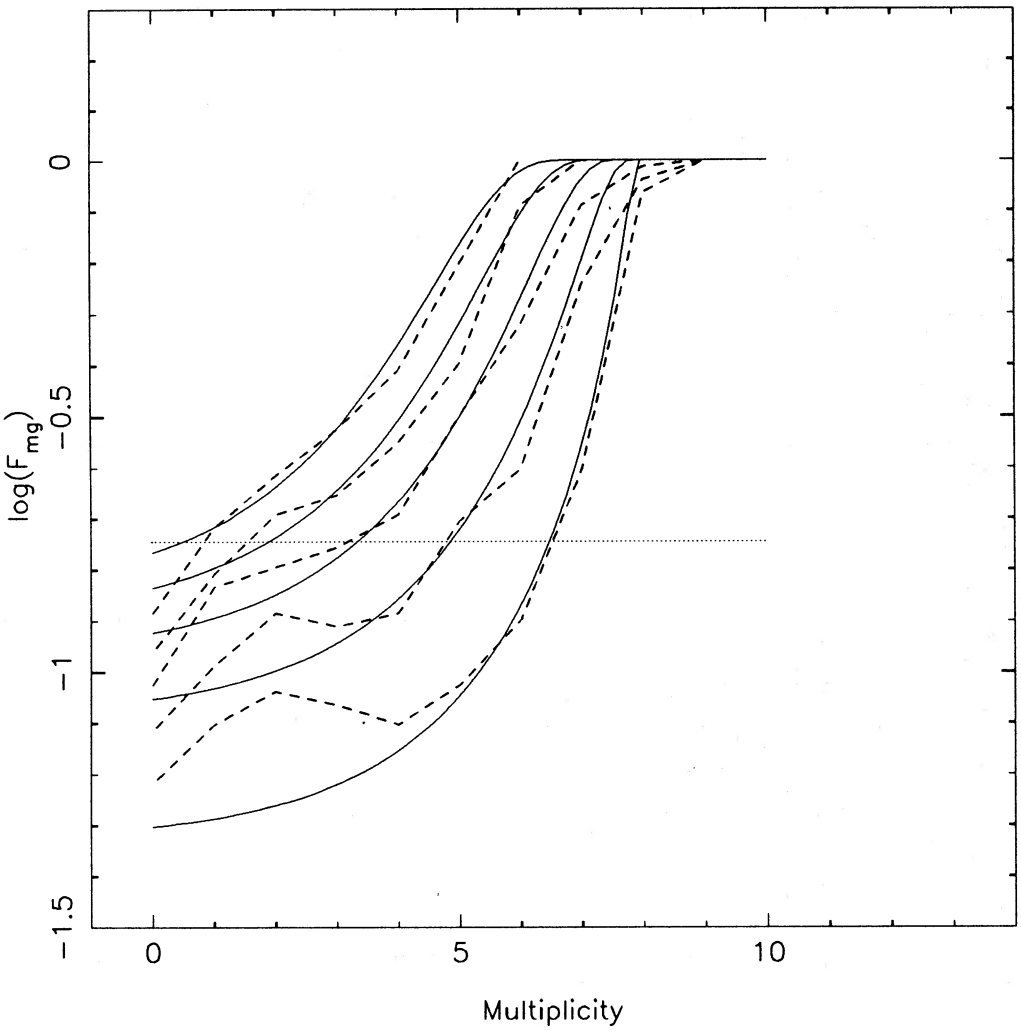

Figure 6. As for Fig. 4, but for the case of $n = -1$.

which this growth may occur: (i) small groups ($M < M_s$) are accreted on to (a few) much larger condensations ($M > M_s$); (ii) a few small groups, all having mass less than M_s, coalesce to produce a new massive condensation, this condensation later merging with other proto-cluster cores of similar size. Which of these mechanisms is dominant depends on the scale mass (M_s), the typical group mass (M_*), the present-day cluster mass (M') and the spectral index (n), and cannot readily be determined without recourse to numerical simulations. We will refer to $\partial F_s/\partial t$ as the *infall rate* – visualizing it as the rate at which matter is infalling into the proto-cluster, either directly by accretion on to pre-existing proto-cluster cores, or by the creation of new proto-cluster cores which subsequently coalesce with their longer established counter-parts. It is important to note that we do not mean to imply that the proto-cluster is dominated by a single very massive condensation.

Up until this point it has not been necessary to distinguish between redshift (or the expansion of the Universe) and time. However, as we will wish to apply the present results to systems in which the natural time-scale is set by the clock of stellar evolution, it is necessary to derive the growth rate of F_s per unit proper time:

$$R_{\rm infall} = \frac{\partial F_s}{\partial t} = \frac{\partial F_s}{\partial z}\frac{dz}{dt} \tag{18}$$

$$= \frac{2}{\sqrt{2\pi}} \frac{[M_s/M_*(0)]^{(n+3)/6}}{\Delta_s} \frac{2}{3}(1+z)^{5/2}$$

$$\times \exp\left\{-\frac{z^2}{2\Delta_s^2}\left[\frac{M_s}{M_*(0)}\right]^{(n+3)/3}\right\}.$$

The infall rate is plotted in Figs 9 and 10 for initial power spectra with $n = 0$ and -2, respectively. We have set $M_s = M_*(0)$; and different line styles show final masses $M'/M_*(0) = 1.5$, 3.0, 10 and 1000. First, we note that 'infall' for a given final mass peaks at a well-defined time. The occurrence of this peak is purely a result of the compression of the relationship between redshift and time – the infall rate per unit redshift increases monotonically to the present day. The epoch at which the peak infall rate occurs depends on final mass in the same manner as z_s. Secondly, we note that

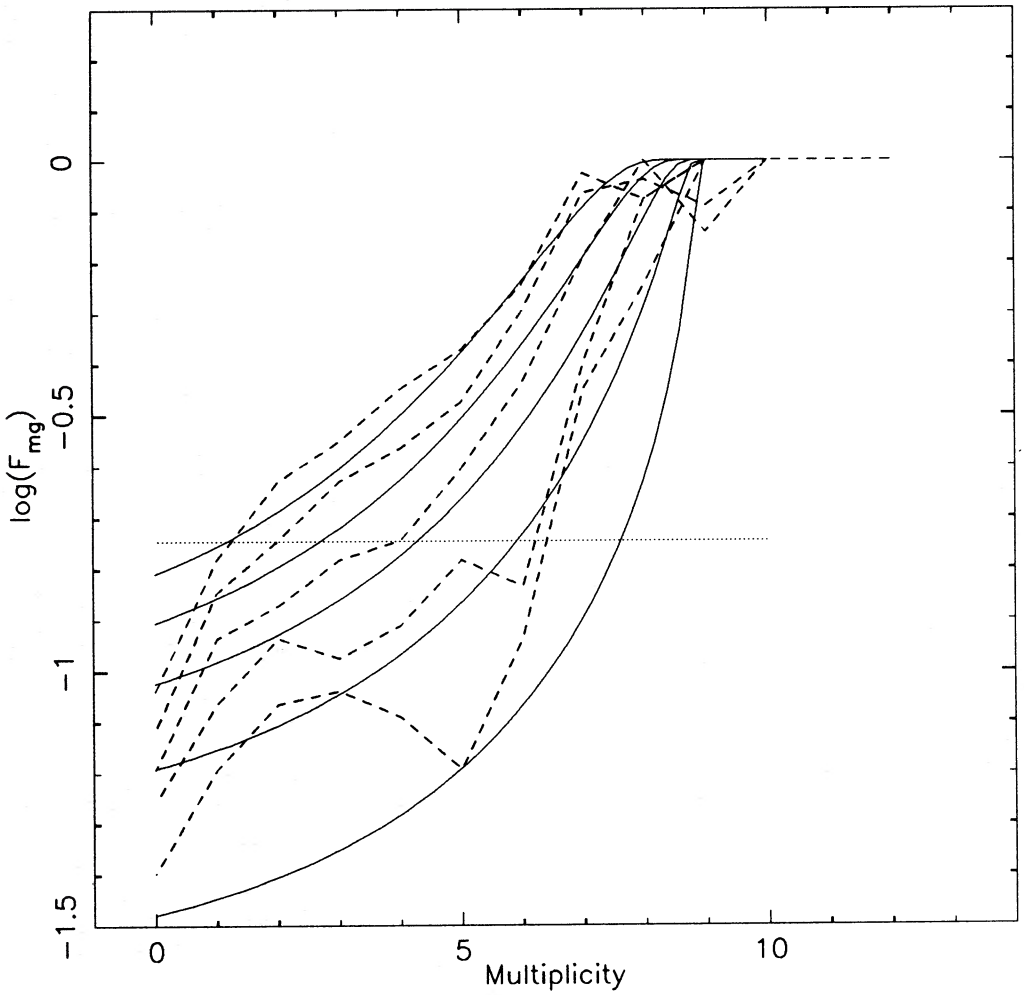

Figure 7. As for Fig. 4, but for the case of $n = -2$.

