

# Formation times and masses of dark matter haloes

Ravi K. Sheth<sup>1</sup> & Giuseppe Tormen<sup>2</sup>

<sup>1</sup> *Departement of Physics & Astronomy, University of Pittsburgh, 3941 O'Hara Street, PA 15260, USA*

<sup>2</sup> *Dipartimento di Astronomia, Vicolo dell'Osservatorio 2, 35122 Padova, Italy*

*Email: rks12@pitt.edu, tormen@pd.astro.it*

Submitted 6 January 2004; in original form 15 October 2003

## ABSTRACT

The most commonly used definition of halo formation is the time when a halo's most massive progenitor first contains at least half the final mass of its parent. Reasonably accurate formulae for the distribution of formation times of haloes of fixed mass have been available for some time. We use numerical simulations of hierarchical gravitational clustering to test the accuracy of formulae for the mass at formation. We also derive and test a formula for the joint distribution of formation masses and times. The structure of a halo is expected to be related to its accretion history. Our tests show that our formulae for formation masses and times are reasonably accurate, so we expect that they will aid future analytic studies of halo structure.

**Key words:** galaxies: clustering – cosmology: theory – dark matter.

## 1 INTRODUCTION

There is a simple analytic approximation for the distribution of halo formation times, when formation is defined as the time when the most massive progenitor first contains at least half the mass of the final object (Lacey & Cole 1993, 1994). (Throughout, we will use the word parent to denote the final object, and the word progenitor to denote the smaller pieces which made up the mass of the parent at some earlier time.) This formula provides a good description of what is seen in numerical simulations of gravitational clustering from Gaussian initial conditions, although recent work indicates that the agreement is not perfect (e.g., Wu 2001; Lin, Jing & Lin 2003). The sense of the discrepancy is that haloes in simulations appear to form slightly earlier than predicted, in qualitative agreement with previous work by Tormen (1998).

A related question is, what is the distribution of the mass of a halo at formation? Absent other information, natural assumptions about this distribution are (i) that it is a delta function centered at one-half, or (ii) that the formation mass is uniformly distributed between one-half and unity. The second assumption is motivated by the fact that halo formation is expected to be a stochastic process; haloes of the same mass may have had different formation histories. The main purpose of the present paper is to derive and test a formula for the joint distribution of formation times and masses. Section 2 studies the distribution of formation masses whatever the formation time. It shows that the distribution of masses just prior to, and just after formation, measured in simulations are both significantly different from delta functions, or from a uniform distribution, but are

rather similar to simple formulae for these quantities derived by Nusser & Sheth (1999). Section 3 studies the conditional distribution of the formation mass, when the formation time is known. This distribution is much better fit by a formula we derive here, than by a delta function or a uniform distribution. A final section summarizes our findings, and discusses possible applications.

## 2 THE DISTRIBUTION OF FORMATION MASSES

For what follows, it is useful to introduce some notation. We will use  $\delta_{\text{sc}}(z)$  to denote the value of the overdensity required for spherical collapse at  $z$ , extrapolated using linear theory to the present time (e.g. Peebles 1993), and  $\sigma^2(m)$  will denote the variance in the initial density fluctuation field when smoothed with a tophat filter of comoving scale  $R = (3m/4\pi\bar{\rho})^{1/3}$ , extrapolated using linear theory to the present time, where  $\bar{\rho}$  is the comoving background density. Thus, the shape of the initial power spectrum determines the relation between  $\sigma$  and  $m$ . At any  $z$ , there is a characteristic mass scale defined by  $\sigma^2(m) = \delta_{\text{sc}}(z)$ . We will use  $M_*(z)$  to denote this mass scale, and will often express masses in units of this characteristic mass.

