Cosmological evolution and hierarchical galaxy formation

Will Percival and Lance Miller

Department of Physics, University of Oxford, Nuclear & Astrophysics Laboratory, Keble Road, Oxford OX1 3RH

Accepted 1999 June 11. Received 1999 June 2; in original form 1998 September 18

ABSTRACT

We calculate the rate at which dark matter haloes merge to form higher mass systems. Two complementary derivations using Press–Schechter theory are given, both of which result in the same equation for the formation rate. First, a derivation using the properties of the Brownian random walks within the framework of Press–Schechter theory is presented. We then use Bayes’ theorem to obtain the same result from the standard Press–Schechter mass function. The rate obtained is shown to be in good agreement with results from Monte Carlo and N-body simulations. We illustrate the usefulness of this formula by calculating the expected cosmological evolution in the rate of star formation that is due to short-lived, merger-induced starbursts. The calculated evolution is well-matched to the observed evolution in ultraviolet luminosity density, in contrast to the lower rates of evolution that are derived from semi-analytic models that do not include a dominant contribution from starbursts. Hence we suggest that the bulk of the observed ultraviolet starlight at \( z > 1 \) arises from merger-induced starbursts. Finally, we show that a simple merging-halo model can also account for the bulk of the observed evolution in the comoving quasar space density.

Key words: galaxies: active - galaxies: formation - galaxies: starburst - cosmology: theory.

1 INTRODUCTION

If the matter content of the Universe is dominated by cold dark matter, then dark matter haloes associated with galaxies are expected to form hierarchically. Bond et al. (1991) have shown that a more rigorous treatment of the work of Press & Schechter (1974) can be used to obtain information about the build-up of structure. Calculations using Press–Schechter (PS) theory are based on a statistical analysis of the initial field of density perturbations and do not include any non-linear dynamical effects, but nonetheless the results appear to agree well with dynamical, collisionless (‘N-body’) simulations (Lacey & Cole 1994; Somerville et al. 1998).

The aim of this paper is to use PS theory to calculate the rate of formation of dark haloes of specified mass. We compare the results with those of Monte Carlo and N-body simulations, and we also use the calculations to argue that the bulk of the observed cosmological evolution in star-formation and quasar activity is a reflection of the evolution in rate of dark halo formation.

A standard application of PS theory is to calculate the mass function given a cosmic epoch of interest. In many cosmological situations we are also interested in the rate at which haloes of some mass form. Consequently, there have been a number of attempts to estimate this rate using PS theory (Lacey & Cole 1993; Sasaki 1994). In this paper we use the theory of random walks within the PS framework to calculate directly the halo formation rate for any cosmology. We also use Bayes’ theorem to derive the same equation from the standard PS mass function. These complementary methods provide important insights into how the formation of haloes can be understood within PS theory.

In Section 5 we use a Monte Carlo realization of actual Brownian random walks to calculate the formation epochs of the haloes they represent. The distribution of these epochs is found to be in good agreement with the formula calculated previously. Following this we show that the formation time distribution of haloes in a large N-body simulation (found using a standard friend-of-friends algorithm) is also well approximated by the PS result.

We then discuss the relevance of this work to our understanding of the observed cosmological evolution in star-formation rate (SFR) and quasar space density. In particular we suppose that there is a causal link between the hierarchical formation of dark haloes and the activation of both quasars and luminous bursts of star formation. The possibility of a connection between these two key evolving quantities has already been suggested by a number of authors (Shaver et al. 1996; Dunlop 1997; Boyle & Terlevich 1998). Mergers and interactions have long been implicated in both luminous starbursts and active galaxies (see the review by Barnes & Hernquist 1992 and references therein) and merging galaxies have been found in which both phenomena are observed (Canalizo & Stockton 1997; Stockton, Canalizo & Close 1998; Brotherton et al. 1999). Investigation of the galaxies found at redshifts...
around 3 by the Lyman-break method (Steidel et al. 1998a) show that much of the ultraviolet light inferred to be due to star formation can be identified with individual galaxies with star formation rates of possibly up to 1000 M\(_\odot\) per year (Steidel et al. 1998b). These galaxies are strongly clustered and hence are likely to be the progenitors of cluster galaxies at the present epoch (Steidel et al. 1998a,b; Adelberger et al. 1998); at high redshifts there is therefore direct evidence for a link between the high level of star formation and host galaxies which are inferred to be merging into higher mass galaxy systems.

The detailed physics of both quasar- and star-formation is complicated and not understood, and in this paper we discuss only the role that the cosmological variation in halo formation rate might have in determining the evolution in observed quantities such as the ultraviolet luminosity density arising from star formation.

2 PRESS–SCHECHTER THEORY

We now describe briefly the principles of Press–Schechter (PS) theory and calculate the number density of haloes at a particular epoch – the standard PS formula; a result which has been shown to be in good agreement with N-body simulations (e.g. Efstathiou et al. 1988).

Dark haloes are assumed to form by the non-linear gravitational collapse of initial density perturbations. We assume that the initial perturbations form a homogeneous, isotropic Gaussian random field. In PS theory such a field is smoothed by convolving with a filter function whose size is related to the mass \( M \) of halo in which we are interested. The fractional overdensity at any location is assumed to grow linearly until a critical overdensity \( \delta_c \) is reached, when that location is considered to have collapsed into a dark halo of mass \( M \), provided that the critical overdensity is not exceeded when the field is filtered on a scale corresponding to a larger mass. In the analysis that follows, instead of viewing the field as growing with time, we consider the field to be fixed and the critical overdensity \( \delta_c \) to decrease with time.

