WHAT DO DARK MATTER HALO PROPERTIES TELL US ABOUT THEIR MASS ASSEMBLY HISTORIES?

ANSON W. C. WONG AND JAMES E. TAYLOR
Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada; a57wong@uwaterloo.ca, taylor@uwaterloo.ca
Received 2011 December 18; accepted 2012 August 1; published 2012 September 6

ABSTRACT

Individual dark matter halos in cosmological simulations vary widely in their detailed structural properties, properties such as concentration, shape, spin, and degree of internal relaxation. Recent non-parametric (principal component) analyses suggest that a few principal components explain a large fraction of the scatter in these structural properties. The main principal component is closely aligned with concentration, which in turn is known to be related to the mass accretion history (MAH) of the halo, as described by its merger tree. Here, we examine more generally the connection between the MAH and structural parameters. The space of mass accretion histories has principal components of its own. The strongest, accounting for almost 60% of the scatter between individual histories, can be interpreted as the age of the system. We give an analytic fit for this first component, which provides a rigorous way of defining the dynamical age of a halo. The second strongest component, representing acceleration or deceleration of growth at late times, accounts for 25% of the scatter. Relating structural parameters to formation history, we find that concentration correlates strongly with the early history of the halo, while shape and degree of relaxation or dynamical equilibrium correlate with the later history. We examine the inferences about formation history that can be drawn by splitting halos into sub-samples based on observable properties such as concentration and shape. Applications include the definition young and old samples of galaxy clusters in a quantitative way, or empirical tests of environmental processing rates in clusters.

Key words: dark matter – galaxies: clusters: general – galaxies: groups: general – large-scale structure of universe

Online-only material: color figures

1. INTRODUCTION

Dark matter halos provide the framework for visible structure in the universe over a span of eight decades in mass, from the scale of rich galaxy clusters down to the scale of individual dwarf galaxies. Mass is normally assumed to be the main determinant of a halo’s baryonic contents, and analytic models such as the halo occupation distribution (Peacock & Smith 2000; Seljak 2000; Ma & Fry 2000; White et al. 2001; Berlind & Weinberg 2002; Cooray & Sheth 2002), the conditional luminosity function (van den Bosch et al. 2003; Yang et al. 2003, 2005; Tinker et al. 2005; van den Bosch et al. 2007), and abundance matching (Vale & Ostriker 2004, 2006; Moster et al. 2010; Guo et al. 2010; Behroozi et al. 2010) make this assumption explicitly. Yet the halos that form in cosmological simulations vary greatly in shape, concentration, spin, substructure, and other structural properties (e.g., Vitvitska et al. 2002; Kasun & Evrard 2005; Allgood et al. 2006; Bett et al. 2007; Ragone-Figueroa et al. 2010; Knebe et al. 2011; Wang et al. 2011; Giocoli et al. 2012). As simulations of increasing size and resolution provide a more and more detailed picture of halo properties, and as observational techniques including weak and/or strong gravitational lensing, X-ray, and Sunyaev–Zel’dovich (SZ) measurements (e.g., Corless et al. 2009; Oguri et al. 2010; Sereno & Zitrin 2012; Morandi et al. 2011) reach a precision where they can measure structural properties reliably, it is important to understand how the structural features of a halo are interrelated, and what they can tell us about its formation and evolution.

Recently, two groups, Skibba & Macciò (2011) and Jeeson-Daniel et al. (2011; S11 and J11 hereafter, respectively), have taken the important step of performing non-parametric principal component analyses of halo properties. Non-parametric analysis allows one to learn about data without assuming a particular model fit. Principal component analysis (PCA) is a particular form of non-parametric analysis that searches for simplifying trends in a complex data set by finding the axes in the multi-dimensional data space that account for the largest fraction of the scatter. In the simplest case, it can uncover linear correlations in the data (e.g., fundamental lines or planes) and reveal hidden patterns or simplifications.

The PCA results of S11 and J11 agree on some basic aspects of halo structural properties. Overall, the scatter in halo properties spans a fairly high-dimensional space, with four principal components required to explain 70%–80% of the scatter. Nonetheless, a few clear patterns emerge. In terms of structural parameters, the first (strongest) principal component is best traced by concentration. The strong correlation with concentration suggests that this first principal component is linked to the overall age of the halo, and the strength of the correlation with $z_{0.5}$, the redshift by which the halo had built up half its final mass, confirms this. The origin of halo concentration has been studied extensively since it is a crucially important factor in many calculations, including strong lensing (e.g., Broadhurst et al. 2005, 2008; Giocoli et al. 2011) and dark matter annihilation (e.g., Taylor & Silk 2003; Pieri et al. 2008). Several analytic models have been developed over the years to explain concentration in terms of formation history (e.g., Bullock et al. 2001; Wechsler et al. 2002; Zhao et al. 2009), so the link between formation history and this particular property is fairly well understood. Older systems, that is, systems which already assembled most of their mass into one or a few progenitors at early times, are more concentrated on average, although the exact connection between age and concentration varies from one analytic model to another.

If the first principal component of halo structural properties is thus linked to age, we might naturally ask what the others
correspond to. Are shape, spin, or degree of relaxation also related to formation history, and if so, how? To put this question in a quantitative framework, we first have to decide how to describe the “formation history” itself—what should we take this to mean exactly, given the complex set of merger and accretion events through which halos form? We can take the mass accretion history (MAH) as a starting point, defining this as the function \( M(z) \equiv M(z)/M(0) \) that describes the mass of the main progenitor of a halo as a function of redshift, normalized to the value at \( z = 0 \) (van den Bosch 2002). Since \( M(z) \) is a continuous function of a real variable, it contains an arbitrarily large amount of information about the history of a halo; equivalently, describing the MAH fully means specifying values of \( M(z) \) at an infinite set of redshifts. To characterize the MAH in simpler terms, we can turn once again to PCA, approximating each individual MAH as a vector of values \( M(z_i) \) corresponding to the MAH evaluated at a finite, fixed set of redshifts \( z_i \). PCA of these vectors can then tell us whether mass accretion histories are well described by a set of basis functions characterized by a single parameter, as suggested by van den Bosch (2002) and Wechsler et al. (2002), or whether they require two or more variables to explain their diversity, as suggested by Tasitsiomi et al. (2004) or McBride et al. (2009).