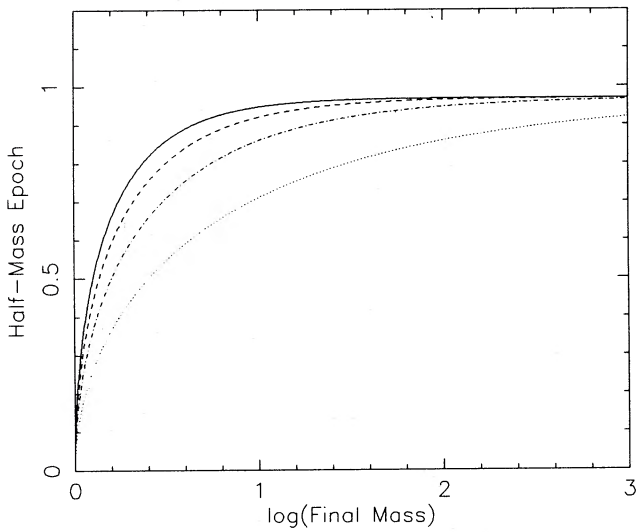

Figure 8. This figure shows the redshift at which half the present-day mass of a group is bound into groups more massive than the 'scale mass', M_s. It is plotted as a function of the log (to base 10) of the present-day mass [in units of $M_*(0)$] for the case $M_s = M_*(0)$. The four curves show the effect of altering the spectral index of the initial perturbations. The solid line corresponds to $n = +1$; the dashed line to $n = 0$; the dot–dashed line to $n = -1$; and the dotted line to $n = -2$. All the curves tend to an asymptotic redshift of 0.97.

although there is a well-defined peak to the infall rate, the fall-off is very slow so that the peak rate is enhanced over that at the present day by a factor of only ~ 3, i.e. in this model the infall of small groups into rich clusters is very much an on-going process.

It is interesting to apply this equation to compare the present-day infall rates of clusters of various masses,

$$R_{\rm infall}(0) = \frac{4}{3\sqrt{2\pi}} \frac{[M_s/M_*(0)]^{(n+3)/6}}{\Delta_s}. \tag{19}$$

Comparison with equation (17) shows that the present infall rate is intimately connected with the epoch at which half the mass of the present-day cluster became bound into groups more massive than M_s, i.e. z_s:

$$R_{\rm infall}(0) \propto \frac{1}{z_s}. \tag{20}$$

5.3 The density of large groups at early times

Yet further information on the evolution of groups may be extracted from the joint multiplicity function, equation (14). We address the question of the density of the groups that are

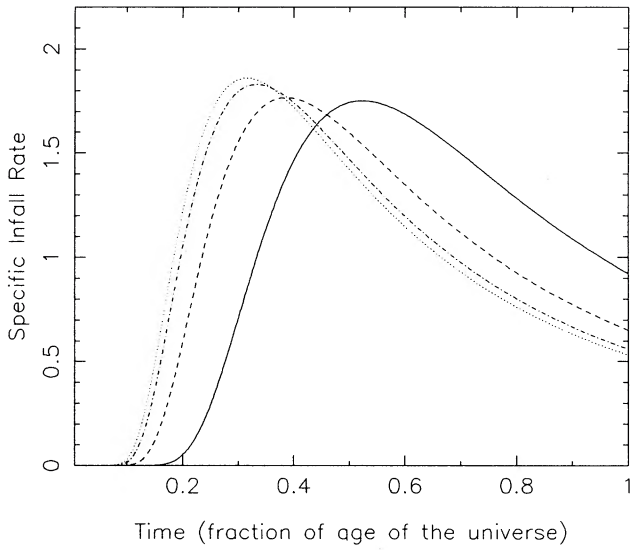

Figure 9. This figure shows the time dependence of the infall rate of small groups into larger ones for a steep ($n=0$) initial power spectrum. Time, expressed as a fraction of the total age of the Universe, advances from left to right, i.e. the present is at 1.0. The rate of infall is expressed as the fraction of the present-day mass of the group infalling per unit time. The four curves show the infall in regions which collapse to form present-day groups of masses 1.5 (solid line), 3.0 (dashed line), 10.0 (dot–dashed line) and 1000.0 (dotted line). All masses have been expressed in terms of the typical present-day mass, $M_*(0)$, and we have divided small and large groups at $M_s = M_*(0)$.

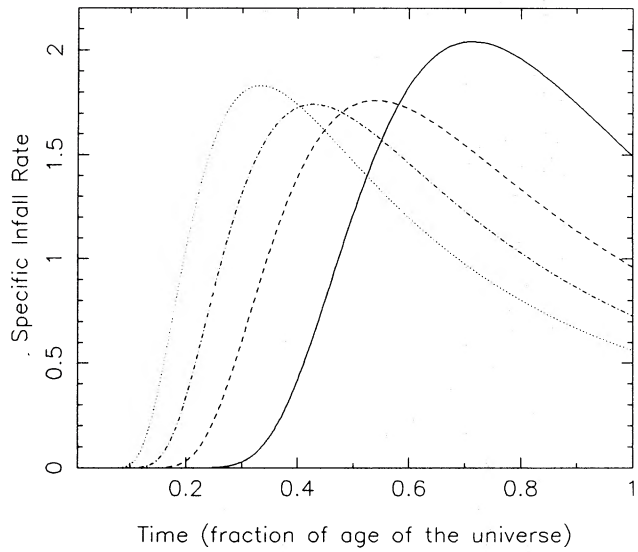

Figure 10. The time dependence of the infall rate for a flat ($n=-2$) power spectrum. The different curves show the effect of varying the present-day mass as described in Fig. 9.

more massive than M_s (an arbitrary mass-scale) at an early epoch. The fraction of these groups that become bound into groups of mass M' at the present epoch is

$$f(M'|>M_s,z)\,dM' = \frac{\int_{M_s}^{\infty}\rho(M,z,M')\,dM\,dM'}{\int_{M_s}^{\infty}\rho(M,z)\,dM}$$

$$\doteq \frac{\int_{u_s}^{\infty}(4/2\pi)(1/\Delta^3)z\exp[-(z^2u^2/2\Delta^2)]\exp[-(u'^2/2)]\,du\,du'}{\int_{u_s}^{\infty}(2/\sqrt{2\pi})(1+z)\exp[-(1+z)^2u^2/2)]\,du},$$

in the notation of Appendix B. The integrals in this equation can be written as complimentary error functions

$$f(M'|>M_s,z)\,dM' = \frac{2}{\sqrt{2\pi}}\exp\left(-\frac{u'^2}{2}\right)$$

$$\times\frac{\mathrm{erfc}[(z/\sqrt{2}\Delta_s)u_s]}{\mathrm{erfc}\{[(1+z)/\sqrt{2}]u_s\}}\,du' \qquad (21)$$

$$=\frac{1}{\sqrt{2\pi}}\left(\frac{n+3}{3}\right)\left[\frac{M'}{M_*(0)}\right]^{(n+3)/6}$$

$$\times\exp\left\{-\frac{1}{2}\left[\frac{M'}{M_*(0)}\right]^{(n+3)/3}\right\}$$

$$\times\frac{\mathrm{erfc}\{(1/\sqrt{2}\Delta_s)[M_s/M_*(z-1)]^{(n+3)/6}\}}{\mathrm{erfc}\{(1/\sqrt{2})[M_s/M_*(z)]^{(n+3)/6}\}}\frac{dM'}{M^*}.$$

This equation is plotted for $n=0$ and $n=-2$ spectral indices in Figs 11 and 12, respectively. Its form can be qualitatively understood as follows. Apart from the factor formed by the ratio of the two error functions, the distribution follows that of the universal multiplicity function at the present epoch. However, the relationship between the mass-scale, M_s and the present-day mass enters through the factor Δ_s in the argument of the upper error function. At values of M' close to M_s, Δ_s is small, forcing the value of the complementary error function towards zero. The effect is to cut-off

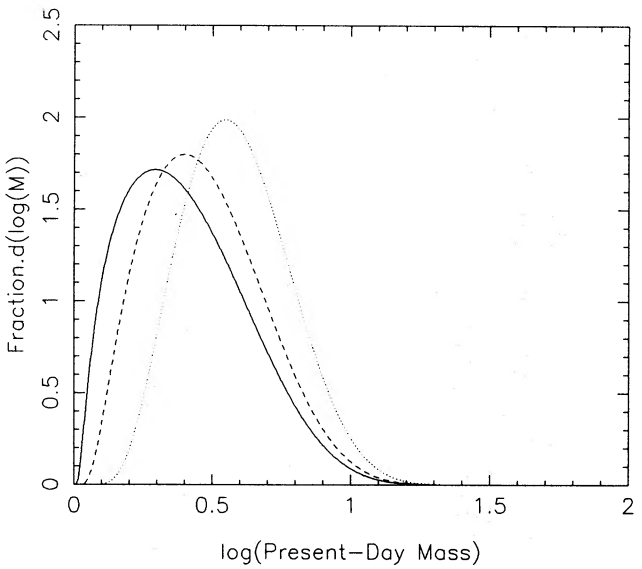

Figure 11. A group which is more massive than M_s at epoch z will evolve to become bound into a group of mass M' at the present epoch. This figure shows the distribution of M' for groups that become more massive than $M_s = M_*(0)$ at redshifts $z=0.5$ (solid line), 1.0 (dashed line) and 2.0 (dotted line). These groups contain respectively 13 per cent, 5 per cent and 0.3 per cent of the total mass of the Universe. In this figure we illustrate the case $n=0$.