If we use a sharp \( k \)-space filter of radius \( R \), \( W(\kappa, R) = [\sin(\kappa R) - (\kappa R) \cos(\kappa R)]/[\kappa R]^3 \), the density of the filtered field at any point is given by a Brownian random walk with the variance of the filtered field being the ‘time’ axis (Peacock & Heavens 1990; Bond et al. 1991). The mass associated with this point in space is given by the position of the first upcrossing of an absorbing barrier at \( \delta = \delta_c(t) \). The probability density function that a trajectory will have its first upcrossing at a mass \( M \) and \( M + \Delta M \), \( P(M|\delta_c) \Delta M \) is given by a solution of the diffusion equation with an absorbing boundary condition (Bond et al. 1991):

\[
P(M|\delta_c) \Delta M = \frac{\delta_c}{(2\pi)^{1/2} \sigma_M^3} \exp \left( -\frac{\delta_c^2}{2\sigma_M^2} \right) \frac{\Delta M}{dM} 
\]

where \( \sigma_M^2 \) is the variance of the filtered field which is given by:

\[
\sigma_M^2 = \frac{1}{2m^2} \int_0^{\#} k^2 P(k) dk 
\]

where \( P(k) \) is the power spectrum of the initial random field. We can use equation (1) to obtain the comoving number density of haloes of mass \( M \) at a given time \( t \), \( n(M|t) \):

\[
n(M|t) M dM = \rho P[M|\delta_c(t)] \Delta M = 2\rho \frac{\delta_c(t)}{(2\pi)^{1/2} \sigma_M^3} \exp \left( -\frac{\delta_c(t)^2}{2\sigma_M^2} \right) \frac{dM}{dM} 
\]

This is the standard PS mass function. It should be noted that when using other filters this result is not correct (Bond et al. 1991; Jedamiski 1995; Yano, Nagashima & Gouda 1996).

For a sharp filter, we can relate the mass of a halo to the filter radius, \( M = 4/3\pi\rho R^3 \). We can also calculate the critical overdensity for collapse for a uniform spherical region; for a flat \( \Omega = 0 \) universe this is given by \( \delta_c = \alpha(1+z) \), where \( \alpha \sim 1.68 \) (Gunn & Gott 1972). For an open \( \Omega = 0 \) universe, \( \delta_c(\zeta) \) is given by Lacey & Cole (1993), and for a flat \( \Omega \neq 0 \) universe by Eke et al. (1996). For a sharp \( k \)-space filter, the relationship between the mass and filter size is less obvious. Following Lacey & Cole (1993), we integrate the filter function over all space to obtain its ‘volume of influence’, which gives \( M = 3/4\pi\rho R^3 \). However, if we use the critical overdensity applicable to the top-hat filter used to normalize the power spectrum the \( k \)-space filter predicts a different number density of haloes. We therefore choose the critical overdensity associated with a sharp \( k \)-space filter to be that which predicts the same number density of haloes for both filters at the mass used to normalize the power spectrum.

3 DERIVATION OF THE HALO FORMATION RATE FROM PS THEORY

We now use PS theory to obtain more information about the build-up of structure, in particular to calculate the rate at which dark haloes of a particular mass form. We cannot calculate this simply by taking the time derivative of equation (3) as, in PS theory, haloes are not only being continually formed at any particular mass but are also continually being lost into haloes of higher mass.

We present new work to calculate the required halo formation rate by examining properties of the Brownian random walks invoked in standard PS theory. We use the fact that the rate of halo formation is equivalent to a probability density function in time.

The analysis of Brownian random walks is important in many fields of pure and applied science and similar results to those derived below can also be found in many introductory books on stochastic processes (e.g. Karlin & Taylor 1975).

Initially, suppose that we are not interested in how the halo formed so we are not interested in the shape of the trajectory up until the epoch of halo formation. We wish to calculate the distribution of cosmic epochs at which haloes of a given mass \( M \) are first formed, \( P(t|M) dt \). In PS theory, this probability density function is related by a variable transformation to \( P(\delta_c|\sigma_M^2) \), the probability that a trajectory has its first upcrossing between \( \delta_c \) and \( \delta_c + \delta_c^\text{lim} \) at a given value of \( \sigma_M^2 \), and it is a formula for this that we now derive using the trajectories approach. We let \( \delta(\sigma_M^2) \) be the position of the walk at \( \sigma_M^2 \) and define:

\[
\delta_\text{max}(\sigma_M^2) = \max\{\delta(\sigma_M^2); 0 \leq \sigma_M^2 \leq \sigma_M^2\}, \\
\delta_\text{lim}(\sigma_M^2) = \delta(\sigma_M^2) - \delta(\sigma_M^2) 
\]
By considering the reflection of trajectories about the line \( \delta = m \) beyond the first upcrossing of this line, we can see that there is a one-to-one correspondence between trajectories for which \( \delta(\sigma^2_m) > 2m - x \) and trajectories for which \( \delta_{\max}(\sigma^2_m) > m \) and \( \delta(\sigma^2_m) < x \) where \( x < m \). For a pictorial representation of this correspondence see Fig. 1. This means that:

\[
P(\delta_{\max}(\sigma^2_m) > m, \delta(\sigma^2_m) < x) = P(\delta(\sigma^2_m) > 2m - x, \delta(\sigma^2_m) < x)
\]

\[
= \frac{1}{2} \left[ 1 - \text{erf} \left( \frac{2m - x}{(2\sigma^2_M)^{1/2}} \right) \right], \quad m > 0, m > x.
\]  