Assuming that a few principal components capture the main features of a halo’s MAH, this will allow us to study correlations between structure and history in a well defined and quantitative way. We note that this is only a first step toward understanding structure in terms of growth history; the MAH does not contain all the information about a halo’s past by any means, and an alternate approach is to study the origin of a particular property such as shape or degree of relaxation in detail, considering the full three-dimensional geometry and dynamics of the physical processes involved (e.g., Vera-Ciro et al. 2011; Power et al. 2012).

In this paper, we generate halos in a set of cold dark matter (CDM) simulations covering three different mass scales. We record the MAH of each well-resolved halo and decompose the space of MAHs, taken as vectors of values \( M(z_i) \) into principal components. These components provide a non-parametric description of the MAH and clarify its basic properties. We also record the structural properties of each halo at the present-day. We study the correlations between these properties themselves (reproducing the trends found by S11 and J11), and between structural properties and the principal components of the merger history. Finally, we discuss an application of our results, showing in particular how samples of halos selected by concentration or shape will have systematically different formation histories. In this way, the observable properties of groups or clusters of galaxies may be used to infer their (unobservable) formation history.

The outline of the paper is as follows. Section 2 describes our N-body simulations and group finding, and explains how the structural properties of halos are defined and measured. In Section 3, we analyze the vector space of MAHs and decompose them into principal components. In Section 4, we analyze the correlations between different structural parameters and derive the principal components of this second vector space. We then explore the connection between structural parameters and MAH parameters. In Section 5, we discuss applications of this work to individual objects or samples of objects. We summarize our results in Section 6.

Throughout the paper, we consider a WMAP7 cosmology with parameters \( \Omega_M = 0.27, \Omega_\Lambda = 0.73, \Omega_b = 0 \) (in our

| Name | \( m_p \) | \( L_{\text{box}}(z = 0) \) | \( \epsilon_{\text{min}} \) | \( \epsilon_{\text{com}} \) | \( \epsilon_{\text{max}} \) |
|------|----------|----------------|----------|---------|---------|
| Sim60 | 0.01206 | 60 | 81.1 | 50 | 5 |
| Sim120 | 0.09648 | 120 | 67.3 | 110 | 11 |
| Sim240 | 0.77184 | 240 | 55.2 | 240 | 23 |

CDM-only simulations), \( H_0 = 72 \text{ km s}^{-1} \text{ Mpc}^{-1} \), \( n_s = 0.96 \), and \( \sigma_8 = 0.80 \).

## 2. MEASURING HALO PROPERTIES

### 2.1. Numerical Simulations

Our simulations were performed using the massively parallel N-body code Gadget2 (Springel 2005), with cosmological initial conditions generated by initial condition code Gadget2 (Bertschinger 2001). Three different boxesizes were simulated with \( N_p = 512^3 \) particles of mass \( m_p \) each, producing halo catalogs that vary in mass by a factor of eight from one simulation to the next.

The softening length used in the simulations was \( \epsilon = \min(\epsilon_{\text{com}}, \epsilon_{\text{max}}/\alpha) \), where \( \epsilon_{\text{com}} \) is a comoving softening length and \( \epsilon_{\text{max}} \) is a (maximum) physical softening length. The softening lengths and other simulation parameters are listed in Table 1.

In each simulation, we output 95 snapshots spaced logarithmically in cosmic scale factor \( a \) between \( z = 8 \) and \( z = 0 \). To detect halos in each snapshot, we use the friends-of-friends (hereafter FOF) algorithm (Davis et al. 1985) with a linking length of \( \delta = 0.2 \) times the mean inter-particle separation \( L_{\text{box}} N_p^{-1/3} \). This conventional choice of linking length should produce halos with a mean overdensity of approximately 180, a value close to the virial overdensity predicted by the spherical collapse model in an Einstein de-Sitter cosmology. We use it here for consistency with previous work, although whether it is the “best” choice of linking length is a complex issue (More et al. 2011). We also use the SMOOTH algorithm (Stadel 1995) to compute the mean local density around every particle. (These tools can be found at http://www-hpcc.astro.washington.edu/tools/tools.html.) From the initial FOF catalog we select all halos with 1000 particles or more, as our experiments suggest this degree of resolution is required for an accurate determination of structural parameters. This reduces the number of usable FOF groups to 1560, 1580, and 1532, respectively, for simulations Sim60, Sim120, and Sim240, giving us a total sample of 4672 halos with sufficient resolution to measure structural parameters reliably.