Later in this section we will compare our results with simulations; these were kindly made available to the public by the Virgo consortium (Frenk et al. 2000). We will analyse results from the set of runs known as the GIF simulations. In particular, we will show results from the SCDM and  $\Lambda$ CDM models, for which  $\Lambda = 1 - \Omega$  and  $(\Omega, \sigma_8) = (1, 0.6)$  and  $(0.3, 0.9)$  respectively. Particle positions and velocities from

arXiv:astro-ph/0402055v1 2 Feb 2004

both simulations were output at a range of redshifts, approximately evenly spaced in logarithmic expansion factor:  $\Delta \ln(1+z) \approx 0.0596$ . For each output time, we identified haloes using the spherical overdensity method (e.g. Lacey & Cole 1994; Tormen, Moscardini & Yoshida 2003) which contained at least twenty particles. The required overdensity is a cosmology dependent factor times the background density, as specified by the spherical collapse model. For the SCDM model, this factor is 178, and it is independent of redshift; for the  $\Lambda$ CDM model, it is 323 at  $z=0$ , and is smaller at higher redshifts (e.g. Peebles 1993). At any given output time  $z_1$ , we selected the halos which were composed of more than two hundred particles, and studied the formation times and masses at formation of these haloes as follows. (For reference, an  $M_*$  halo at  $z=0, 0.5$  and  $1.0$  has 1289, 170 and 31 particles in the SCDM run, and 807, 185 and 40 particles in the  $\Lambda$ CDM run, so the high redshift runs mainly probe the formation times and masses of objects much larger than  $M_*$ .)

Given a halo of mass  $M_1$  (i.e., containing  $N_1$  particles) at  $z_1$ , we go to the previous output time ( $z_1 + dz_2$ , say), identify the object which contributes the most number of particles to  $N_1$ , and call it the most massive progenitor at  $z_1 + dz_2$ . Suppose this most massive progenitor had  $N_2$  particles. We then go to the preceding output step ( $z_1 + dz_2 + dz_3$ , say) and identify the most massive progenitor,  $N_3$ , of  $N_2$ . We continue in this way until the number of particles in the most massive progenitor first falls below  $N_1/2$ . If the mass just before formation is  $N_n$ , then the mass just after formation is  $N_{n-1}$ , and the redshift of formation is  $z_1 + \dots + dz_{n-1}$ . We store these values for each halo  $M_1$  at  $z_1$ .

The main quantity of interest in what follows is  $p(m, z_f)$ , the joint distribution of formation masses and times, where formation is defined to be the time when one of the subclumps of a halo first accounts for at least half the final mass  $M_1$  of its parent. Because of this definition of formation,  $m/M_1$  is distributed between one-half and unity (recall that the mass of the most massive progenitor must exceed half the mass of its parent).

The formation time distribution of haloes which have final mass  $M_1$  at redshift  $z_1$ ,

$$p(z_f) dz_f = \int p(m, z_f) dm, \quad (1)$$

is expected to be well approximated by

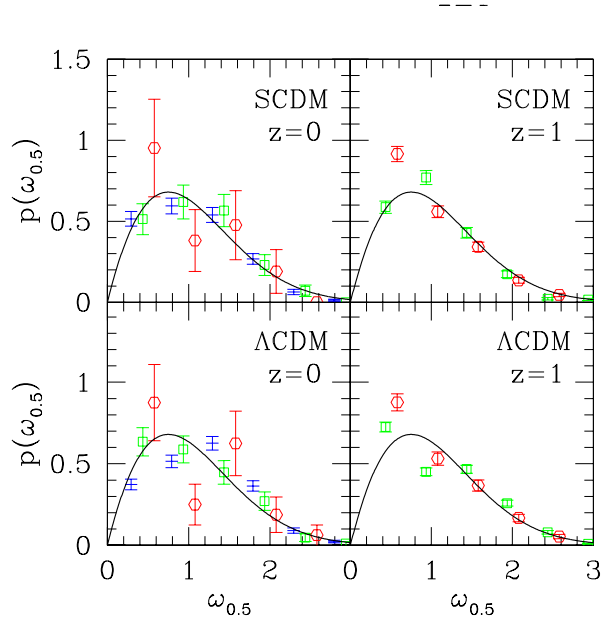
$$p(z_f) dz_f = p(\omega) d\omega = 2\omega \operatorname{erfc}\left(\frac{\omega}{\sqrt{2}}\right) d\omega \quad (2)$$

(Lacey & Cole 1993), where  $\omega^2 \equiv (\delta_{cf} - \delta_{c1})^2 / (S_f - S_1)$ ,  $\delta_{cf} = \delta_{sc}(z_f)$ ,  $\delta_{c1} = \delta_{sc}(z_1)$ , and  $S_f = \sigma^2(M_1/2)$ . As Lacey & Cole note, this formula is only well-behaved for white-noise initial conditions (for which  $\sigma^2(m) \propto 1/m$ ), although it provides a reasonable approximation in the more general case.