(6)

We differentiate with respect to \( x \) and then with respect to \( m \), changing the sign, to get the joint density function for \( \delta_{\max}(\sigma^2_m) \) lying in the interval \( m, m + dx \) and \( \delta(\sigma^2_m) \) lying in the interval \( x, x + dx \). This gives:

\[
f(m, x|\sigma^2_m) \, dm \, dx = \sqrt{\frac{2}{\pi}} \frac{2m - x}{(2\sigma^2_M)^{3/2}} \exp \left[ -\frac{(2m - x)^2}{2\sigma^2_M} \right] \, dm \, dx,
\]

\[
0 \leq m, x \leq m.
\]  

(7)

To obtain the joint density function for \( \delta_{\max}(\sigma^2_m) \) lying in the interval \( m, m + dm \) and \( \delta_{\min}(\sigma^2_m) \) lying in the interval \( y, y + dy \) we note that:

\[
P(\delta_{\max}(\sigma^2_m) < a, \delta_{\min}(\sigma^2_m) < b|\sigma^2_m) = \int_0^b \int_0^a f(m, m - y) \, dy \, dm.
\]  

(8)

Figure 1. The solid line represents a random walk consisting of 70 equally spaced steps of \( \sigma^2_M \) in the interval \( 0 < \sigma^2_M < 1.4 \). The random walk was chosen from a set of such walks and is such that \( \delta(\sigma^2_m) > 2m - x \). The dotted line represents the walk obtained by reflecting the initial trajectory about the line \( \delta = m \) after the first upcrossing of this line. Note that it has the property that \( \delta(\sigma^2_m) > 2m - x \). Similarly, any trajectory which is such that \( \delta(\sigma^2_m) > 2m - x \) can undergo a reflection to obtain a trajectory with \( \delta_{\max}(\sigma^2_M) > m \) and \( \delta(\sigma^2_m) < x \). There is a one-to-one correspondence between walks with \( \delta_{\max}(\sigma^2_M) > m \) and \( \delta(\sigma^2_m) < x \) and walks with \( \delta(\sigma^2_m) > 2m - x \).
related to the formation of a halo, thus provides mass and formation time distributions for the progenitors of the halo. As noted above, the form of the trajectory continuing from such a formation point is independent of the position of that point, and consequently the mass distribution of progenitors immediately prior to the formation event is independent of the formation epoch. Turning this argument around, we see that the effect of placing constraints on the progenitors of the halo immediately prior to the formation event does not affect the distribution of formation times. We thus have the important result that, even if we only consider formation events resulting from mergers with specified progenitors, such as haloes resulting from a merger of two approximately equally sized objects, the distribution of times at which these mergers occur is still that given by equation (13). It is for this reason that we consider the cosmic variation in halo formation rate also to be a measure of the cosmic variation in the rate of mergers between dark haloes. We should note that because standard PS theory does not account for halo sub-structure, the derivation presented here excludes possible mergers between low-mass halo sub-units forming part of a larger collapsed halo.

4 AN ALTERNATIVE DERIVATION

We will now show that $P(t|M) \, dt$ as given by equation (13) can be calculated from the conditional probability density function $P(\delta|\sigma_M) \, d\delta$ (Bond et al. 1991) using Bayes' theorem. Firstly, we wish to calculate the prior for $\delta_c$ which is equivalent to asking the question "Given no knowledge of $\sigma_M$, what is the distribution of upcrossing points in $\delta_c$?". We note that all trajectories must have an upcrossing point of any given line $\delta = \delta_c$. Now, in any two equally sized intervals, $d\delta_1$ and $d\delta_2$, there must be equal probability of such a crossing existing because the walk does not alter its form at different $\delta_c$ (all walks which pass through a given point can be thought of as new walks starting from that point). Thus, given no a priori information about $\sigma_M^2$, all values of $\delta_c$ are equally likely and we should assume a uniform prior for $\delta_c$. In order to be mathematically rigorous, we must make $\delta_c$ bounded. We see later that we can remove these bounds without affecting the result. So we have that:

$$P(\delta_c) \, d\delta_c = \begin{cases} \frac{d\delta_c}{\delta_{\text{max}} - \delta_{\text{min}}} & \delta_{\text{min}} \leq \delta_c \leq \delta_{\text{max}} \\ 0 & \text{otherwise} \end{cases}$$

(16)

Applying Bayes' theorem and using equation (1), we can find the joint probability of $\delta_c$ and $\sigma_M^2$:

$$P(\sigma_M^2, \delta_c) \, d\sigma_M^2 \, d\delta_c = \frac{1}{(\delta_{\text{max}} - \delta_{\text{min}})} \times \frac{d\delta_c}{2\pi^{1/2}(\sigma_M^2)^{1/2}} \exp\left(-\frac{\delta_c^2}{2\sigma_M^2}\right) \, d\sigma_M^2 \, d\delta_c.$$  

(17)