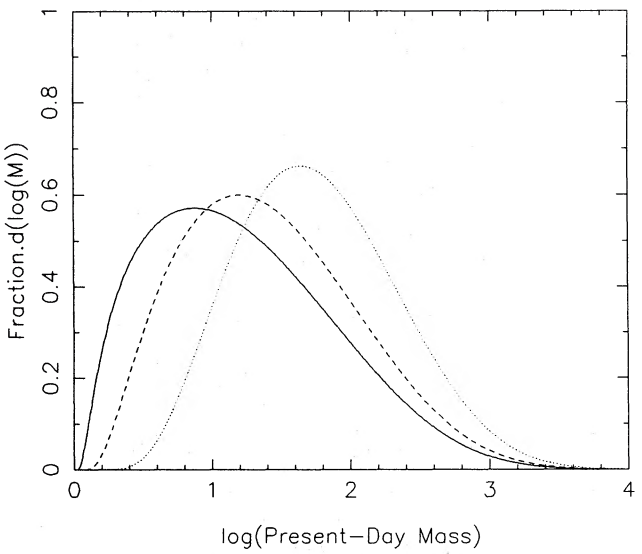

Figure 12. This figure presents the same information as Fig. 11, but for the case of a flat ($n=-2$) power spectrum. Because we have chosen the special case of $M_s = M_*(0)$, the groups we distinguish in each of the curves contain the same fraction of the total mass of the Universe.

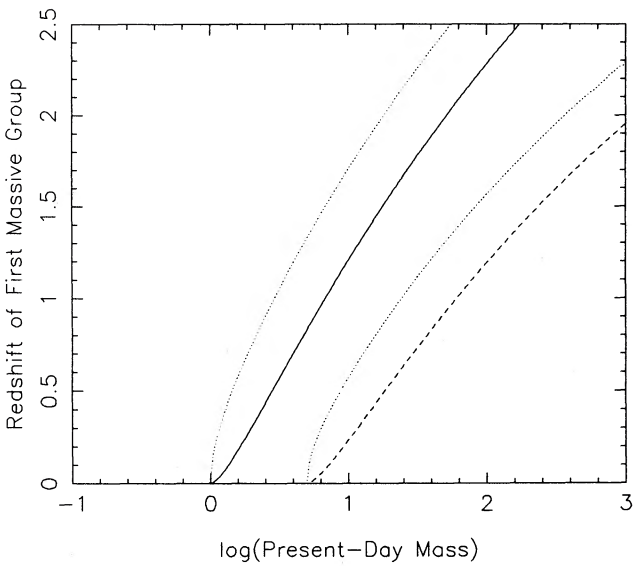

Figure 13. The redshift at which the first massive group (i.e. $M > M_s$) is formed in a region that collapses to become a present-day cluster of mass M'. The curves shown illustrate the dependence on cluster mass for a flat power spectrum ($n=-2$), the distinction between high- and low-mass groups being made at masses of $M_s = 1.0$ (solid line) and 5.0 (dashed line). Dotted lines show, for each value of M_s, the epoch at which a group larger than M_s forms in 20 per cent of regions destined to collapse to become a cluster of mass M'. It can be seen that this epoch is not much different from z_{1M_s}.

the distribution from zero to slightly above M_s. For values of M' considerably larger than M_s, $\Delta \to 1$ and the ratio of the error functions reaches a constant value. This factor re-normalizes the distribution so that the total area under each of the curves remains at unity. From the study of the figures, it can be seen that the cut-off imposed by Δ_s becomes less sharp as the redshift at which the high-mass groups are selected increases. Therefore, the groups that become massive at early times tend to evolve to become the most massive groups later on. Very few of the massive groups that we identify at epoch z will fail to grow significantly in their subsequent evolution. This effect is closely related to the *natural biasing* mechanism discussed by White *et al.* (1987).

In the previous sections, we showed that the more massive present-day clusters *do not* have significantly different average evolutionary histories (on group and galaxy mass-scales) from their less massive counter parts. The present result shows that there is a strong tendency for the most massive groups at early times to be incorporated into the most massive clusters at the present-day. These two results may appear contradictory, but they are not. Confusion has arisen because we have not been particular to distinguish between masses that have been specified as a fraction of the present-day cluster mass (e.g. in our calculation of z_s, equation 17), and those that have been specified in absolute value (e.g. in equation 21).

For example, following the derivation of (17), we may calculate the epoch at which 1 per cent of the present-day cluster mass is bound into massive groups,

$$z_{1\%} = \frac{3.65}{[M_s/M_*(0)]^{(n+3)/6}}\,\Delta_s.$$

This function defines a series of curves similar to those displayed in Fig. 8, but with the redshift axis rescaled. $z_{1\%}$ grows rapidly above M_s, but turns over so that there is little variation for $M' > \sim 10\ M_s$. It should, however, be noted that in a present-day group of mass 10 M_s the mass in large ($> M_s$) groups at $z_{1\%}$ is implied to be 0.1 M_s. Clearly, this must be interpreted in a statistical sense, i.e. approximately 1 in 10 of these present-day groups contained one group more massive than M_s at $z_{1\%}$. We may now ask a subtly different question: 'At what epoch does the region that collapsed to form a typical present-day cluster of mass M' contain a single group of mass $> M_s$?' From equation (16), by setting $F_s = M_s/M'$ (i.e. we require a mass of 1 M_s to be bound into a group of mass $> M_s$) we obtain

$$z_{1M_s} = \frac{\sqrt{2\Delta_s}}{[M_s/M_*(0)]^{(n+3)/6}}\,\mathrm{erfc}^{-1}[M_s/M'],$$

where erfc^{-1} denotes the inverese of the complementary error function. This function is plotted for $n=-2$ and $M_s = 1.0$ and 5.0 in Fig. 13. It can be seen that this epoch is strongly correlated with M' – the same effect as that seen in Figs 11 and 12. We emphasize again that this is not in conflict with the effect seen in $z_{1\%}$ (and z_s) as the contribution to the final mass of the cluster made by this early group becomes progressively smaller with increasing M'. The different abundances of high- and low-mass groups are unable to remove the strong correlation seen in Fig. 13 from $f(M' | > M_s, z)$ (*cf.* dotted lines in Fig. 13). The dominant component of the final mass is, however, made up from groups with average properties.

6 APPLICATION TO THE BUTCHER–OEMLER EFFECT

We now focus our attention on the evolution of the galaxy populations of rich clusters. This serves not only as an

interesting investigation in its own right, but also to illustrate the power of the analytic theory that we have developed.

6.1 Quantifying the Butcher–Oemler effect

Relatively recently, it has become possible to study the colours and spectra of galaxies in rich clusters at redshifts corresponding to cosmologically interesting look-back times. The first study of this type, Butcher & Oemler (1978), was based on broad-band photometry. A highly significant trend was found in the sense that $z > 0.3$ rich clusters appeared to contain a much higher fraction of blue galaxies than similar clusters nearby. Subsequently, many analogous studies have been carried out [e.g. Couch & Newell 1984; Butcher & Oemler 1984 (BO84), and references therein].