The distribution of formation masses, obtained by marginalizing over the formation time distribution, is

$$p(m) dm = \int p(m, z_f) dz_f. \quad (3)$$

Nusser & Sheth (1999) describe a model for the evolution of the mass of the most massive progenitor which is able to reproduce the formation time distribution of equation (2).



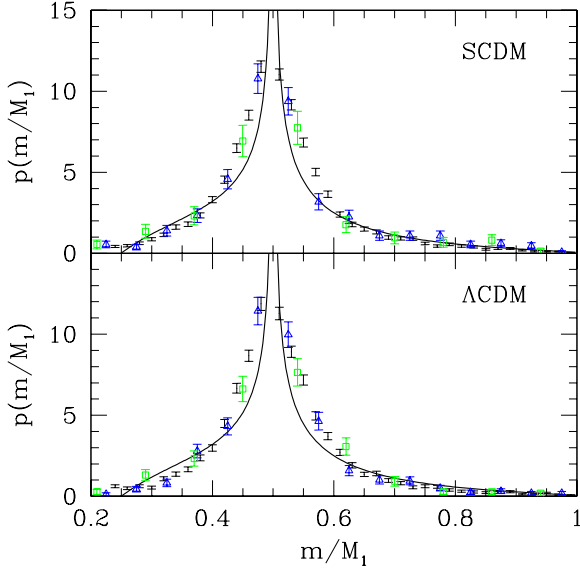
**Figure 1.** Distribution of scaled formation times in two different cosmological models, for haloes identified at two different redshifts. In these scaled units, the formation time distribution is expected to be independent of halo mass and final time. Solid curve shows the precise form which this universal formation time distribution is expected to have (equation 2). In all panels, squares and hexagons show the simulation results for parent haloes with masses in the range  $4 \leq M_1/M_*(z_1) < 8$  and  $16 \leq M_1/M_*(z_1) < 32$ . Simple bars in the panels on the left show results for slightly lower halo masses:  $M_1/M_*(z_1) \leq 2$ . Error bars were estimated assuming Poisson counts. Evidently, equation (2) provides a reasonable, but not perfect description of halo formation times in the simulations.

In their model, it is possible to derive an expression for the associated formation mass distribution of equation (3). In particular, they argue that

$$p(\mu) d\mu = \frac{2}{\pi} \sqrt{\frac{1-\mu}{2\mu-1}} \frac{d\mu}{\mu^2}, \quad \text{where } 1/2 \leq \mu \leq 1, \quad (4)$$

and  $\mu \equiv m/M_1$  (equation A15 in Nusser & Sheth 1999). Strictly speaking this formula, like equation (2), is valid for white-noise initial conditions, but Nusser & Sheth argued that it should provide a good approximation even if the initial spectrum has more large-scale power (see their Fig. A2 and associated discussion).

Figure 1 compares the formation time formula, equation (2), with measurements in the GIF simulations. (We have used the notation  $\omega_{0.5}$  to emphasize that formation is when the largest progenitor subclump contains at least half the mass of the final parent halo. Our requirement that parent haloes have at least two hundred particles means that we only probe the formation statistics of the most massive haloes at high redshift.) Although lower mass haloes identified at a given time tend to have formed at higher redshifts than more massive haloes (cf. Figure 4 below), equation (2) suggests that, when appropriately rescaled, all dependence on mass, time and the shape of the power spectrum should be removed. The different panels in the figure show that the scaled formation time distributions in the SCDM and



**Figure 2.** The distribution of masses  $m$  at formation, for parent haloes which have mass  $M_1$  at  $z_1 = 0$ . Symbols show the simulation results for  $M_1/M_*(z_1) \leq 1$  (dots),  $2 \leq M_1/M_*(z_1) < 4$  (triangles), and  $M_1/M_*(z_1) \geq 8$  (squares). Error bars were estimated assuming Poisson counts. Curves on the right and the left of  $m/M_1 = 1/2$  show the distributions in equations (4) and (5) respectively. There is no obvious trend with  $M_1$ , although haloes in simulations appear to have  $m/M_1 \approx 1/2$  slightly more frequently than the model predicts. Results for formation masses of parent haloes identified at other redshifts are similar.