We can integrate this equation over all $\delta_c$ to obtain the probability density function for $\sigma_M^2$:

$$P(\sigma_M^2) \, d\sigma_M^2 = \frac{1}{(2\pi)^{1/2}(\sigma_M^2)^{1/2}(\delta_{\text{max}} - \delta_{\text{min}})} \times \exp\left(-\frac{\delta_c^2}{2\sigma_M^2}\right) \right] \, d\sigma_M^2.$$  

(18)

and again use Bayes' theorem to obtain the conditional probability

$$P(\delta_c|\sigma_M^2):$$

$$P(\delta_c|\sigma_M^2) \, d\delta_c = \frac{\delta_c}{\sigma_M^2} \exp\left(-\frac{\delta_c^2}{2\sigma_M^2}\right) \, d\delta_c \times \left[ \exp\left(-\frac{\delta_{\text{min}}^2}{2\sigma_M^2}\right) - \exp\left(-\frac{\delta_{\text{max}}^2}{2\sigma_M^2}\right) \right].$$  

(19)

In the limit as $\delta_{\text{min}} \to 0$ and $\delta_{\text{max}} \to \infty$ this becomes:

$$P(\delta_c|\sigma_M^2) \, d\delta_c = \frac{\delta_c}{\sigma_M^2} \exp\left(-\frac{\delta_c^2}{2\sigma_M^2}\right) \, d\delta_c.$$  

(20)

This is the probability that given a particular value of $\sigma_M^2$, the first upcrossing at this $\sigma_M^2$ has probability $P(\delta_c|\sigma_M^2)$ of being between $\delta_c$ and $\delta_c + d\delta_c$. This probability is identical to that calculated directly from the Brownian random walks in Section 3.

5 COMPARISON WITH RESULTS FROM MONTE CARLO SIMULATION

As a check on the validity of equation (14) we have performed Monte Carlo simulations to calculate random trajectories and determine the distribution of formation epochs. Approximately $10^6$ random walks were constructed, each consisting of $2^{11}$ uniform steps in $\sigma_M^2$ between 0 and $\sigma_M^2$ equivalent to a mass of $1.3 \times 10^{13} M_\odot$ for a standard CDM power spectrum with shape parameter $\Gamma = 0.25$ normalized to $\sigma_8 = 0.64$. Of these walks, $10^6$ were recorded with first upcrossing points at the final value of $\sigma_M^2$.

The distribution of these upcrossing points was recorded in uniform bins in time and is compared with the expected distribution given by equation (14) in Fig 2. Good agreement is demonstrated.

Figure 2. Comparison of Monte Carlo results with Press–Schechter predictions of the halo formation rate at fixed final mass. Open symbols show Monte Carlo results of the formation time of a halo of mass $1.3 \times 10^{13} M_\odot$ for a standard CDM power spectrum with shape parameter $\Gamma = 0.25$ normalized to $\sigma_8 = 0.64$. The solid curve shows the prediction of equation (14) at this mass. The results are normalized to give a peak formation rate of 1.
6 COMPARISON WITH RESULTS FROM N-BODY SIMULATION

We present results from a simulation run using the Hydra N-body, hydrodynamics code (Couchman, Thomas & Pearce 1995). The simulation used $128^3$ dark matter particles in a flat universe with $\Omega_M = 1$, $\Omega_{\Lambda} = 0$, $h = 0.5$, and standard CDM power spectrum with shape parameter $\Gamma = 0.25$. The power spectrum was normalized to $\sigma_8 = 0.64$. Groups of particles were found using a standard friends of friends (FOF) algorithm which links two particles together if the mean overdensity of a grid of particles with the same separation is greater than that predicted at the moment of virialization for spherical top-hat collapse (Peebles 1980).

In order to calculate the rate at which haloes form we analysed results from the N-body run at 104 different times, separated by equal intervals of time. By comparing the FOF results at each output time to results from earlier times we were able to build up a picture of the hierarchical growth of structure. Haloes with at least half of the component particles not observed in a halo of equal or higher mass at an earlier time were recorded as being formed in the time interval between this output time and the previous one.

There is a problem in this analysis that we may miss formation events – a halo formed in the time interval of interest may have already merged into a halo of higher mass when we analysed the simulation. However we can estimate the potential error caused by this effect by calculating the maximum number of haloes which could have been formed from the data on all new haloes. The distribution of new haloes of between 45 and 55 particles is shown in Fig. 3. The symbols mark the average of the minimum and maximum mass which could have been involved in mergers per unit time. The minimum is just the recorded mass in new haloes of between 45 and 55 particles, the maximum is the mass in all new haloes of greater than 45 particles which could have passed through the interval of interest. The error bars mark the positions of these maximum and minimum points. Note that these error bars only denote the error in measuring the mass formation rate from the N-body simulation results – they do not include errors intrinsic to the N-body simulation.

In order to count enough formation events we had to use a reasonably large interval in mass, corresponding to between 45 and 55 particles. However, the curve predicted from equation (14) is similar within this mass range, and the error caused by this effect will be small. In Fig. 3 we plot the prediction from equation (14) at a mass equivalent to 50 particles which shows remarkably good agreement with the results of the N-body simulation. Note that we have not renormalized any of the parameters of the PS models other than by the self-consistent method given in Section 2. One caveat we should note, however, is that we have only been able to test the agreement between the PS result and N-body simulation at relatively high halo masses.