In a sample defined solely by FOF linking, many groups will be multi-component systems caught in the act of merging. This can introduce significant scatter in the measured structural properties of halos. We can flag unrelaxed systems by using the minimum value of \( x_{\text{CoM}} \) (see Equation (3) below) as a goodness-of-fit indicator. Alternately, the asymmetry of the mass distribution can also be used to distinguish relaxed and unrelaxed systems. The center of each halo is taken to be the position of the particle with the highest local mean density within the FOF group, \( x_\rho \), as centering on this point generally appears to produce smoother fits to the density profile than if we use the center of mass of the group, \( x_{\text{CoM}} \). Halos that have undergone recent major mergers are commonly left with an asymmetric mass distribution. The offset between \( x_\rho \) and \( x_{\text{CoM}} \) relative to the size of the halo is a convenient measure.
of asymmetry, and thus relaxedness. By excluding from the sample systems with large offsets, we can filter our sample to select virialized halos with smooth and symmetric density profiles. Defining $x_{\text{off}} \equiv |x_0 - x_{\text{CoM}}|/r_{\text{vir}}$, we select from our full sample (all three simulations combined) a sub-sample of 3290 relaxed halos for which $x_{\text{off}} < 0.07$ and $\chi^2_{\text{NFW}} < 0.5$. These correspond to the criteria used by Macciò et al. (2007) and by S11 (our $\chi^2_{\text{NFW}}$ is their $\rho_{\text{rms}}$). We note that they are fairly inclusive, classifying 70% of our full sample as relaxed; Power et al. (2012), for instance, suggest using the stricter criterion $x_{\text{off}} < 0.04$ to select relaxed systems. In what follows, we will consider results both for the full sample of 4672 halos and for the relaxed sub-sample of 3290 halos.

2.2. Density Profiles

To determine the spherically averaged density profile, we bin the halo particles radially in equal-sized bins, starting from the center of the halo (i.e., $x_0$) and proceeding outward until we reach the virial radius, $r_{\text{vir}}$, defined by

$$M_{\text{vir}} = \frac{4\pi r_{\text{vir}}^3}{3} \Delta \rho_c,$$

where $M_{\text{vir}}$ is the virial mass of the halo, taken to be the FOF mass, and $\Delta \rho_c = 200$. We note that an average of about 16% of the outer particles linked by FOF lie outside $r_{\text{vir}}$ at low redshift, so this indicates the mean overdensity of the FOF groups is slightly less than 200.

We fit a Navarro–Frenk–White (NFW, Navarro et al. 1996) density profile to each halo over the radial range $[0, 0.1 r_{\text{vir}}, r_{\text{vir}}]$. We do not include the innermost 1% of the NFW profile in the fit, since resolution effects and softening may systematically alter the density profile there. The NFW profile is

$$\rho_{\text{NFW}}(r) = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2},$$

where $r_s$ is the scale radius (the radius at which the logarithmic density slope is $-2$) and $\rho_s$ is the characteristic density. Best-fit values of these parameters are obtained by minimizing the quantity

$$\chi^2_{\text{NFW}} = \frac{1}{N_{\text{bins}}} \sum_{i=1}^{N_{\text{bins}}} \left[ \ln \rho(r_i) - \ln \rho_{\text{NFW}}(r_i) \right]^2,$$

where $N_{\text{bins}}$ is the number of bins used in the fit. This minimization technique is logarithmic rather than linear in order to give similar weight to the fitting near the center and at the outer edges of the halo (Jing 2000; Macciò et al. 2007). The minimization code uses the Levenberg–Marquardt method to find the best-fit values of $r_s$ and $\rho_s$. As mentioned previously, we find that using the particle with the highest local density as the center of the halo produces good NFW fits to the density profile, while fits using the center of mass have larger residuals. Bin size also affects our results somewhat; after some experimentation we chose to use a bin size of 15 particles to measure the profile and determine halo concentration.

2.3. Concentrations

Given an NFW density profile fit, the concentration parameter is defined as $c \equiv r_{\text{vir}}/r_s$. Here $r_{\text{vir}}$ is the virial radius, that is the radius around the halo center that encloses mass $M_{\text{vir}}$, as defined in Equation (1). Physically, concentration is a measure of the size or density of the central region relative to the size or density of the whole halo.

We could determine the concentration of a halo by measuring $r_{\text{vir}}$ and $r_s$ separately and taking the ratio of the two. In practice, a slightly different technique appears to give more robust results. Integrating the NFW profile with respect to radius out to $r_{\text{vir}}$, the enclosed mass is given by

$$M_{\text{vir}} = 4\pi \rho_s r_s^3 \left[ \ln(1 + c) - \frac{c}{1 + c} \right].$$

Comparing the definition of the virial mass (Equation (1)) with Equation (4), we can derive the nonlinear relation

$$f(c) = \frac{1}{c^3} \left[ \ln(1 + c) - \frac{c}{1 + c} \right] = \left( \frac{\Delta_{\text{vir}},r_{\text{vir}}}{3\rho_s} \right).$$

Solving this equation then gives us the concentration $c$, as in Klypin et al. (2002) and Gavazzi et al. (2003). We will use this second method throughout the rest of the paper, although we find that the two methods for computing concentration are generally consistent with one-another to within 2%–3%. Typical errors in concentration are large (20%–30%) for our smallest halos, but drop below 10% for halos with more than 3000 particles.

The mass–concentration relation measured in our simulations (cf. the top panel of Figure 14) is in good agreement with previous determinations (e.g., Macciò et al. 2007; Neto et al. 2007).

2.4. Mass Accretion Histories and Ages

To determine the MAH for a given halo, we construct its merger tree, working backward from the final snapshot at $z = 0$. The merger tree links the final halo to its progenitors at each previous redshift step. To determine whether a halo $H_{\text{later}}$ found at a later time is related to an earlier halo $H_{\text{earlier}}$, we require that more than half of the particles in $H_{\text{earlier}}$ be contained in $H_{\text{later}}$. If this condition is satisfied then we say that halo $H_{\text{earlier}}$ is a parent halo (or progenitor) of the child halo $H_{\text{later}}$. This method of defining parent halos restricts all earlier halos to have at most one child halo in the consecutive snapshot. The method does not restrict the number of parent halos a child halo may have, so in order to construct the MAH, we define the main parent to be the one with the largest contribution of particles to the child halo. This then produces a single sequence of parent halos in the merger tree, so we can assign a well-defined MAH to the halo in the final snapshot.