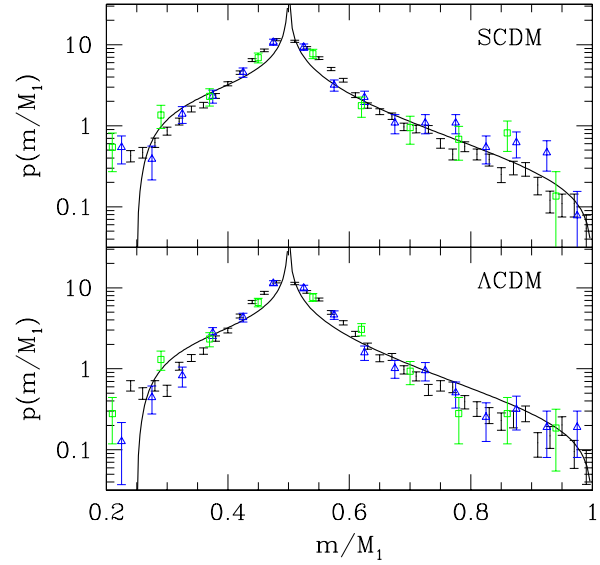
$\Lambda$ CDM runs are reasonably, but not perfectly-well described by equation (2).

Our next task is to test the accuracy of the formation mass formula, equation (4). The symbols in Figure 2 show the distribution of masses  $m$  at formation for haloes in the GIF simulations which have final mass  $M_1$  at  $z_1 = 0$ ; dots show  $M_1/M_*(z_1) < 1$ , open triangles show  $2 \leq M_1/M_*(z_1) < 4$ , and squares show  $M_1/M_*(z_1) > 8$ . Error bars were estimated assuming Poisson counts. The figure shows no clear trend with  $M_1$ . A similar analysis of the formation masses, using haloes identified at  $z = 0.5$  and  $z = 1$ , yields similar results. The solid curves which span the range  $1/2 \leq m/M_1 \leq 1$  in the two panels of Fig. 2 show equation (4). Although the formation mass distribution measured in the simulations is significantly different from either a delta function, or a uniform distribution, equation (4) is able to provide a reasonable description of its shape.

We also studied the mass of the most massive progenitor just before the formation time; these are shown by the symbols which span the range  $1/4 \leq m/M_1 \leq 1/2$  in the two panels. Once again, the measured distribution is neither a delta function nor is it uniform. In this case, also, there is a simple formula for the distribution of formation masses:

$$q(\mu) d\mu = \frac{d\mu/\mu^2}{\pi(1-\mu)} \left( \sqrt{\frac{\mu}{1-2\mu}} - \sqrt{1-2\mu} \right), \quad (5)$$

where  $1/4 \leq \mu \leq 1/2$ , and  $\mu \equiv m/M_1$  as before (equation A19 in Nusser & Sheth 1999). The curves which span the range  $1/4 \leq m/M_1 \leq 1/2$  in the two panels of Fig. 2



**Figure 3.** Same as previous plot, but now shown logarithmically, to emphasize the discrepancy near the peak and in the tails.

show this formula; it provides a reasonable description of the measurements in the simulations.

Although the analytic formulae provide a reasonable description of the measurements, haloes in the simulations appear to have slightly more occurrences of  $m/M_1 \sim 0.45$ , and  $m/M_1 \sim 0.55$  than the formulae predict. Some of the discrepancy may arise because the simulation outputs are not spaced arbitrarily closely in time (the typical redshift steps are of order  $\Delta z \sim 0.1$ ). As a result, the measured distributions almost certainly smooth-out the divergence around  $\mu \sim 1/2$ . (To better illustrate the behaviour around the peak, Figure 3 shows the same distributions, but this time on a logarithmic scale.) In principle, the analysis in Nusser & Sheth (1999) can be used to estimate this smearing-out (their equations A14 and A18 actually depend on the redshift difference), but we believe it will be better to use simulations with better time resolution instead, as these will be available shortly.

Some of the discrepancy may be associated with the fact that the approach leads to an underestimate of the mean formation redshift. This discrepancy could plausibly affect the formation mass distribution, since, if formation happens at higher redshift when the basic building blocks are smaller, then the formation masses are less likely have values as large as  $m/M_1 \sim 1$ . Moreover, equations (2)–(5) are derived from an approach which predicts fewer massive parent haloes than are actually observed in simulations (e.g. Sheth & Tormen 1999). If the abundance of parent haloes is modified so that it is in better agreement with simulations, then the formation mass and time distributions will also be modified (for reasons made explicit in Sheth & Tormen 2002). Accounting for this is left for future work, since the agreement between the model and the simulations is quite good.