While these studies generally confirm the original results, it is apparent that the photometric method may suffer from residual contamination of the cluster by foreground field galaxies. In order to quantify this problem, a number of spectroscopic studies have been made (e.g. Dressler & Gunn 1982, 1983 (DG83), 1987 (DG87); Dressler, Gunn & Schneider 1985 (DGS); Lavery & Henry 1986; Couch & Sharples 1987 (CS); Mellier *et al.* 1988). The primary purpose of these studies is to determined redshifts for the blue galaxies in the field of the cluster, and hence to determine whether they are bona fide cluster members. The results vary from cluster to cluster. For example, many of the blue galaxies in the cluster 3C 295 were found to belong to foreground groups (DG83). On the other hand, the large excess of blue galaxies in the cluster Cl 0024 + 1654 is fully confirmed (DGS). Therefore, while in a few clusters spectroscopy has reduced the significance of the photometric Butcher–Oemler effect, the widespread trend of an increasing blue galaxy fraction remains.

In addition to their employment in determining redshifts, these spectra are also able to give some rudimentary information on the stellar populations of the galaxies. The most comprehensive studies at sufficiently high signal-to-noise ratio and spectral resolution have been made by CS and DG87. Contrary to expectation, only a few of the blue galaxies exhibited spectra corresponding to normal spiral types. The majority of the spectra corresponded to galaxies seen during, or shortly after an intense burst of star formation. Comparable spectra are very rare in present-day clusters. We will refer to the increasing fraction of star burst galaxies as the *spectroscopic Butcher–Oemler effect*. Examples of a third class of blue spectrum, corresponding to galaxies with active nuclei, were also found. However, because of the small numbers of these systems, it is difficult to determine whether this galaxy fraction is also increasing with redshift.

The rate of evolution of the blue galaxy or star burst fraction is best quantified by the 'dimensionless' gradient $(1/f_0)(\Delta f/\Delta z)$, where f_0 is the galaxy fraction at $z \approx 0$. In Table 2 we present values for this gradient derived from a selection of sources. For each entry, we present: (i) the redshift range of the distant clusters used to determine the effect; (ii) the type of galaxies that have been compared; (iii) the author's determination of f_0; (iv) and the inferred value of the gradient. We do not present the formal errors in these determinations as cluster-to-cluster variations and differences in the authors' definition of the galaxy types are of greater importance. It can be seen that the evolution is extremely strong. In all of the sources, we obtain $(1/f_0)(\Delta f/\Delta z) > 10$. If the evolution is parameterized as some power of look-back time, $f = f_0(1+z)^m$, then we require $m > 4.0$ in order to account for the observed effect.

Before proceeding to make a comparison with our theoretical work, a note of caution is required. Although the clusters at high and low redshift have been selected to have similar richness and concentration, Newbury, Kirshner & Boroson (1988) have noted that there is a tendency for the clusters' velocity dispersions and surface densities to increase with redshift. While this suggests that the clusters in the high- and low-density samples may not be directly comparable, these changes in cluster properties probably result from an increase in apparent substructure. Such behaviour is to be expected of rich clusters that are being formed in a hierarchical collapse (i.e. the rich galaxy concentrations identified at higher redshift are likely to result from the line-of-sight superposition of smaller clusters that are associated but have not yet coalesced). We therefore find no contracticton in associating the distant clusters in the studies reviewed above with the progenitors of today's rich clusters.

6.2 An infall model for the Butcher–Oemler effect

Several models have been proposed to account for the sharp rise seen in the fraction of blue, or starburst, galaxies in rich clusters at moderate redshifts. Lavery & Henry (1988) have proposed that galaxy–galaxy collisions are responsible. Although the high-velocity dispersion of the cluster makes such interactions ineffective, Lavery & Henry suggest that the effect occurs in small galaxy groups that are compressed as they infall into the cluster potential. Alternatively, Dressler & Gunn (1983) (*cf.* Gunn 1989, for a more comprehensive review) have suggested that the infalling galaxies are triggered into star burst activity by the ram pressure of the intracluster medium. In either case, it is necessary to determine how this infall evolves with redshift. This problem may be addressed using the extension of the Press–Schechter theory that we have presented in the previous sections of this paper.

Table 2. Quantitative estimates of the Butcher–Oemler effect.

Source	Redshift Range	Type of Galaxy Fraction	f_0	$\frac{1}{f_0}\frac{\Delta f}{\Delta z}$	Comments
Butcher & Oemler (1984)	0.0–0.5	blue galaxies	3%	14.	based on photometry only: 29 clusters with $C > 0.35$
Lavery & Henry (1986)	0.2	blue cluster members	3%	28.	based on spectroscopy of 43 galaxies in 3 clusters
Couch & Sharples (1987)	0.31	blue cluster members	5%	16.	based on spectroscopy of 152 galaxies in 3 clusters
Dressler & Gunn (1987)	0.4–0.5	on-going/recent star formation	5%	11.	based on spectroscopy of 236 galaxies in 7 clusters

We argue that f_b can be directly related to the infall rate. The environment of a rich cluster is hostile to star formation due to the ram pressure of the intra-cluster medium (*cf.* Gisler 1979). Therefore, we argue that the blue galaxies must have only recently arrived, and that the number of these galaxies is proportional to the infall rate. Since the gas stripping time-scale, and the time-scale for the galaxy colour to become red after the termination of star formation are of order 1 Gyr, we should expect f_b and the infall rate per Gyr to be of a similar order of magnitude. This is indeed the case: in near-by rich, condensed clusters BO84 obtain $f_b \approx 0.03$; assuming an age for the Universe of 14 Gyr, the infall rate into a typical cluster (per unit cluster mass, per Gyr) is ~ 0.04.

We now determine how the blue galaxy fraction should be expected to vary in this simple dynamical model. Comparison with the observed Butcher–Oemler effect allows us to assess the relative importance of the internal evolution of the galaxies. If the evolution of the infall rate were sufficiently rapid that rich clusters were effectively 'formed' at a redshift of 0.3–0.5, and had subsequently reached a quiescent state at the present-day, then this model would provide an adequate explanation for the observed evolution of f_b. Our study shows, however, that the infall rate does not vary sufficiently quickly. Inspection of equation (18) reveals that the infall of small groups per unit time initially increases with redshift as $(1+z)^{5/2}$. This growth is not long sustained, the infall peaking at $z \approx \Delta_s[M_s/M_*(0)]^{-(n+3)/6}$ due to the sharp decline of the Gaussian term.

In order to make a more quantitative comparison with the observational results, we must determine reasonable values for the parameters n, $M_*(0)$ and M_s. Bahcall (1979) and Moore, Frenk & White (1990) have used the CfA redshift survey to determine the combined luminosity function of galaxies, groups and clusters (we refer to this as the All Galaxy Systems, AGS, luminosity function). If the mass-to-light ratio of galaxies is universal, then $L\Phi_{AGS}(L)$ may be directly compared with the universal (Press–Schechter) multiplicity function in order to determine values for the parameters $M_*(0)$ (given a value for M/L) and the density field spectral index, n. Moore *et al.* (1990) find that the best fit is given with $n=-1.5$ and $L_{*,AGS} = 3.9\times10^{10}\ L_\odot \approx 5\ L_{*,galaxy}$ (using $H_0 = 100$ km s^{-1} Mpc^{-1}). A large cluster today, for example Coma, has total luminosity (Moore *et al.* 1990) of about 60 $L_{*,AGS}$. M_s, the group mass at which ram pressure and/or galaxy collisions become an important driving force in galaxy evolution, is much more difficult to determine directly. However, as a typical group at the present epoch contains five average galaxies, it seems reasonable to set $M_s = M_*(0)$. Our results are not sensitive the choice of this, or any other, parameter.

In Fig. 14, we have plotted the photometrically determined blue fractions of BO84. For comparison, our theoretical infall rate, equation (18), is drawn with the choice of parameters outlined in the previous paragraph (solid line). The infall rate peaks at $z \approx 1$, well above the observed range of clusters. However, the predicted increase in the infall rate is clearly insufficient to fully account for the Butcher–Oemler effect that is observed. The insensitivity of this result to our choice of model parameters is illustrated by the dotted lines. These show the effect of altering $M_s/M_*(0)$ between 0.1 and 5.0.

We conclude that the Butcher–Oemler effect cannot be explained in terms of a 'formation epoch' for rich clusters: the increased rate at which field galaxies infall into rich clusters at moderate redshifts makes only a small contribution to the observed change in f_b. This result agrees with the conclusions that have been drawn from spectroscopic studies. As we have previously described, these studies suggest that galaxies at moderate redshifts are much more susceptible to strong bursts of starformation activity than galaxies of similar magnitude today. This leads to a higher proportion of infalling galaxies being identified as having 'blue' colours. Below we make an estimate of the importance of this effect.