We fit our final MAHs with the two-parameter model proposed by Tasitsiomi et al. (2004) and McBride et al. (2009)

$$M(z) \equiv M(z)/M(0) \sim (1 + z)^\beta e^{-\gamma z},$$

where $\beta$ and $\gamma$ are free parameters to be fit and $M(0)$ is the mass of the halo at $z = 0$. This form is a generalization of the earlier one-parameter fit proposed by Wechsler et al. (2002). The physical interpretation of the parameters $\beta$ and $\gamma$ can be seen by noting that

$$\frac{d}{dz} M(z) = -(1 + z)^\beta e^{-\gamma z} [\beta (1 + z)^{-1} - \gamma]$$

$$= \gamma - \beta \quad \text{for} \quad z \sim 0.$$ (8)

Thus $(\gamma - \beta)$ is the logarithmic growth rate of the halo in the limit $z \to 0$ or $a \to 1$. 


We perform the model fits to each MAH by minimizing the quantity

$$
\chi^2_M(\beta, \gamma) = \frac{1}{N_{\text{snap}}} \sum_{i=1}^{N_{\text{snap}}} \left[ \frac{M(z_i)}{M(0)} - (1 + z)^{\beta} e^{-\gamma z} \right]^2,
$$

(9)

where $N_{\text{snap}}$ is the number of snapshots for which the halo has existed in our sample. We note there are many other quantities beyond $\beta$ and $\gamma$ that we could extract from the MAHs, notably the time of the last major merger, defined as the last point where the MAH increased by more than some fraction in a single step. We will focus here on the smoother features of the MAH and leave mergers for future work.

2.5. Shape and Triaxiality

In order to describe the shape of a halo, one can diagonalize its moment of inertia tensor

$$
I_{ij} = \sum_{n=1}^{N} m_n (r_{i,n}^2 \delta_{ij} - r_{i,n} r_{j,n}^2),
$$

(10)

where $r_{i,n}$ is the $i$th Cartesian component (relative to the halo center) of the $n$th particle of the halo. This will produce eigenvalues $\lambda^2_1$ that are related to the relative axis lengths $a, b, c$ of the halo (assuming an ellipsoidal mass distribution) by

$$
\lambda_1^2 = \frac{1}{5} M_{\text{halo}} (b^2 + c^2),
$$

(11)

$$
\lambda_2^2 = \frac{1}{5} M_{\text{halo}} (a^2 + c^2),
$$

(12)

$$
\lambda_3^2 = \frac{1}{5} M_{\text{halo}} (a^2 + b^2).
$$

(13)

With simple linear combinations of Equations (11)–(13), one can then obtain the relative axis lengths.

A more direct—but equivalent—method of obtaining the relative axis lengths $a, b, c$ is to diagonalize the tensor describing the second moment of the mass distribution

$$
I_{ij}^2 = \sum_{n=1}^{N} m_n r_{i,n} r_{j,n}.
$$

(14)

The three eigenvalues $\lambda_i^2$ from the diagonalization of the matrix in Equation (14) will be the squares of the relative axis ratios, so the relative lengths of the principal axes can thus be determined as $a, b, c = \{ |\lambda_1^2|^{\frac{1}{2}}, |\lambda_2^2|^{\frac{1}{2}}, |\lambda_3^2|^{\frac{1}{2}} \}$. By convention, throughout the rest of the paper we will re-label the axis lengths as necessary so that $a > b > c$. We note there are several other possible ways to measure shape—see Zemp et al. (2011) for a discussion of the relative advantages of different techniques.

As well as taking ratios of the individual axis lengths, we will consider the triaxiality index $T \equiv (a^2 - b^2)/(a^2 - c^2)$. Oblate spheroids have $T = 0$, while purely prolate halos have $T = 1$. Halos have a strongly skewed distribution of $T$ values, however, so we also consider the elongation parameter $E \equiv (b^2 + c^2)/(2a^2)$, whose distribution is more symmetric and Gaussian.

Our use of three shape parameters $T, E,$ and $c/a$ may seem somewhat redundant; we consider all three parameters to show that they do indeed correlate well with each other, and to allow direct comparison with the results of S11, J11, and future work.

2.6. Summary of Measured Halo Properties

In studying correlations between halo properties, we will consider a large set of properties so as to overlap with the two recent PCA studies, S11 and J11. Clearly, not all the parameters listed below are independent. For example, we will assess several different definitions of shape, but they all ultimately depend on the two axis ratios $c/a$ and $b/a$. Our main goal will be to reproduce the patterns seen in S11 and J11, and then relate them to aspects of the MAH. Any redundant parameters will appear as strong correlations in our analysis and can be combined or ignored accordingly. In the analysis, we have taken logarithms of each halo parameter to obtain a more Gaussian distribution, although our experiments show that this does not have a significant effect on the final results of the PCA.

1. Virial mass $M_{\text{vir}}$. This is the total mass of halo within a virial radius $r_{\text{vir}} = r_{200c}$ of the halo center, which is taken to be the particle of highest local density within the linked group found by FOF. The virial radius is defined such that the halo has an average density of $200 \rho_c$. We use units of $h^{-1} M_{\odot}$ throughout.

2. Concentration $c \equiv r_{\text{vir}}/r_s$. The concentration $c \equiv r_{\text{vir}}/r_s = r_{200c}/r_s$ is determined using Equation (5), as explained above.

3. Formation redshift $z_\ast$ for $x \in (0, 1)$. This is the redshift at which the MAH reaches $M(z)/M(0) = x$. The most common example is $z_{\ast 0.5}$, at which $M(z_{\ast 0.5}) = 0.5 M(0)$.

4. Mass fraction history ($M/M_0$), for $z > 0$. The mass fraction history is a complementary measure of age, equal to $M(z)$ for a specified value of $z$.

5. Relaxedness $x_{\text{off}}$. This is the distance between the center of mass of the halo and the particle of highest local density, divided by the virial radius of the halo $r_{\text{vir}}$. Relaxed halos without any recent mergers will typically have a small value for $x_{\text{off}}$. Throughout this paper, we will use $x_{\text{off}} < 0.07$, $x_{\text{off}} < 0.05$ to define the relaxed sample.