7 COMPARISON WITH PREVIOUS RESULTS

We consider that the formation of a halo occurs when all of the mass in the halo is assembled. Blain & Longair (1993) and Sasaki (1994) also used the same definition of ‘formation’, and tried to calculate the rate of halo formation from the standard PS mass function (equation 3) and its derivative. However, they both made assumptions beyond standard PS theory. Blain & Longair (1993) calculated the formation rate empirically assuming a form for the distribution of mergers which occur. Sasaki (1994) obtained an equation for the formation rate assuming the destruction rate (from which they then calculate the formation rate) for a power-law spectrum has no characteristic mass scale. These assumptions lead to different forms for the formation rate from that derived above where we have not made any assumptions beyond those of standard PS theory.

Lacey & Cole (1993) defined the formation time as the time when the largest progenitor of a halo first contains at least half the mass of the halo. Using our definition of formation, we see that the distribution of such times is equivalent to the formation time distribution of progenitors of mass greater than $M/2$ given that they are the first such progenitors for a particular halo. The formation time distribution of progenitors of mass $M$ given by equation (15) is only dependent on knowing that a progenitor of mass $M$ is formed at some epoch. For instance by setting $M = M'/2$ in equation (15) we can calculate the distribution of times at which progenitors of mass $M'/2$ form, given that a progenitor of mass $M'/2$ is formed at some time. However, we do not know that the first subclump has mass $M'/2$, only that it lies in the range $M'/2 < M < M'$ – this is a consequence of PS theory containing mass jumps. Lacey & Cole (1993) give a counting argument (their section 2.5.2) which converts from a probability density in mass to a distribution in the number of progenitors in order to set the condition that we are only interested in the first halo to contain at least half the mass of the final halo. Without such an argument it is difficult to see how to distinguish the first progenitor from subsequent ones and so the results of this paper are not directly comparable to those of Lacey & Cole.

8 THE PREDICTED HALO FORMATION RATE

We can now estimate the rate at which mergers occur to create haloes of any particular mass for a number of cosmologies. We
investigate four different cosmologies, based on those used by the Virgo consortium (Springel et al. 1998; Jenkins et al. 1998) summarized in Table 1.

Fig. 4 shows how the functions vary with halo mass for the ΛCDM model. Curves are plotted at four different values of halo mass, $10^{10.5} - 10^{12} M_{\odot}$. Here it is easy to see the hierarchical build up of structure — low mass haloes predominantly form first, with their peak formation rate occurring at higher redshift than that for higher mass haloes.

For each cosmology chosen, for a particular mass of halo, the redshift at which the formation rate peaks is highly dependent on the $\sigma_8$ normalization chosen. High values of $\sigma_8$ lead to early formation and the peak in the formation rate is at higher redshift. However by choosing a higher halo mass we can recover the same shape of curve. This is easy to see in the SCDM and ΛCDM cosmologies, where equation (14) applies, and the curve shape is only dependent on the value of $\beta$ which can be kept constant by varying the halo mass with $\sigma_8$. Fig. 5 shows the mass merging rate for each of the cosmological models, but with halo masses chosen such that the models approximately agree on the redshift of the peak of halo formation. Viewed in this way, the OCDM model produces less evolution than the other models, but the other three have rather similar amounts of evolution.

| model | $\Omega_M$ | $\Omega_{\Lambda}$ | $\Gamma$ | $h$ | $\sigma_8$ |
|-------|------------|-------------------|---------|-----|-----------|
| OCDM  | 0.3        | 0                 | 0.15    | 0.5 | 0.85      |
| ΛCDM  | 0.3        | 0.7               | 0.15    | 0.5 | 0.91      |
| SCDM  | 1          | 0                 | 0.5     | 0.5 | 0.60      |
| ΛCDM  | 1          | 0.25              | 0.5     | 0.5 | 0.60      |

Table 1. Table of parameters used for each of the four cosmological models chosen. Here $h = H_0/100$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_M = (8\pi G/3H_0^2)\rho_{m0}$, $\Omega_{\Lambda} = (\Lambda/3H_0^2)$, and $\Gamma$ is the CDM power spectrum shape parameter.

Figure 4. The formation rate of galaxy haloes as calculated using equation (13) for the ΛCDM cosmological model. Curves are plotted for halo masses in the range $10^{10.5} - 10^{12} M_{\odot}$ and are normalized to give the same peak formation rate.

Figure 5. Comparison of the four different cosmological models. Because of the differing normalizations of the models the formation rates are plotted for different values of the halo mass, chosen to provide the closest match between the models. The masses are: OCDM, $10^{12.7} M_{\odot}$; ΛCDM, $10^{12.0} M_{\odot}$; SCDM, $10^{12.1} M_{\odot}$; ΛCDM, $10^{11.1} M_{\odot}$. The curves are normalized to the same peak formation rate.