At the present time, field galaxies have a mean blue fraction of only 41 ± 10 per cent. Therefore, only 40 per cent of infalling galaxies need contribute to the cluster f_b at the present epoch. Given the increased responsiveness of galaxies at higher redshift, it is possible that *all* infalling galaxies contribute to the cluster f_b at, say, $z=0.5$. A dramatic increase in f_b is obtained, f_b being increased by a factor of $1/0.4 = 2.6$ at $z=0.5$. The theoretical gradient, $(1/f_0)/(\Delta f/\Delta z)$, is increased to 10.6 – a value in acceptable aggreement, given that various subtle systematic effects may be at work (*cf.* Koo 1987), with the values quoted in Table 2. To further aid comparison of this theoretical evolution with the data, we have illustrated the growth of f_b by a dotted line in Fig. 14.

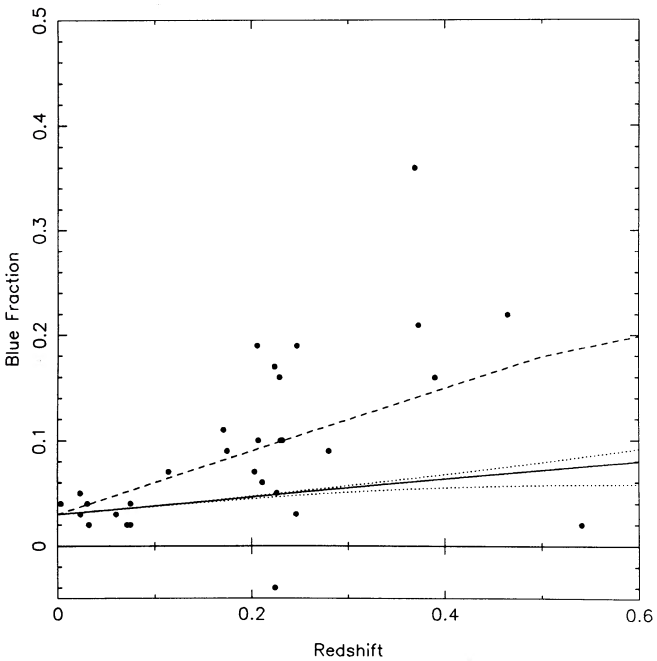

Figure 14. Comparison between the evolution of the infall rate and the evolution of the blue fraction of galaxies in rich clusters. Data points are taken from Butcher & Oemler 1984. The solid curve shows the infall rate, normalized to match the observations at $z \approx 0$. The parameters used in equation (18) are : $n=-1.5$; $M'=100\ M_*(0)$; $M_s = M_*(0)$. The dotted lines show the effect of setting $M_s = 0.1\ M_*(0)$ and $5.0\ M_*(0)$. It can be seen that the increase in the infall rate considerably underestimates the Butcher–Oemler effect seen in the blue galaxy fraction. The dashed line illustrates the growth in f_b that may be anticipated if allowance is made for the increased star formation activity of galaxies at higher redshifts.

7 DISCUSSION

In the previous sections, we have presented several basic mathematical results that describe the evolution of the galaxy clustering hierarchy. As the results have been interspersed with complex mathematical derivations, it is convenient to collect them together here.

We start by parameterizing the initial conditions in the Universe in terms of a power law of density fluctuations with spectral index n. The first of these parameters is directly related to the typical mass of collapsed objects at the present time [$M_*(0)$], and it is convenient to make this substitution in all subsequent formulae. In terms of the parameters $M_*(0)$ and n, the Press–Schechter theory predicts the form of the *Universal Multiplicity Function* of groups at epoch z to be

$$\rho(M,z)\,dM=\rho_0\frac{1}{\sqrt{2\pi}}\left(\frac{n+3}{3}\right)\left[\frac{M}{M_*(z)}\right]^{(n+3)/6} \times\exp\left\{-\frac{1}{2}\left[\frac{M}{M_*(z)}\right]^{(n+3)/3}\right\}\frac{dM}{M}, \tag{2}$$

where $M_*(z)=(1+z)^{-6/(n+3)}M_*(0)$. The evolution of this distribution is self-similar; i.e. its shape remains unchanged, the distribution only altering in its scale [defined by $M_*(z)$]. The evolution is more rapid for smaller, or more negative, values of n.

In Section 3.3, we have extended the PS formalism to calculate the multiplicity fraction of groups given the added constraint that they must merge to form a cluster of size M' at the present epoch. We refer to this as the *Conditional Multiplicity Fraction*

$$\tilde{f}(M,z|M')\,dM=\frac{1}{\sqrt{2\pi}}\left(\frac{n+3}{3}\right)\frac{1}{\Delta^3}\left[\frac{M}{M_*(z-1)}\right]^{(n+3)/6} \times\exp\left\{-\frac{1}{2\Delta^2}\left[\frac{M}{M_*(z-1)}\right]^{(n+3)/3}\right\}\frac{dM}{M}, \tag{12}$$

where $\Delta=[1-(M/M')^{(n+3)/3}]^{1/2}$. This differs from the previous function in two respects: (i) the evolution in a region of large final mass ($M'\gg M$) is 'accelerated' over that in the Universe as a whole; (ii) the correlation between density fluctuations on different scales suppresses the formation of subclumps close to M' before the final epoch. The interplay between these two effects is sensitive to the spectral index n.

Combining the Conditional Multiplicity Fraction (equation 12) with the distribution of the masses of present-day clusters (M') specified by the Universal Multiplicity Function (equation 2) as is described in Section 3.4, we obtain the multiplicity function of masses M at epoch z that combine to form a group of mass M' at the present epoch. We refer to this as the *Joint Multiplicity Function*,

$$\rho(M,z,M')\,dM\,dM'=\frac{\rho_0}{2\pi}\left(\frac{n+3}{3}\right)^2\frac{1}{\Delta^3}\left[\frac{M}{M_*(z-1)}\right]^{(n+3)/6} \times\exp\left\{-\frac{1}{2\Delta^2}\left[\frac{M}{M_*(z-1)}\right]^{(n+3)/3}\right\}\left[\frac{M'}{M_*(0)}\right]^{(n+3)/6} \times\exp\left\{-\frac{1}{2}\left[\frac{M'}{M_*(0)}\right]^{(n+3)/3}\right\}\frac{dM}{M}\frac{dM'}{M'}. \tag{14}$$

It is important to note that this multiplicity function is consistent with the universal distributions of M and M' – in Section 4.1 we demonstrate that integration of equation (14) over M' gives the Universal Multiplicity Function of M at epoch z, and vice versa.

There is an important and fundamental difference between equation (14) and any similar equation suggested by the k-split approach. In the formulae presented here, the enhancement in the hierarchy of group masses (caused by the presence of the long-wavelength perturbation that will eventually collapse to form M') is independent of M'; M' enters only through the cross-correlation between the mass-scales of M and M' (e.g. through the factor denoted Δ in the formulae). If these correlations were taken to be negligible (as in the k-split approximation), then it is tempting to allow for the total mass of the larger scale region by varying the over-density of the long-wavelength perturbation as a function of M'. This procedure is invalid – if the overdensity is averaged over the whole of M', it must be found to be $\delta_c(0)$ (i.e. the critical overdensity for collapse at the present epoch); if it were found to be larger then the region would be absorbed into the collapse of an even larger object. As a result of the manner in which M' enters our formulae, the distribution of M is largely independent of M' if the scales are sufficiently widely separated that cross-correlation effects are unimportant (i.e. in the region where the k-split is a good approximation). This has important consequences when we come to compare the evolutionary histories of groups of differenting M'.

In order to illustrate the evolution of groups by a simple example, we divide groups into two mass regimes at M_s. In doing this we have in mind a simple model for the environmentally driven evolution of galaxies. We calculate the following.