### 3 CONDITIONAL DISTRIBUTION OF FORMATION MASS AND TIME

The joint distribution of formation mass and time for parent haloes with mass  $M_1$  at  $z_1$  is

$$p(m, z) dm dz = ds \int_{S_f}^{s_m} dS p(S, z + \Delta z | s, z) p(s, z | S_1, z_1), \quad (6)$$

where  $s \equiv \sigma^2(m)$ ,  $S_1 \equiv \sigma^2(M_1)$ ,  $S_f \equiv \sigma^2(M_1/2)$ ,  $s_m \equiv \sigma^2(m/2)$ ,

$$p(s, z | S_0, z_0) ds = \frac{d\nu}{\nu} \sqrt{\frac{\nu}{2\pi}} \exp(-\nu/2), \quad (7)$$

with  $\nu \equiv [\delta_{sc}(z) - \delta_{sc}(z_1)]^2 / (s - S_1)$ , and a similar expression holds for  $p(S, z + \Delta z | s, z)$ . When inserted in equation (1), equation (6) yields equation (2), and when inserted in equation (3) it yields equation (4).

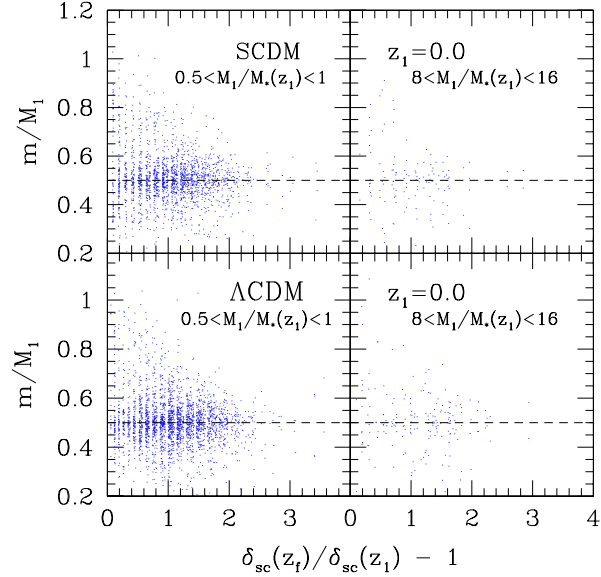
In the limit of small time steps ( $\Delta z \ll 1$ ), and a white-noise power spectrum, equation (6) simplifies considerably. A little algebra shows that, for haloes of fixed mass  $M_1$ , the conditional distribution of formation masses  $m$  when it is known that the formation time was  $z_f$  is given by

$$p(\mu | z_f) d\mu \equiv \frac{p(\mu, z_f) d\mu}{p(z_f)} = \frac{p(\mu) d\mu}{s/S_1 - 1} \frac{\exp\left[-\frac{\omega^2}{2} \frac{(S_f - S_1)}{(s - S_1)}\right]}{2 \operatorname{erfc}(\omega/\sqrt{2})}, \quad (8)$$

where  $\mu = m/M_1$ ,  $s \equiv \sigma^2(m)$ ,  $S_1 \equiv \sigma^2(M_1)$ ,  $S_f \equiv \sigma^2(M_1/2)$ , and  $\omega$  was defined in equation (2). For a white-noise spectrum,  $s/S_1 = 1/\mu$  and it is straightforward to verify that this distribution is correctly normalized. For more general power spectra,  $s/S_1 \sim \mu^{-\alpha}$ , say, this conditional distribution must be multiplied by a normalization factor which depends on  $\alpha$ . We have checked that use of the white-noise expression is a good approximation to the curves associated with  $\alpha < 1$ , provided we insure that the distribution is correctly normalized to unity. Thus, although equation (8) only holds for a white-noise power spectrum, we expect it to be more generally applicable for the same reasons that our equations (4) and (5) are more generally applicable. In what follows, therefore, we will simply set  $s/S_1 = 1/\mu$  and  $S_f/S_1 = 2$ . In this approximation, our expression for the conditional distribution of formation masses is independent of power spectrum.