9 THE MERGER-INDUCED STAR FORMATION RATE

A complete description of the formation of stars in hierarchically forming galaxies requires us to understand mergers between dark haloes, the fate of the baryonic matter in such mergers, and the processes of cooling, star formation and feedback and stellar evolution that all affect the observed luminosity of a galaxy and its time variation. Some recent attempts to explain specifically the strong evolution in star formation have used ad hoc models of the rate of generation of new galaxies as a source function, which have then been combined with models describing the remaining processes and tested against the observed data in sub-mm, infrared, optical and ultraviolet wavebands (Madau et al. 1996, 1998). An alternative method is to simulate the merger histories of galaxies based on the work of Lacey & Cole (1993, 1994), add heuristic recipes for the formation of stars and predict quantities such as the galaxy luminosity function or the cosmic evolution in star formation rate (Kauffmann, White & Guiderdoni 1993; Cole et al. 1994; Baugh et al. 1998). The success of these latter models has been the broad general agreement between the observations and the models, in that evolution in the predicted star formation rate is seen, with a maximum rate at redshifts about unity (Madau et al. 1996, 1998; Baugh et al. 1998). However, when considered in more detail, it is clear that those models do not reproduce the amount of evolution that is seen. That amount appears to be at least a factor of about 10 (Lilly et al. 1996; Connolly et al. 1997; Madau et al. 1996, 1998), and the sub-mm data suggest that...
factors of about 30 may be required (Hughes et al. 1998). Such large evolution in star formation rate would be consistent with the depletion of neutral gas that is inferred from the evolution in intervening absorption towards distant quasars (Pei & Fall 1995). The models cited above appear to predict evolution only by a factor about 5. Semi-analytic models incorporating an additional evolving component of star formation associated with short-lived starbursts have now been produced (Guiderioni et al. 1998) in which cosmological evolution by a factor \((1+z)^2\) is assumed, and these models do appear to give a better representation of the observed SFR.

Our aim in this section is to provide a complementary analysis to the above work by considering only the role of the cosmological evolution in halo formation rate. Observed luminous starbursts at low redshifts are probably rather young, and simulations indicate that bursts of star formation associated with mergers between either disc or bulge/halo systems with a variety of relative masses all have lifetimes which are short \((< 10^8 \text{ yr})\) (Mihos & Hernquist 1994, 1996). We have already argued that the rate of halo formation should have the same cosmological evolution as the rate of mergers between dark haloes, and we should therefore expect any star formation which is associated with mergers to have cosmological evolution consistent with the calculated evolution in halo formation rate. Similarity between the observed and calculated evolution would indicate that dark halo mergers are an important factor in star formation at high redshifts and we would then predict that most of the star formation in high-redshift galaxies should be short lived relative to the Hubble time.

The inferred star formation rates shown in Fig. 6 are derived primarily from the observed luminosity density at a rest-frame wavelength of 280 nm. Models of an evolving galaxy after a burst of star formation (Bruzual & Charlot 1993) show that the ultraviolet light decays approximately exponentially with a timescale of about 0.6 Gyr and it is this timescale which therefore dominates the light curve of short-lived starbursts. Thus, to compare the cosmic evolution in formation rate with the observed evolution we convolve the rate of halo formation as derived above with an exponential function with this decay time-scale. Note that this simple approach is not designed to provide a complete description of the physics of star formation in these galaxies: merely to test the relative importance of the cosmic variation in merger rate.

A sample of such curves are plotted with recent star formation rate data in Fig. 6: to be consistent with this previously-published data we adopt the \(\Omega_M = 1, \Omega_L = 0\) cosmology and we show a range of halo masses for the \(\Lambda CDM\) model. Since the fraction of halo mergers which result in an observable starburst is unknown, we treat the normalization of the models as being a free parameter and normalize the curves to the zero-redshift data point.

We see excellent agreement between the data and the cosmic variation in halo formation rate for any of the dark halo masses considered at redshifts up to about 1. At higher redshifts the data are very uncertain, and will remain so until corrections for optical extinction in starbursts are better understood, but we expect the relative contributions from a range of halo masses to be important in influencing the observed form of the evolution at high redshift. As an illustration, we combine formation rates for haloes of a range of final masses, weighting the contribution from each mass by a Gaussian distribution in \(\log M\). The relative normalization for each mass is determined at \(z = 0\) using equation (3). Fig. 7 shows the resulting evolution compared with the star formation data, for a Gaussian weighting function centred on \(10^{10.6} M_\odot\) with a logarithmic dispersion \(\sigma = 1.5\). Combining the curves in this way flattens the high redshift evolution as mergers creating lower mass haloes become increasingly important. As expected, the \(0 < z < 1\) evolution is the same as that shown in Fig. 6, but there is now much better agreement with the high-redshift data. A thorough treatment of this issue is beyond the scope of this paper, and would require additional modelling such as may be obtained from the semi-analytic approach. The illustration presented here is sufficient to show that the observed high-redshift behaviour can be reproduced by considering the star-forming galaxies to be dominated by merger-induced starbursts, and that it is the cosmic variation in halo formation rate that drives the evolution in star formation.

### 10 Quasar Density Evolution

The comoving space density of quasars has also decreased by a factor of order 100 from a redshift of 2 to the present day (Boyle et al. 1990). The quasar comoving space density appears to have a maximum at a redshift about 2–3 (Schmidt, Schneider & Gunn 1995; Shaver et al. 1996). Fig. 8 shows a compilation of measurements of the comoving space density of quasars selected in one decade of luminosity, shown as a function of redshift. The compilation shows quasars selected in radio, optical and X-ray wavebands: the consistency in their evolution is a strong argument that the form of the observed evolution is intrinsic to the quasar
Figure 7. The evolution predicted by combining the formation rate of different mass haloes (convolved with an expected lifetime of 0.6 Gyr) weighted by a Gaussian in $d/d\log M$ (solid line). The Gaussian is centred on a halo mass of $10^{10.6}M_\odot$ and has $\sigma = 1.5$. The evolution expected from halo formation events resulting in haloes only of mass $10^{10.6}M_\odot$ is also plotted (dotted line). Both curves are normalized to the data at $z = 0$. Data are plotted as in Fig. 6.

population and not due to either observational selection effects or to intervening obscuration (see also Shaver et al. 1996). The luminosity range that has been chosen is that in which the greatest amount of cosmological evolution is seen in the range $0 < z < 2$, and this therefore provides the most stringent test of the model we discuss here. Previous authors have chosen instead to plot the comoving quasar luminosity density integrated over a wide range in luminosity (e.g. Boyle & Terlevich 1998); this produces a lower amount of evolution, in fact more comparable to the inferred evolution in SFR. We adopt the more pessimistic approach here.