(i) The average epoch, z_s, at which galaxies make the transition between groups above and below M_s (given that at $z=0$ they become bound into a group of mass M'),

$$z_s=\frac{0.97}{[M_s/M_*(0)]^{(n+3)/6}}\Delta_s, \tag{17}$$

where the dependence of z_s on the mass M' enters through

$$\Delta_s=\left[1-\left(\frac{M_s}{M'}\right)^{(n+3)/3}\right]^{1/2}.$$

(ii) The rate (in units of the age of the Universe, and scaled to the mass of the present-day cluster) at which galaxies from low-mass groups infall into the growing proto-cluster,

$$R_{\rm infall}=\frac{2}{\sqrt{2\pi}}\frac{[M_s/M_*(0)]^{(n+3)/6}}{\Delta_s}\frac{2}{3}(1+z)^{5/2} \times\exp\left\{-\frac{z^2}{2\Delta_s^2}\left[\frac{M_s}{M_*(0)}\right]^{(n+3)/3}\right\}. \tag{18}$$

(iii) The rate at which galaxies from low-mass groups infall into a cluster at the present time [this is a special case of

equation (2)],

$$R_{\rm infall}(0)=\frac{4}{3\sqrt{2\pi}}\frac{[M_{\rm s}/M_*(0)]^{(n+3)/6}}{\Delta_{\rm s}} \tag{19}$$

(iv) The present-day distribution (in group mass M') of the mas contained in groups that exceed the mass-scale $M_{\rm s}$ at, or before, redshift z,

$$f(M'|>M_{\rm s},z)\,dM'=\frac{1}{\sqrt{2\pi}}\left(\frac{n+3}{3}\right)\left[\frac{M'}{M_*(0)}\right]^{(n+3)/6} \tag{21}$$

$$\times\exp\left\{-\frac{1}{2}\left[\frac{M'}{M_*(0)}\right]^{(n+3)/3}\right\}$$

$$\times\frac{\mathrm{erfc}\{1/\sqrt{2}\Delta_{\rm s}[M_{\rm s}/M_*(z-1)]^{(n+3)/6}\}dM'}{\mathrm{erfc}\{1/\sqrt{2}[M_{\rm s}M_*(z)]^{(n+3)/6}\}\quad M'}$$

Inspection of results (1)–(3) above shows that the history of a group of present-day mass M' (on the scale set by $M_{\rm s}$) is strongly dependent on M' only if M' is close to $M_{\rm s}$. As we have discussed above, this effect arises because the mass of the large-scale overdense region only modulates the smaller scale fluctuations through the cross-correlation of the density fluctuations (when averaged throughout the entire region). If M' is significantly larger than $M_{\rm s}$, the effect is very weak. Study of the equations also shows that the modulation is stronger for flatter power spectra (i.e. more negative n). However, it should be remembered that the rate of universal evolution is also more rapid in this case.

While we have shown that the more massive present-day clusters do not have significantly different average evolutionary histories (on group and galaxy mass-scales) from their less massive counter parts, there is a strong tendency for the most massive groups at early times to be incorporated into the most massive clusters at the present-day (result 4). These two statements may appear contradictory, but they are not. Confusion has arisen because we have not been particular to distinguish between masses that have been specified as a fraction of the present-day cluster mass (e.g. in our calculation of $z_{\rm s}$, equation 17), and those that have been specified in absolute value (e.g. in equation 21).

In Section 7 of this paper, we have illustrated the utility of our results by considering the Butcher–Oemler effect. We have examined whether the increasing blue fraction of galaxies in rich clusters at moderate redshifts can be explained solely by the evolution of infall rate into these clusters. Our analysis shows that while the infall rate of field galaxies does increase over these look-back times, the effect is not, by itself, sufficient to account for the rapid rise in blue fraction that is observed. If, however, allowance is made for the increased star formation activity seen in the spectra of the galaxies, it is possible to bring the model into good quantitative agreement with observation.

It is readily possible to apply the mathematical results presented in this paper to further problems in modern cosmology. Such applications will be considered in subsequent papers.

ACKNOWLEDGMENTS

I would like to thank, and acknowledge, Carlos Frenk for the many long and detailed discussions that we have held during the course of my post-graduate study at Durham University. In addition, I would like to extend many thanks to Simon White for his comments on the preprint version of this paper. The receipt of an SERC studentship is also gratefully acknowledged. I am much indebted to Mike Fitchett, Stephen Lorrimer and Karen Brazier for their help in the preparation of this manuscript.

REFERENCES

Adler, R. J., 1981. *The Geometry of Random Fields,* Wiley, Chichester.
Bahcall, N. A., 1979. *Astrophys. J.,* **232,** 689.
Bardeen, J. M., Bond, J. R., Kaiser, N. & Szalay, A. S., 1986. *Astrophys. J.,* **304,** 15.
Bond, J. R. & Efstathiou, G., 1984. *Astrophys. J.,* **285,** L45.
Bond, J. R., Cole, S., Efstathiou, G. & Kaiser, N., 1990. *Astrophys. J.,* submitted.
Butcher, H. & Oemler, A., 1978. *Astrophys. J.,* **219,** 18.
Butcher, H. & Oemler, A., 1984. *Astrophys. J.,* **285,** 426 (BO84).
Couch, W. J. & Newell, E. B., 1984. *Astrophys. J. Suppl.,* **56,** 153.
Couch, W. J. & Sharples, R. M., 1987. *Mon. Not. R. astr. Soc.,* **229,** 423.
Davis, M., Efstathiou, G., Frenk, C. S. & White, S. D. M., 1985. *Astrophys. J.,* **292,** 371.
Dressler, A. & Gunn, J. E., 1982. *Astrophys. J.,* **263,** 533.
Dressler, A. & Gunn, J. E., 1983. *Astrophys. J.,* **270,** 7 (DG83).
Dressler, A. & Gunn, J. E., 1987. In: *IAU Symposium No. 130, Large Scale Structure of the Universe,* p. 311, eds Audouze, J., Pelletan, M. & Szalay, A., Reidel, Dordrecht.
Dressler, A., Gunn, J. E. & Schneider, D. P., 1985. *Astrophys. J.,* **294,** 70 (DGS).
Efstathiou, G., Frenk, C. S., White, S. D. M. & Davis, M., 1988. *Mon. Not. R. astr. Soc.,* **235,** 715 (EFWD).
Gisler, G. R., 1979. *Astrophys. J.,* **228,** 385.
Gunn, J. E., 1989. In: *The Epoch of Galaxy Formation,* p. 167, eds Frenk, C. S. *et al.,* Kluwer, Dordrecht.
Gunn, J. E. & Gott, J. R., 1972. *Astrophys. J.,* **176,** 1.
Guth, A. H., 1981. *Phys. Rev. D,* **23,** 347.
Kaiser, N., 1986. *Mon. Not. R. astr. Soc.,* **222,** 323.
Koo, D. C., 1987. In: *Towards Understanding Galaxies at Large Redshift,* p. 275, eds Kron, R. & Renzini, A., Reidel, Dordrecht.
Lavery, R. J. & Henry, J. P., 1986. *Astrophys. J.,* **304,** L5.
Lavery, R. J. & Henry, J. P., 1988. *Astrophys. J.,* **330,** 596.
Lynden-Bell, D., 1967. *Mon. Not. R. astr. Soc.,* **136,** 101.
Mellier, Y., Soucail, G., Fort, B. & Mathez, G., 1988. *Astr. Astrophys.,* **199,** 13.
Moore, B., Frenk, C. S. & White, S. D. M., 1990. In press.
Moore, B., Frenk, C. S. & White, S. D. M., 1990. *Mon. Not. R. astr. Soc.,* submitted.
J., **335,** 629.
Peacock, J. A. & Heavens, A. F., 1990. *Mon. Not. R. astr. Soc.,* **243,** 133.
Peebles, J., 1980. *The Large Scale Structure of the Universe,* Princeton University Press, Princeton.
Press, W. H. & Schechter, P., 1974. *Astrophys. J.,* **187,** 425 (PS).
Shandarin, S. F., Doroshkevich, A. G. & Zel'dovich, Ya. B., 1983. *Soviet Phys. Usp.,* **26,** 46.
White, S. D. M. & Frenk, C. S., 1990. *Mon. Not. R. astr. Soc.,* submitted.
White, S. D. M., Davis, M., Efstathiou, G. & Frenk, C. S., 1987. *Nature,* **330,** 451.

APPENDIX A: COMPUTATION OF THE CROSS-CORRELATION BETWEEN SCALES

In order to establish the method by which we will determine the co-variance of the density within V and V' ($\sigma^2_{VV'}$), we review the technique (Peebles 1980) by which the variance of the fluctuations on scale V can be determined. Initially, we adopt the notion that the Universe is periodic on some scale much larger than V. This allows us to decompose the density field into discrete Fourier components. The overdensity at some point, $\boldsymbol{x}$, may then be written as

$$\delta(\boldsymbol{x}) = \sum_{\boldsymbol{k}} \delta_{\boldsymbol{k}} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}}.$$

The average density inside a volume V ('centred' at $\boldsymbol{x}_0$) is

$$\delta_V = \frac{1}{V}\int_V \sum_{\boldsymbol{k}} \delta_{\boldsymbol{k}} \mathrm{e}^{-i\boldsymbol{k}\cdot(\boldsymbol{x}_0+\boldsymbol{x}')}\, \boldsymbol{d}^3\boldsymbol{x}'.$$