The factor which multiplies  $p(\mu)$  is largest at  $s/S_1 - 1 = \omega^2$ , so objects which form at redshifts which are lower than the mean value for that mass (i.e.,  $\omega < 1$ ), are expected to have formation masses which are biased towards  $\mu \approx 1$  (i.e.,  $s \approx S_1$ ). Conversely, objects which form at abnormally high redshifts ( $\omega > 1$ ) are expected to have formation masses which are closer to the minimum value allowed:  $\mu \approx 1/2$ . Presumably, this is a consequence of the fact that, to have  $\mu \approx 1$  requires two pieces each of size  $\mu \approx 1/2$ . In a hierarchical model, the building blocks available to form the parent halo are, on average, smaller at early times: when the probability of having an object of mass  $\mu \approx 1/2$  is small, the chance of having two such objects is smaller still. In effect, our formula (8) quantifies the importance of this effect.

Figure 4 shows the joint distribution of formation mass and time for parent haloes identified at  $z_1 = 0$  in the SCDM (top) and  $\Lambda$ CDM (bottom) simulations. (The stripes are a result of the fact that simulation outputs are written to file only at finitely many time-steps.) The axis labels use the



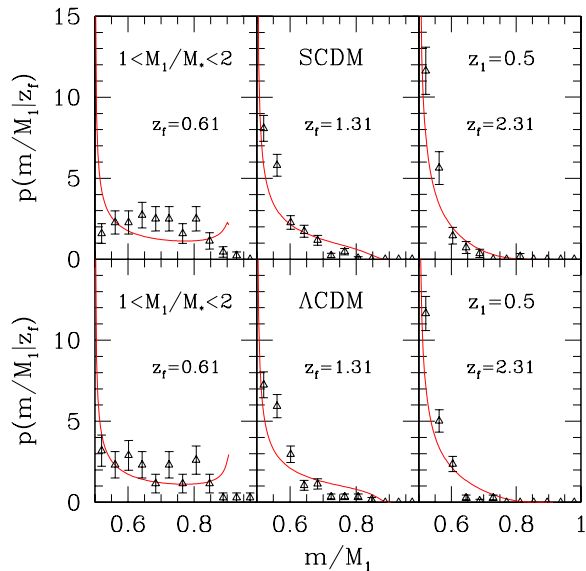
**Figure 4.** Joint distribution of formation times and masses measured in the simulations. Haloes which form at higher redshifts appear to have a smaller spread in formation masses. This is quantified in the next figure, which includes a comparison with the model predictions.

notation  $m/M_1$  to denote the ratio of the formation mass to final mass,  $z_f$  the formation redshift, and  $z_1$  the redshift at which the parent object was identified. The two panels for each simulation show results for different choices of the parent halo mass. Analogous plots for  $z_1 = 0.5$  and  $z_1 = 1.0$  look very similar, provided we scale the formation time axis to  $\delta_{sc}(z_f)/\delta_{sc}(z_1) - 1$  as we have done, rather than simply show  $z_f$ . We have chosen to not include them here. (The natural rescaling would have been to show  $\omega$ , defined in equation 2, along the x-axis. This would differ from the rescaling we show by a factor of  $(S_f - S_1)/\delta_{sc}(z_1)$ . We chose not to scale by this additional factor because one of the points we wish to emphasize is that the formation mass formulae turn out to be approximately independent of power spectrum.)

The formation time distribution discussed in the previous section is obtained by summing up all haloes with the same  $z_f$  whatever their value of  $m/M_1 > 1/2$ . The formation mass distributions studied in the previous section were obtained by summing up all haloes with the same  $m/M_1$  whatever their value of  $z_f$ .

Notice that there appears to be a tendency for the objects with large  $z_f$  to have small values of  $m/M_1$ , but because there are many fewer haloes with high formation redshifts, it is not obvious if this trend is real, or if it is simply a consequence of small-number statistics.