Our knowledge of the physics of quasar activation is even worse than our knowledge of the activation of intense bursts of star formation. We follow on from work by Elsathiuq & Rees (1988) and Haehnelt & Rees (1993) and assume that merger events activate quasars, although we are not concerned with the actual mechanism of the activation. As for the induced starbursts we wish to establish whether the cosmic evolution in halo formation rate can account for much of the evolution in quasar numbers.

We have little knowledge of the lifetime of quasars, but we can see that their lifetime must be short if the model is to reproduce the observed amount of evolution. The association of a few quasars with starbursts might be taken as evidence that their lifetimes are similar, and the ages of extended radio sources are also thought to be short ($< 10^8$ yr). We have therefore plotted the formation rate convolved with the same exponential lifetime as used to model starburst evolution with the quasar data in Fig. 8. The normalization has again been treated as a free parameter. Recent observations of luminous quasars suggest that they live in massive host galaxies with a narrow range of host luminosities (McLeod & Rieke 1994a,b; 1995; Taylor et al. 1996; Hooper, Impey & Foltz 1997) and by inference a narrow range of host masses. Here we compare the model at a single halo mass to the data. In this case the OCDM model does not produce enough evolution for any mass of halo and does not fit the data well. Reasonable fits can be obtained for the remaining three models, although all of them have a deficit in the amount of evolution between $z = 2$ and $0$ of about a factor 4. We have chosen to plot the ACDM model as, with its high $\sigma_8$, haloes of masses $10^{11.8}M_\odot$ peak at $z \sim 2.5$. The SCDM model also fits reasonably with a high value for halo mass, but can probably be excluded on other grounds (Gawiser & Silk 1998). The TCDM model fits with a value for the halo mass of $10^{10.6}M_\odot$: a value which is rather lower than would be implied by the observation that luminous quasars exist in massive host galaxies. The halo mass would be increased to $10^{11.8}M_\odot$ if $\sigma_8$ were increased to 0.9: such high power-spectrum normalizations are indicated by the COBE data (Gawiser & Silk 1998). Finally, we note that increasing $h$ from 0.5 to 0.7 decreases the inferred halo masses by a factor about 3.

Figure 8. The measured comoving space density of quasars is plotted for flat-spectrum radio-loud quasars (Shaver et al. 1996) with $P_{270.8} > 10^{27.0}WHz^{-1}sr^{-1}$ (stars), optically selected UVX quasars (Goldschmidt & Miller 1998; Boyle et al. 1990) with $M_{R < -25.4} > M_R > -27$ (open triangles), optically selected high-redshift quasars (Schneider, Schmidt & Gunn 1994) with $10^{48} < P_{270.8}W < 10^{49}$ (open circles), and a composite of ROSAT soft X-ray selected quasars (Boyle, Wilkes & Elvis 1998; Boyle et al. 1994; Schmidt et al. 1998; Read et al., in preparation) with $10^{37.7} < P_{0.5-2.4keV}W < 10^{38.7}$ (filled circles). The cosmological parameters of the ACDM model have been assumed: the shape of the evolution shows little variation between the three cosmologies parametrized in Table 1. As each waveband selects a different subset of the overall quasar population, and also covers a slightly different range of bolometric luminosity, the normalization of each of the samples has been adjusted such that they lie on a consistent curve passing through all the data and then normalized to the X-ray data. The curve plotted is the predictions of the ACDM model for a halo mass of $10^{11.8}M_\odot$ with normalization of the curve adjusted to match the data (see Section 10).
11 CONCLUSIONS

We have derived a simple formula for the rate at which structure merges to form new haloes within Press–Schechter theory. This formula has been tested and shown to be in good agreement with the results of numerical simulations.

We have also considered whether the strong cosmological evolution in star formation is primarily driven by the cosmic variation in the rate of halo formation. In order to produce a complete model of the evolution of merger-induced star formation we need to know much more physics than we have used in this paper. However, it is expected that the evolution in the rate at which dark matter haloes merge to form higher mass haloes will have a large influence on the evolution of the merger-induced SFR. Examining Fig. 6 we saw that, although it is expected that a range of halo masses are important in producing starbursts, for halo masses $10^{10}$-$10^{11.5} M_{\odot}$ the observed drop in the SFR from redshifts $z \sim 1$ to $z \sim 0$ is independent of mass and the evolution predicted from any linear combination of these curves shows strong evolution. The similarity of this evolution to that observed is remarkable and, given that quiescent star formation predicted from semi-analytic models does not provide enough evolution in this redshift range (Guiderdoni et al. 1998), indicates that merger-induced starbursts are extremely important for star formation at $z \sim 1$ and are perhaps the principal mechanism leading to the observed star formation at high redshifts. This model does not preclude the undoubted existence of quiescent star formation as well: it merely states that because the starburst component evolves so strongly it dominates at high redshift. At higher redshifts a more physically motivated model is needed to deduce the relative contributions of a range of halo masses, but we have shown that a simple combination of such a range can produce evolution which is in good agreement with the data of Steidel et al. (1999).