The variance of the density inside V is the mean square value of δ_V; the average of δ_V itself being zero by definition. Writing this explicitly we have

$$\sigma^2(V) = \frac{1}{V^2}\left\langle \left| \int_V \sum_{\boldsymbol{k}} \delta_{\boldsymbol{k}} \mathrm{e}^{-i\boldsymbol{k}\cdot(\boldsymbol{x}_0+\boldsymbol{x}')}\, \boldsymbol{d}^3\boldsymbol{x}' \right|^2 \right\rangle,$$

where the average, denoted by the angle brackets, is taken over all possible positions, $\boldsymbol{x}_0$, of the volume, V (i.e. the entire Universe). Expansion of the square produces terms of the form $\mathrm{e}^{-i(\boldsymbol{k}_1-\boldsymbol{k}_2)\boldsymbol{x}_0}$. Only if $\boldsymbol{k}_1 = \boldsymbol{k}_2$ will the term make a net contribution to $\sigma^2(V)$ when the average is taken over $\boldsymbol{x}_0$. We can therefore simplify the expression to

$$\sigma^2(V) = \frac{1}{V^2}\sum_{\boldsymbol{k}} |\delta_{\boldsymbol{k}}|^2 \left| \int_V \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}'}\, \boldsymbol{d}^3\boldsymbol{x}' \right|^2.$$

In the continuous limit, this becomes

$$\sigma^2(V) = \int_{\boldsymbol{k}} |\delta_{\boldsymbol{k}}|^2 |W_{\boldsymbol{k}}|^2\, \boldsymbol{d}^3\boldsymbol{k}, \tag{22}$$

where the window function, $W_{\boldsymbol{k}} = \int_V \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}'}\, \boldsymbol{d}^3\boldsymbol{x}' / \int_V \boldsymbol{d}^3\boldsymbol{x}'$, is the Fourier transform of the top-hat function that defines the volume V. (The term 'top-hat function' is used loosely here – we do not mean to imply that the volume V needs to be spherical.)

The cross-correlation between the overdensities in two volumes, one contained inside the other, is calculated in a similar manner. Starting from the definition of the co-variance as $\sigma^2(VV') = \langle \delta_V \delta_{V'} \rangle$ (the average being taken over all admissible pairs of volumes V and V') we have

$$\sigma^2(VV') = \frac{1}{VV'}\left\langle \left\langle \left[\int_V \sum_{\boldsymbol{k}} \delta_{\boldsymbol{k}} \mathrm{e}^{-i\boldsymbol{k}\cdot(\boldsymbol{x}_0+\boldsymbol{x}')}\, \boldsymbol{d}^3\boldsymbol{x}' \right] \times \left[\int_{V'} \sum_{\boldsymbol{k}} \delta_{\boldsymbol{k}} \mathrm{e}^{-i\boldsymbol{k}\cdot(\boldsymbol{y}_0+\boldsymbol{y}')}\, \boldsymbol{d}^3\boldsymbol{y}' \right]^* \right\rangle_{\boldsymbol{x}_0} \right\rangle_{\boldsymbol{y}_0},$$

where the average over $\boldsymbol{y}_0$ is taken over all space, and the average over $\boldsymbol{x}_0$ is taken over values of $\boldsymbol{x}_0$ for which the volume V is contained in V' (as we discuss later, there is also an independence criterion that limits the choice of $\boldsymbol{x}_0$). We continue by separating $\boldsymbol{x}_0$ into two components ($\boldsymbol{x}_0 = \boldsymbol{y}_0 + \boldsymbol{u}_0$):

$$\sigma^2(VV') = \left\langle \left\langle \left[\sum_{\boldsymbol{k}} \delta_{\boldsymbol{k}} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{y}_0} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{u}_0}\, \boldsymbol{d}^3\boldsymbol{x}'\, W_{\boldsymbol{k}}(V) \right] \times \left[\sum_{\boldsymbol{k}} \delta_{\boldsymbol{k}} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{y}_0}\, \boldsymbol{d}^3\boldsymbol{x}'\, W_{\boldsymbol{k}}(V') \right]^* \right\rangle_{\boldsymbol{u}_0:\, V\in V'} \right\rangle_{\boldsymbol{y}_0}.$$

We are now free to swap the order in which the averages are taken. Continuing as previously, we obtain

$$\sigma^2(VV') = \int_{\boldsymbol{k}} |\delta_{\boldsymbol{k}}|^2\, W_{\boldsymbol{k}}(V)\, W^*_{\boldsymbol{k}}(V') \langle \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{u}_0} \rangle_{\boldsymbol{u}_0}\, \boldsymbol{d}^3\boldsymbol{k}. \tag{23}$$

Care is required at this point to define the meaning of the 'average' over $\boldsymbol{u}_0$ more exactly. $\boldsymbol{u}_0$ is a randomly chosen point for which the volume V that surrounds it is entirely enclosed in the volume V'. However, it is not clear whether we should: (i) choose a point at random, test to see if V is contained in V', then reselect another random point if this is not the case; or (ii) select a random set of points $\boldsymbol{u}_0$ for which all the volumes V are independent and enclosed in V', then choose a point at random from this set. The practical difference is that if method (i) is adopted, the overdensity at a point near the outer edge of V' is less likely to contribute to the mean overdensity of V than a point that is close to the centre; with method (ii) all points in V' have an equal chance to contribute to the mean overdensity of V. Later, when we come to extend the Press–Schechter theory, we will wish to apply the present calculation to determine the probability of collapse of a region (e.g. the volume V). On collapse, the matter in this region is not available to be included in an adjacent region (unless both regions collapse to form a single object on a larger scale, this circumstance being taken into account explicitly in the theory). We therefore argue that the volumes V in the averaging process must also be given this property of independence, and use method (ii) to define the process by which we average over $\boldsymbol{u}_0$. The important consequence of this decision is that close to the edge of $\tilde{V}$ (the volume covered by the points $\boldsymbol{u}_0$ for which V lies inside V') there are more independent choices for V per volume element, $\boldsymbol{d}^3\boldsymbol{u}_0$, as part of V lies outside $\tilde{V}$ and therefore does not exclude other possible values of $\boldsymbol{u}_0$.

The averaged term, $\langle \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{u}_0} \rangle_{\boldsymbol{u}_0}$, can be written to appear like another window function,

$$\tilde{W}_{\boldsymbol{k}}(\tilde{V}) = \int_{\tilde{V}} w(\boldsymbol{u}_0) \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{u}_0}\, \boldsymbol{d}^3\boldsymbol{u}_0 \Big/ \int_{\tilde{V}} w(\boldsymbol{u}_0)\, \boldsymbol{d}^3\boldsymbol{u}_0,$$

where $\tilde{V}$ is the volume covered by the points $\boldsymbol{u}_0$ for which V lies inside V'; $w(\boldsymbol{u}_0)$ is a weight function that gives increased weight to outer values of u_0 to account for the fact that the selection has been made as described in method (ii).

It is possible for us to express $\tilde{W}_{\boldsymbol{k}}(\tilde{V})$ in terms of the window functions of V and V'. We introduce the functions $\Theta(\boldsymbol{x}; V, \boldsymbol{x}_0)$, that we define as

$$\Theta(\boldsymbol{x}; V, \boldsymbol{x}_0) = \begin{cases} 1/V & \text{if } \boldsymbol{x} \text{ is contained in the volume } V \text{ 'centred' on } \boldsymbol{x}_0, \\ 0 & \text{otherwise.} \end{cases}$$

The k-space window functions W_k are the Fourier transforms of these real-space window functions

$$W_k(V)=\int_V \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}'}\,d^3\boldsymbol{x}'=\int_\infty \Theta(\boldsymbol{x}';V,0)\mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{x}'}\,d^3\boldsymbol{x}',$$

where we have arbitrarily chosen to 'centre' the volume V at the origin. By analogy, we also introduce the function $\tilde{\Theta}(\boldsymbol{x};\tilde{V},\boldsymbol{x}_0)$, which we define as

$$\tilde{\Theta}(\boldsymbol{x};\tilde{V},\boldsymbol{x}_0)=\begin{cases} w(\boldsymbol{x})/\int_{\tilde{V}} w(\boldsymbol{x})\,d^3\boldsymbol{x} & \text{if } \boldsymbol{x} \text{ is contained in the volume } \tilde{V} \\ & \text{'centred' on } \boldsymbol{x}_0; \\ 0 & \text{otherwise.}\end{cases}$$

Note that the normalization of these functions is such that $\int_\infty \Theta(\boldsymbol{x}';V,\boldsymbol{x}_0)\,d^3\boldsymbol{x}'=1$, and similarly for $\tilde{\Theta}(\boldsymbol{x};\tilde{V},\boldsymbol{x}_0)$.