6. Triaxiality $T$. Defined as $T = (a^2 - b^2)/(a^2 - c^2)$ where $a > b > c$ are the lengths of the principal axes of the halo. $T$ measures the prolateness or oblateness of a halo. Spherical and sausage-shaped halos have $T \approx 1$, while disk-shaped halos have $T \approx 0$.

7. Elongation $E$. This is defined as $E = [(b/a)^2 + (c/a)^2]/2$. The distribution of values of $E$ is more Gaussian than the distribution of triaxiality values.

8. Sphericity $c/a$. This is a measure of the sphericity of the halo, dividing the shortest axis length $c$ by the longest axis length $a$.

9. Spin $\lambda$. To provide a measure of rotation, we use a dimensionless spin parameter $\lambda \equiv I_{\text{vir}}/(2GM_{\text{vir}}^3)$, where $I_{\text{vir}}$, $M_{\text{vir}}$, and $r_{\text{vir}}$ are the total angular momentum, virial mass, and virial radius, respectively. A different possible definition of the spin parameter is discussed in Macciò et al. (2007).

10. Environment $D_{n,f}$ for $n \in \mathbb{Z}$ and $f > 0$. The dimensionless environment parameter $D_{n,f}$ is defined as distance of the $n$th nearest halo that has a virial mass greater than $f M_{\text{vir}}$, divided by the virial radius of the neighbor $r_{\text{vir}}$. This definition is physically motivated because $D_{n,f}$ scales as the tidal force exerted by the neighbor to the $-1/3$ power (Haas et al. 2012). To match the results of J11, we choose $n = 1$ and $f = 0.1$.

Prior to S11 and J11, correlations between these parameters had been established by many authors. Of particular interest
are the correlations between MAH or structural parameters and environment, since these have implications for halo occupation models and similar methods relating observed galaxies to the dark matter distribution. Generally speaking in the Press–Schechter or excursion-set models of halo growth, historical and structural properties, which trace a halo’s past, should be almost independent of environment, which determines its future. Thus, it was surprising to discover (Sheth & Tormen 2004; Gao et al. 2005) that formation time (taken to be \( z_{\text{form}} \)) correlates with environment, in the sense that halos in dense environments form earlier, and early-forming halos are more clustered.

This “assembly bias” has since been confirmed in many other simulations (Avila-Reese et al. 2005; Harker et al. 2006; Wechsler et al. 2006; Mautz et al. 2007; Faltenbacher & White 2010), and its implications for the galaxy distribution have also been studied extensively (Zhu et al. 2006; Croton et al. 2007; Reed et al. 2007). The trend appears to be due to a combination of tidal forces (which limit the late-time growth of halos in dense regions; Hahn et al. 2007) and the shortcomings of the simplest excursion-set picture (Dalal et al. 2008), although the strength of the trend is sensitive to the definition of formation time (Li et al. 2008).

In this work, we are considering a fairly local definition of environment, so we expect to be sensitive to some of these effects (e.g., tidal ones) but not others. In general, the environmental correlations produced by assembly bias are not very strong in our halo sample; we will show this below in the results of the structural PCA.

3. ANALYSIS OF THE MASS ACCRETION HISTORIES

3.1. The (\( \beta, \gamma \)) Fit

For each of our well-resolved halos, we fit the functional form in Equation (6) to the MAH to determine values of \( \beta \) and \( \gamma \). Figure 1 shows the resulting distribution of parameter values. (The distribution for each individual simulation is similar, so we show results for all three simulations together.) The color coding indicates the value of the relaxedness parameter \( x_{\text{off}} \) (left-hand panel) or the formation redshift as \( \log_{10}(z_{\text{form}}) \) (right-hand panel). Finally, the points on the right-hand panel of Figure 1 indicate the values of \( \beta \) and \( \gamma \) used to fit the curves in Figure 2 (see below).

While the functional form of Equation (6) produces a reasonable fit to most of our MAHs, we can see that \( \beta \) and \( \gamma \) are not necessarily a natural choice of variables, as there is a strong correlation between them for most MAHs. These parameters are also degenerate in many cases, with different combinations of values providing almost identical fits to the MAH, as discussed in Taylor (2012). Finally, for a minority of our MAHs (mainly those with rapid growth at late times), the fit to the actual trajectory is poor and the resulting \( (\beta, \gamma) \) model loses mass over significant range of redshift. We discard these fits (roughly 13% of the MAHs) when analyzing the smooth fitted trajectories below.

An alternative variable is the quantity \( (\gamma - \beta) \), which corresponds to the logarithmic growth rate as \( z \to 0 \). Given the simplicity of the \( (\beta, \gamma) \) fit, this quantity also measures the overall age of the system, as can be seen clearly from the color scale in the right-hand panel of Figure 1. We will discuss a non-parametric representation of the MAHs, which captures age in the most direct way, in the next section.

3.2. Principal Component Analysis of the MAHs

PCA is a non-parametric technique for decomposing a set of \( N_P \) possibly correlated variables into an equal or smaller number of independent variables. Because the final number of variables may be smaller than the initial number, PCA has the ability to reduce the dimensionality of a data set and uncover hidden patterns.