The evolution of the quasar population at optical absolute magnitudes $M_B \sim -26$ is larger than that of the star formation rate, although the evolution of the total quasar luminosity density is not (Boyle & Terlevich 1998). We have shown that, provided that quasar lifetimes are finite but short (around 0.6 Gyr), then mergers to form the haloes of massive galaxies can reproduce the principal features of the observed quasar evolution, consistent with the observation that luminous quasars consistently live in massive host galaxies. But whilst the curve predicts very well all the basic features of the observations, it fails by about a factor of $\sim 4$ to produce enough evolution between redshifts of 0 and 2. In this sense the model does not provide a full explanation of the observed evolution, but as the PS theory does not include any non-linear dynamical effects we consider even this agreement to be surprisingly good. Previous work has suggested that the expected cosmic variation in galaxy velocity dispersion (Carlberg 1990) or in halo circular velocity (Haehnelt & Rees 1993) could contribute to the quasar evolution, and only a modest amount of additional evolution would be needed. In the latter case we should expect to see some cosmological evolution in the host masses associated with a given quasar luminosity. Some additional component of quasar luminosity evolution (Boyle et al. 1990; Goldschmidt & Miller 1998) would also have the desired effect. Quasar lifetimes significantly longer than about 1 Gyr would smooth out the evolution in these models and would produce worse agreement with the data.

From the above comparisons between observations and the calculated rate of dark halo formation we conclude that hierarchical merging is highly important for the strong cosmological evolution that is observed in both star formation and quasar activity. Determination of the masses of haloes involved in star formation at high redshifts (e.g. Pettini et al. 1998) and quasars would provide good constraints on allowed cosmological models and their parameters, and estimation of the ages of the star-forming systems at high redshift would be a direct test of the proposition that starbursts account for much of the observed star formation.

ACKNOWLEDGMENTS

We are grateful for the use of the Hydra N-body code (Couchman, Thomas & Pearce 1995) kindly provided by the Hydra consortium. WJP acknowledges a PPARC studentship.

REFERENCES

Adelberger K., Steidel C., Giavalisco M., Dickinson M., Pettini M., Kellogg M., 1998, ApJ, 505, 18
Barnes J. E., Hernquist L., 1992, ARA&A, 30, 705
Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 1998, ApJ, 498, 504
Blain A. W., Longair M. S., 1993, MNRAS, 264, 509
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Boyle B. J., Fong R., Shanks T., Peterson B. A., 1990, MNRAS, 243, 1
Boyle B. J., Shanks T., Georgantopoulos I., Stewart G. C., Griffiths R. E., 1994, MNRAS, 271, 639
Boyle B. J., Wilkes B. J., Elvis M., 1998, MNRAS, 285, 511
Boyle B. J., Terlevich R. J., 1998, MNRAS, 293, L49
Brotherton M. S. et al., 1999, ApJ, 520, L87
Bruzual A. G., Charlot S., 1993, ApJ, 405, 538
Canalizo G., Stockton A., 1997, ApJ, 480, L5
Carlberg R. G., 1990, ApJ, 350, 505
Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781
Connolly A. J., Szalay A. S., Dickinson M., SubbaRao M. V., Brunner R. J., 1997, ApJ, 486, L11
Couchman H. M. P., Thomas P. A., Pearce F. R., 1995, ApJ, 452, 797
Dunlop J., 1997, in Bremer M., Jackson N., Perez-Fournon I., eds, Observational Cosmology with the New Radio Surveys. Kluwer, Dordrecht
Efstathiou G., Rees M. J., 1988, MNRAS, 230, 5p
Efstathiou G., Frenk C. S., White S. D. M., Davis M., 1988, MNRAS, 235, 715
Eke V. R., Cole S., Frenk C. S., 1996, MNRAS, 282, 263
Galloj G., Zamorano J., Aragon-Salamanca A., Rege M., 1995, ApJ, 455, L1
Gawiser E., Silk J., 1998, Sci, 280, 1405
Glazebrook K., Blake C., Economou F., Lilly S., Colless M., 1998, MNRAS, in press (astro-ph/9808276)
Goldschmidt P., Miller L., 1998, MNRAS, 293, 107
Gun J. E., Gott J. R., 1972, ApJ, 176, 1
Guiderdoni B., Hivon E., Bouchet F. R., Maffei B., 1998, MNRAS, 295, 877
Haehnelt M. G., Rees M. J., 1993, MNRAS, 263, 168
Hooper E. J., Impey C. D., Foltz C. B., 1997, ApJ, 480, L95
Hughes D. et al., 1998, Nat, 394, 241
Jedamiski K., 1995, ApJ, 448, I
Jenkins A. et al., 1998, ApJ, 499, 20
Karlin S., Taylor H. M., 1975, A first course in stochastic processes, 2nd ed. London Academic Press, London
Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lacey C., 1994, MNRAS, 271, 676
Lilly S. J., Le Fèvre O., Hammer F., Crampton D., 1996, ApJ, 460, L1
Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, MNRAS, 283, 1388