We have defined the weighting function, $w(\boldsymbol{x})$, of $\tilde{V}$ so that every point in the large volume V' has an equal chance of contributing to the average density measured in V. This is equivalent to stating that we can write the function $\Theta(\boldsymbol{x};V',\boldsymbol{x}_0)$ as the convolution of $\Theta(\boldsymbol{x};V,\boldsymbol{x}')$ and $\tilde{\Theta}(\boldsymbol{x}';\tilde{V},\boldsymbol{x}_0)$:

$$\Theta(\boldsymbol{x};V',0)=\int_\infty \tilde{\Theta}(\boldsymbol{x}';\tilde{V},0)\Theta(\boldsymbol{x}-\boldsymbol{x}';V,0)\,d^3\boldsymbol{x}'.$$

Then in Fourier space

$$W_k(V')=\tilde{W}_k(\tilde{V})\,W_k(V). \tag{24}$$

In the general case, however, it is not possible to find a real bounded weighting function $w(\boldsymbol{x})$ that has this property. For example, if V and V' are of comparable size, all volumes V will include points near the centre of V'. Nevertheless, it is possible to define $w(\boldsymbol{x})$ so that the convolution of $\Theta(\boldsymbol{x};V,\boldsymbol{x}')$ and $\tilde{\Theta}(\boldsymbol{x}';\tilde{V},\boldsymbol{x}_0)$ closely approximates $\Theta(\boldsymbol{x};V',\boldsymbol{x}_0)$ in all but its high-$\boldsymbol{k}$ components. For spherical top-hat functions, the approximation is accurate up to $\boldsymbol{k}_m>4.5/R'$. Such high-$\boldsymbol{k}$ waves make negligible contribution to the integral in equation (23).

Returning to our expression for the co-variance of the volumes V and V', equation (23), and using the relation between the window functions that we have derived above (equation 24), we obtain

$$\sigma^2(VV')=\int_{\boldsymbol{k}} |\delta_k|^2\,|W_k(V')|^2\,d^3\boldsymbol{k}=\sigma^2(V'). \tag{6}$$

As is illustrated in the main text, this strikingly simple result is trivial to derive if the mean overdensity of the small volumes V contained in V' is assumed to equal the overdensity of region V' as a whole.

APPENDIX B: INVERSION OF THE JOINT MULTIPLICITY FUNCTION

In order to demonstrate the self-consistency of this work, we are required to show that integration of the joint multiplicity function over all possible final masses, M', recovers the universal multiplicity function, $\rho(M,z)$. The required integral may be written explicitly as

$$\begin{aligned} I=\int_{M'=M}^{\infty}\rho(M,z,M')\,dM\,dM' &= \frac{\rho_0}{2p}\left[\frac{(n+3)}{3}\right]^2\frac{dM}{M} \\ &\times\int_{M'=M}^{\infty}\frac{1}{\Delta^3}\left[\frac{M}{M_*(z-1)}\right]^{(n+3)/6} \\ &\times\exp\left\{-\frac{1}{2\Delta^2}\left[\frac{M}{M_*(z-1)}\right]^{(n+3)/3}\right\}\times\left[\frac{M'}{M_*(0)}\right]^{(n+3)/6} \\ &\times\exp\left\{-\frac{1}{2}\left[\frac{M'}{M_*(0)}\right]^{(n+3)/3}\right\}\frac{dM'}{M'}. \end{aligned} \tag{25}$$

While it is not possible to solve this integral by standard analytic techniques, numerical integration shows that it is at least a very close approximation to $\rho(M,z)\,dM$. Below, we are able to demonstrate, by unusual analytic reasoning, that the inversion is indeed exact.

As a first two, we simplify the expression for the joint multiplicity function by making the transformations

$$u'=[M'/M_*(0)]^{(n+3)/6},$$

and

$$\begin{aligned} u&=[M/M_*(0)]^{(n+3)/6} \\ &=z^{-1}[M/M_*(z-1)]^{(n+3)/6} \\ &=(1+z)^{-1}[M/M_*(z)]^{(n+3)/6}. \end{aligned} \tag{26}$$

In these variables,

$$\rho(u,z,u')\,du\,du'=\frac{4\rho_0}{2\pi}\frac{1}{\Delta^3}z\exp\left(-\frac{z^2u^2}{2\Delta^2}\right)\times\exp\left(-\frac{u'^2}{2}\right)du\,du',$$

with $\Delta=\sqrt{1-u^2/u'^2}$. Our aim is to integrate this expression over $u'=u$ to ∞. Progress can be made by noting that in these variables the universal multiplicity function (equation 2) becomes

$$\rho(u,z)\,du=\frac{2}{\sqrt{2\pi}}1\,(1+z)\exp\left[-\frac{(1+z)^2u^2}{2}\right]du. \tag{27}$$

This suggests that we attempt to simplify the exponent of the integrand by the substitution

$$x^2=\frac{z^2u^2}{(1-u^2/u'^2)}-(1+z)^2u^2+u'^2.$$

This expression reduces to

$$x=\frac{u'^2-(1+z)u^2}{\sqrt{u'^2-u^2}}. \tag{28}$$

Differentiating, we obtain

$$\frac{dx}{du'}=\frac{u'^3}{(u'^2-u^2)^{3/2}}\left[1+\frac{u^2}{u'^2}(z-1)\right]=\frac{1}{\Delta^3}\left[1+\frac{u^2}{u'^2}(z-1)\right].$$

This is close to the Δ^{-3} factor in the integrand. For the special case $z=1$ the agreement is exact and the integral, rewritten in terms x, is of standard form.

We have not yet considered the limits to the integral in x-space. As $u' \to u$, $x \to -\infty$; at the other limit $x \to +\infty$. Also, $x=0$ at $u'=(1+z)^{1/2}u$, but note that u'^2 is not a single-valued function of x^2, so that it is best to proceed by splitting the integral into two parts corresponding to the two branches of u':

$$I=\frac{4}{2\pi} z\, e^{-(1+z)^2u^2/2} \times\left\{\int_{0,\, u'^2>(1+z)u^2}^{\infty} \frac{e^{-x^2/2}}{[1+(u^2/u'^2)(z-1)]}\,dx + \int_{0,\, u'^2<(1+z)u^2}^{\infty} \frac{e^{-x^2/2}}{[1+(u^2/u'^2)(z-1)]}\,dx\right\}. \tag{29}$$

For $z=1$, both parts integrate to $\sqrt{2\pi}/2$, and since, in this case, $2z \equiv z+1$, we recover the universal distribution of M exactly. For other values of z, we note that the major contribution to the integral comes from $x \sim 0$.

We continue by combining the two integrals, using the notation u'^2_+ and u'^2_- to distinguish the two values of u'^2 [one greater than $(1+z)u^2$, the other less] corresponding to each value of x. Squaring and expanding equation (28) we obtain the quadratic equation for which u'^2_+ and u'^2_- are the roots,

$$u'^4-[2(1+z)u^2+x^2]u'^2+[(1+z)^2u^2+x^2]u^2=0.$$

The roots of this equation (i.e. u'^2_+ and u'^2_-) satisfy the relations

$$u'^2_+u'^2_-=[(1+z)^2u^2+x^2]u^2,$$

and,

$$u'^2_++u'^2_-=[2(1+z)u^2+x^2]$$

Combining the separate integrals in equation (29) we have

$$I=\frac{4}{2\pi} z\, e^{-(1+z)^2u^2/2} \times\int_{x=0}^{\infty} e^{-x^2/2}\left\{\frac{1}{[1+(u^2/u'^2_+)(z-1)]}+\frac{1}{[1+(u^2/u'^2_-)(z-1)]}\right\}dx \tag{30}$$

(30)

The bracketed part of the integrand may be rewritten by expressing the fractions in terms of a common denominator and using the expressions for the sums and products of u'^2_+ and u'^2_- to eliminate u':

$$\frac{2+(z-1)[x^2+2(1+z)u^2/x^2+(1+z)^2u_2]}{1+(z-1)[x^2+2(1+z)u^2/x^2+(1+z)^2u^2]+[u^2(z-1)^2/x^2+(1+z)^2u^2]}$$

After considerable manipulation, this expression can be reduced to

$$\frac{(1+z)x^2+4z(1+z)u^2}{zx^2+4z^2u^2}=\frac{1+z}{z},$$

i.e. this term is in fact independent of x.

Replacing the bracketed term in equation (30) we obtain

$$I=\frac{4}{2\pi} z\, e^{-(1+z)^2u^2/2}\int_{x=0}^{\infty}\frac{1+z}{z}e^{-x^2/2}\,dx$$

$$=\frac{2}{\sqrt{2\pi}}(1+z)\exp\left[-\frac{(1+z)^2u^2}{2}\right].$$

Comparison with equation (27) shows that we have *exactly* recovered the universal multiplicity function for the masses M. We emphasize that this is an extremely striking result.