Technically, PCA involves determining the eigenvectors (principal axes or principal components \( P_i \)) of the covariance matrix and their respective eigenvalues \( \lambda_i \). Using the convention that principal components are arranged by decreasing eigenvalue such that \( \lambda_1 > \lambda_2 > \ldots > \lambda_{N_P} \), the eigenvalue \( \lambda_i \) gives the fraction of the total variance in the direction of the principal axis \( P_i \). Thus PCA provides us with principal axes \( P_1 \) pointing in the direction of greatest variance in the data, \( P_2 \) pointing in the direction of second greatest variance orthogonal to \( P_1 \), etc. Because the individual properties analyzed by PCA may have different dimensionality and/or magnitude, it is important to transform the initial data such that the mean of each parameter is 0 and the variance is 1, assuring that all the input fields will be treated equally. Normalizing the variables in this way will also scale the eigenvalues to sum to \( N_P \), so the relative contribution of the principal component \( P_i \) to the variance is the fraction \( \lambda_i / N_P \).
We can use PCA to derive a more fundamental decomposition of the MAHs, taking as the input data vectors of values $\mathcal{M}(z_i)$ for a specified set of 70 redshift steps $z_i$. The resulting principal axes will then capture the basic shape variations of the MAHs. For this analysis, we consider a subset of MAHs for 817 halos extracted from Sim120 that are in the well-resolved sample (1000 particles or more, corresponding to masses of $10^{12} h^{-1} M_\odot$ or more) at $z = 0$ and have parents detected above the basic FOF limit of 200 particles or more (corresponding to masses of $2 \times 10^{11} h^{-1} M_\odot$ or more) for at least 70 outputs (i.e., going back to $z \sim 4$). We note that as a result of the latter restriction, this sample is more limited than the one shown in Figure 1 and will be slightly biased to older halos.

Figure 2 shows the mean MAH for these systems (central black curve), plotted in terms of the scale factor $a$. The variance associated with each MAH principal component is given in Table 2.

The first principal component (MAH-PC$_1$), which accounts for 58.4% of the total variance in the MAHs, represents the most important variations away from this average. We can determine the nature of MAH-PC$_1$ by adding or subtracting from the mean MAH the first principal component vector, normalized to 1 or 2 times the rms scatter along the MAH-PC$_1$ axis. The resulting curves, illustrating $\pm 1\sigma - 2\sigma$ deviations along the MAH-PC$_1$ axis, are indicated by the upper (red) and lower (blue) curves in the left-hand panel of Figure 2. Note the sign convention in PCA is arbitrary; here we choose the positive sign for MAH-PC$_1$ to correspond to values of $\mathcal{M}$ above the mean.

Figure 2. Mean MAH for 817 halos from Sim120 (central black solid line), and its variation as one moves $\pm 1$ and $\pm 2$ standard deviations along the first (left-hand panel) and second (right-hand panel) principal axes defined by PCA. In the left-hand panel, the upper (red) curves represent older halos and the lower (blue) curves represent younger halos. In the right-hand panel, they represent halos whose accretion rate has decelerated or accelerated close to $z = 0$, respectively. The gray lines indicate $(\beta, \gamma)$ fits to the curves (see Equation (6)).

(A color version of this figure is available in the online journal.)

| Principal Axis | Contribution to the Total Variance (%) |
|----------------|-----------------------------------------|
| $n = 1$        | 58.4                                    |
| $n = 2$        | 25.1                                    |
| $n = 3$        | 5.4                                     |
| $n = 4$        | 3.3                                     |
| $n \geq 5$     | 7.8                                     |

Clearly, MAH-PC$_1$ corresponds to the overall age of the halo, with the upper (red) curves indicating older systems and the lower (blue) curves indicating younger systems. Both the mean and the $\pm 1\sigma - 2\sigma$ deviations are all well fit by the $(\beta, \gamma)$ model (gray curves). The best-fit values of $\beta$ and $\gamma$ are plotted as triangles on the right-hand panel of Figure 1. Roughly speaking, they are spread out in the $(\gamma + \beta) = \text{constant}$ direction, but the precise positions appear complicated in this parameter space. This may be the result of the degeneracies in the $(\beta, \gamma)$ fit discussed earlier.

Since MAH-PC$_1$ accounts for the largest fraction of the scatter between individual MAHs, and the MAH-PC$_1$ coordinate tracks age monotonically, the value of this coordinate for an individual MAH represents the best way of defining the age of the corresponding halo, at least in terms of overall mass assembly. Unfortunately, the $(\beta, \gamma)$ model does not provide a clear, monotonic, single-parameter measure of age defined in this way. After some experimentation, we have determined an alternative single-parameter analytic fit to the derivative of $\mathcal{M}(a)$:

$$
\frac{d}{da} \mathcal{M}(a) = f(a) = \exp[-Sa] \exp \left[ -\frac{1}{g(S)a} \right],
$$

where $S$ is the fitted age parameter and $g(S)$ is a constant related to $S$ by

$$
g(S) = 5.0 \exp \left[ -\frac{S}{4.5} \right] \exp \left[ -\frac{1}{5S} \right].
$$

Integrating this function over scale factor $a$ then gives $\mathcal{M}(a)$ or equivalently $\mathcal{M}(z)$. Obviously this analytic fit is less elegant than the form proposed by McBride; however, it provides a single-parameter measure of age that is roughly equivalent to the coordinate value on the MAH-PC$_1$ axis.

Figure 3 shows this new fit relative to $\pm 1\sigma - 2\sigma$ deviations along the MAH-PC$_1$ axis. The fit is excellent for the average MAH and for negative values of the MAH-PC$_1$ coordinate (i.e., recently assembled systems). For positive values, it is good up to the $+1\sigma$ deviation curve, but not as good beyond that. (Varying the function $g(S)$ would produce a slightly better fit, at the expense of introducing a second parameter.) The $\pm 1\sigma - 2\sigma$ curves shown in Figures 2 and 3 treat deviations along the MAH-PC$_1$ axis as Gaussian, however, which is not quite correct, as
Figure 3. Same as Figure 2(a), but showing the new single-parameter fit given in Equation (15) (light gray curves) over the range $-2\sigma$ to $+1\sigma$. The dashed gray line shows the relatively poor fit to the $+2\sigma$ deviation.  
(A color version of this figure is available in the online journal.)