To address this in more detail, Figure 5 shows  $p(\mu | z_f)$ , the conditional distribution of formation masses at fixed formation time. The plot was made by choosing all haloes with masses in the range  $1 < M_1/M_*(z_1) < 2$  at  $z_1 = 0.5$ , and then studying the mass at formation in the subset which formed at  $z_f = 0.61, 1.31$  and  $2.31$ . Histograms show the measurements in the simulations. Comparison of the different panels shows that the objects which form at higher red-



**Figure 5.** Conditional distribution of masses  $m$  at formation, given that the mass of the parent halo was in the range  $1 < M_1/M_*(z_1) < 2$  at  $z_1 = 0.5$ , for a range of choices of the redshift of formation (labeled in the middle of each panel). Symbols show the measurements in the simulations, and curves show equation (8).

shifts have formation masses which are close to  $1/2$ , whereas there is an obvious tail of higher formation masses at lower formation redshifts. The smooth curves show equation (8); it reproduces the trend with formation redshift seen in the simulations quite well. We find similar agreement for other choices of  $M_1$ ,  $z_1$  and  $z_f$ , so we conclude that equation (8) provides a reasonable description (by which we mean it is a better fit than is a delta function, or a uniform distribution) of the conditional distribution of formation mass when the formation time is known.

#### 4 DISCUSSION

We presented evidence that formulae for the distribution of formation masses (equations 4 and 5), were reasonably accurate (Figure 2). These formulae do not depend on the shape of the underlying power spectrum, so they are simple to use. We then derived an expression for the conditional distribution of formation masses if the formation time is known (equation 8), and showed that it was also in quite good agreement with measurements made in simulations (Fig. 5). Application of Bayes' rule then gives the joint distribution of formation mass and time.

Our results indicate that haloes which form at abnormally early times are more likely to have formation masses of order one-half that of the final mass of the parent, whereas haloes which form at abnormally late times are more likely to have formation masses which are closer to that of the parent. One consequence of this is that haloes which form late are more likely to have experienced a recent major merger.

We argued that this was a generic consequence of hierarchical formation.

Our formulae for the joint distribution of formation masses and times will find use in studies which attempt to relate the structure of a halo to its formation history (e.g. Tormen 1997, 1998; Tormen, Diaferio & Syer 1998; van den Bosch 2002; Wechsler et al. 2002; Zhao et al. 2003). For instance, haloes which formed recently with large formation masses are almost certainly further from equilibrium than haloes which formed at higher redshift with formation masses of order fifty-percent. Such haloes (i.e. ones which have suffered major-mergers recently) may plausibly be less centrally concentrated than haloes of the same mass which had more quiescent accretion histories. Addressing such issues is the subject of on-going work. If these formulae do prove to be useful, it will become necessary to modify them slightly so that they are more fully consistent with the parent halo mass function described by Sheth & Tormen (1999).

We would like to thank the Aspen Center for Physics for support, and for providing the stimulating environment in which this work was completed. We would also like to thank the Virgo consortium for making the simulation data used here publically available at <http://www.mpa-garching.mpg.de/Virgo>. RKS was supported by the DOE and NASA grant NAG 5-7092 at Fermilab when work on this project began, and acknowledges support from NSF grant AST-0307747.

#### REFERENCES

- Frenk C. S., Colberg J. M., Couchman H. M. P., et al., 2000, *astro-ph/0007362*  
 Lacey C., Cole S., 1993, *MNRAS*, 262, 627  
 Lacey C., Cole S., 1994, *MNRAS*, 271, 676  
 Lin W. P., Jing Y. P., Lin L., 2003, *MNRAS*, 344, 1327  
 Nusser A., Sheth R. K., 1999, *MNRAS*, 303, 685  
 Peebles J. E. P., 1993, *Principles of Physical Cosmology*, Princeton Univ. Press, Princeton, NJ  
 Sheth R. K., Tormen G., 1999, *MNRAS*, 308, 119  
 Sheth R. K., Tormen G., 2002, *MNRAS*, 329, 61  
 Tormen G., 1997, *MNRAS*, 290, 411  
 Tormen G., 1998, *MNRAS*, 297, 648  
 Tormen G., Diaferio A., Syer D., 1998, *MNRAS*, 299, 728  
 Tormen G., Moscardini L., Yoshida N., 2004, *MNRAS*, in press (*astro-ph/0304375*)  
 van den Bosch F. C., *MNRAS*, 331, 98  
 Wechsler R., Bullock J. S., Primack J. R., Kravtsov A. V., Dekel A., 2002, *ApJ*, 568, 52  
 Wu J. P., 2001, *MNRAS*, 327, 629  
 Zhao D. H., Jing Y. P., Mo H. J., Börner G., 2003, *ApJ*, 597, L9