Figure 4. Distribution of the MAH-PC$_1$ coordinate value for 817 well-resolved halos from Sim120. The top axis shows the coordinate value in units of the standard deviation of the distribution, $\sigma = (0.584 \times 70)^{1/2} \sim 6.4$.  
(A color version of this figure is available in the online journal.)

Figure 5. Relative contribution of MAH-PC$_1$ to the total deviation from the mean MAH vector, for 817 well-resolved halos from Sim120. The clustering around $\pm 1$ indicates that individual MAH tend to lie close to the MAH-PC$_1$ axis.  
(A color version of this figure is available in the online journal.)

shown in Figure 4 and discussed below. In practice, only the oldest 5% of all systems have values of the MAH-PC$_1$ coordinate in the range where it is poorly fit by our formula. Nonetheless, we suggest that the results of the new fit be treated with caution when applied to the oldest systems.

The second principal component, MAH-PC$_2$, accounts for a further 25% of the variance. The right-hand panel of Figure 2 shows the effect of $\pm 1\sigma - 2\sigma$ deviations away from the mean along the MAH-PC$_2$ axis. This variation is more complicated than that of MAH-PC$_1$, but roughly speaking, it represents acceleration or deceleration in the MAH. Physically, acceleration in the MAH might correspond to an initially isolated halo falling into or being accreted by a dense region, such that the growth rate increases at late times. Conversely, deceleration might correspond to a single region collapsing rapidly but then becoming isolated from other regions, so that the growth rate decreases at late times. Neither the $(\beta, \gamma)$ model nor our new form are
good fits to this family of curves, particularly not to the ones that accelerate rapidly at late times. The relationship between the fitted values of the $\beta$ and $\gamma$ is much simpler in this case, as indicated by the squares on the right-hand panel of Figure 1, but given the poor fit to the curves overall, it is not clear whether this is significant.

In what follows, we will often refer to the value of the MAH-PC$_1$ coordinate for an individual MAH, that is to say the projection of the individual MAH vector onto the MAH-PC$_1$ axis. The distribution of MAH-PC$_1$ coordinates for our set of MAHs is approximately Gaussian, though slightly skewed, as shown in Figure 4. Given the MAH-PC$_1$ coordinate measures age, we expect the details of the distribution (mean and scatter) to depend on the ages and redshifts of sample of halos considered. The cuts to our sample (requiring a resolved progenitor at $z = 4$) may also explain some of the skewness. The top axis shows the coordinate value normalized by the standard deviation of the distribution; we will use this normalized scale in subsequent figures.

Since MAH-PC$_1$ captures much of the variation between MAHs, the deviation of individual MAHs from the mean tends to be reasonably aligned or anti-aligned with the MAH-PC$_1$ axis. Figure 5 shows that the MAH-PC$_1$ coordinate divided by the length of the MAH vector clusters around the values $\pm 1$ as a result.

We have examined the higher principal components of the MAH. Generally speaking, they appear to be Fourier-like decompositions of the curve with increasing numbers of oscillations about the mean MAH. Given that individually they account for relatively little of the scatter (5.4% for MAH-PC$_3$, 3.3% for MAH-PC$_4$, $< 7.8\%$ for subsequent PCs), we will not discuss them any further. The total power in these higher components is only 15%, so our set of MAHs are reasonably well fit by smooth curves, despite the jumps produced by major mergers. This may be less true for the full set of MAHs that disappear below our resolution limit before $z = 4$.

3.3. Relation to Excursion Set Theory

We can ask to what extent the basic shape of the mean MAH, or the variations of MAH-PC$_1$ around it, reflect the statistics of halo growth predicted by excursion-set (or extended
Press–Schechter; EPS) theory. In excursion set theory, the growth of a halo can be considered a random walk in the space of variables $\sigma^2(M)$ and $\omega(z)$, where $\sigma^2(M)$ is the variance of the primordial density field smoothed on a mass scale $M$, and $\omega(z) = \delta_c(z)$ is the height of the barrier for spherical collapse at redshift $z$ (Lacey & Cole 1993). The distribution of jumps $\Delta \sigma^2$ over a redshift step corresponding to $\Delta\omega$ is a simply a Gaussian in $(\Delta\omega/\sqrt{\Delta\sigma^2})$ (cf. Lacey & Cole 1993, Equation (2.15)). This suggests that MAHs should look much simpler in the coordinate system $(\Delta\omega, \Delta\sigma^2)$. Figure 6 shows the mean MAH and the variations along the MAH-PC1 axis in $(\Delta\omega, \Delta\sigma^2)$ coordinates. As anticipated, the different curves from Figure 2 are now almost straight lines in these new coordinates. On closer examination, however, there is still a slight residual curvature to the lines. Tests with artificial MAHs suggest that this is at least partly due to the difference between the variable $\Delta \sigma^2 = \sigma^2(M_j) - \sigma^2(M_0)$, which is plotted here, and $\Delta \sigma^2 = \sigma^2(M_j) - \sigma^2(M_0)$, which is assumed in EPS theory, for a section of the trajectory going from $M_j$ to $M_i > M_j$. The curvature may also reflect from corrections introduced to EPS statistics by ellipsoidal collapse (Sheth et al. 2001; Sheth & Tormen 2002). Overall, however, the transformation from $(\Delta\omega, \Delta\sigma^2)$ to $(a, M)$ coordinates accounts for most of the characteristic curvature of MAHs in $(a, M)$ space. (Note that given the relatively smooth variation of $\sigma(M)$ with $\log(M)$, this also explains why $\log[M(z)]$ is almost linear in $\Delta\omega$; cf. Neistein & Dekel 2008b.)

This approximate result is useful, as it provides a rough way of predicting MAH distributions as a function of background cosmology. Predicting MAH distributions and merger rates more precisely is much more challenging, however, and may be beyond current analytic methods (e.g., Neistein & Dekel 2008a; Zhang et al. 2008a, 2008b; Neistein et al. 2010; Fakhouri et al. 2010; de Simone et al. 2011).

4. ANALYSIS OF STRUCTURAL PROPERTIES

4.1. Correlations between Structural Parameters

The simplest way to search for relationships between the structural properties of halos is to measure the correlation of one property with another. We will use the Spearman rank correlation coefficient as a quantitative measure of the correlation strength. The Spearman rank correlation coefficient, $\rho_{scc} \in [-1, 1]$, is a non-parametric measure of correlation which determines the strength of the relationship between two variables based on whether the data can be characterized by a monotonic function. The values $\rho_{scc} = +1, -1$ correspond to perfect correlation or anti-correlation, respectively, i.e., monotonic relationships with no scatter, while $\rho_{scc} \sim 0$ indicates uncorrelated variables.

The full correlation analysis of all 11 halo properties is shown in the Appendix, Figure 14. Broadly speaking, the strongest correlations are between concentration $c$ and age indicators such as the formation redshift $z_0.5$. Many parameters, such as $z_0.5$ and $M(0.8)$, are nearly degenerate. Relatively significant correlations also exist between mass and age indicators (e.g., mass and $z_0.2$, with $\rho_{scc} \sim -0.51$), and between shape and concentration (e.g., $E$ and $c$, with $\rho_{scc} \sim 0.51$). The former is easily understood as a consequence of hierarchical structure formation (more massive halos have formed more recently; e.g., Mo & White 2002). The latter appears to be a newly discovered pattern, although it is consistent with previous studies of halo shape (e.g., Allgood et al. 2006), that found that older halos are rounder.

Correlations between halo properties have already been analyzed by J11. Seven of our parameters—$c$, $z_0.5$, $x_{off}$, $c/a$, $T$, $\lambda_{spin}$, and $D_{env}$—match theirs directly. (We omit substructure because we lack the resolution to detect it convincingly in many of our halos.) Overall our results for these parameters are almost identical to theirs; concentration and age are the most fundamental parameters, with mass and shape following behind. We also find as they do that environment is not strongly correlated with any of the other halo properties. Our measured values of $\rho_{scc}$ generally match theirs to within 0.05–0.1, despite the differences between the samples (they consider only halos with $10^4$ particles or more, but cover roughly the same halo mass range by using five simulations rather than three). The only results that differ significantly are for the spin parameter $\lambda$, where we find more correlation with mass and less with age-related parameters such as $c$, $z_0.5$, or $x_{off}$. The origin of this discrepancy may be numerical resolution, since their halos are better resolved than ours.

4.2. Principal Component Analysis

We also perform PCA, as described in Section 3.2, on the 11 halo properties described in Section 2.6. Our results are broadly consistent with the previous analyses of S11 and J11, but we include them here for completeness. The strongest structural principal components (S-PCs) with their respective principal axes are listed in Table 3 for the full sample and Table 4 for the relaxed sub-sample. Numbers in bold indicate significant contributions to the principal component, taken here to be coordinate values $|v_i| > 0.3$. We note that several of our parameters are almost degenerate, so our S-PCs have significant contributions from more parameters than in the previous analyses of S11 and J11.

In principle, PCA is unbiased only if the underlying single-parameter distributions are Gaussian. Some of our parameters (e.g., triaxiality $T$) are clearly non-Gaussian, however, which could affect our results. We have checked for bias by re-running the analysis excluding the least Gaussian parameters and find that these do not have a significant effect on the final results. We conclude that our main PCs are robust to the inclusion of non-Gaussian halo parameters. We summarize the properties of the four strongest PCs below.
1. $S-PC_1$. It includes concentration and the various age indicators ($z_{0.2}$, $z_{0.5}$, $(M/M_0)_{0.5}$) as its strongest contributors. Mass and elongation also contribute significantly to $S-PC_1$ for the relaxed sample; for the full sample relaxedness $x_{\text{off}}$ is more important than mass. The overall weight of $S-PC_1$ and its composition are roughly in agreement with the PCA from S11 and J11; we verify that concentration, age, mass, and shape seem to be related, fundamental parameters. The most likely explanation is that concentration and shape are determined by the detailed processes of halo assembly and are thus correlated with assembly history and therefore age, whereas age is correlated with mass through the separate physics of hierarchical structure formation. We will discuss some of the evidence for this interpretation in Section 5.1.

2. $S-PC_2$. It contributes 30%–40% as much variance as $S-PC_1$ and includes significant contributions from shape, age, and spin. In fact, the contribution from shape parameters ($T$, $E$, $c/a$) is stronger than in $S-PC_1$. This indicates that while shape is correlated with age and thus the MAH, it is also partly determined by an independent second parameter, distinct from age.

3. $S-PC_3$. It is then fairly distinct from the previous components, depending mainly on mass and spin. In the case of relaxed halos, environment also contributes at a lower but significant level to $S-PC_3$. As mentioned previously, the contribution of mass to $S-PC_3$ may be partly an artifact of numerical resolution, since J11 found less correlation between mass in spin in a sample of higher resolution halos covering the same mass range as ours.

4. $S-PC_4$. It consists mainly of environment, showing that environment as we have defined it is a fairly independent parameter, or conversely that structural properties correlate only weakly with our locally defined environmental parameter. For the relaxed subset there is a greater contribution to the vector from relaxedness and triaxiality, showing that these do correlate with local environment in relaxed systems. In general, however, we do not find a strong contribution from concentration or the age indicators, as might be expected from the results on assembly bias discussed in Section 2.6. This may be in part because our halo span a range of masses both above and below the characteristic structure formation mass $M_*$, and many assembly bias trends appear to switch sign at $M_*$ (e.g., Faltenbacher & White 2010).

4.3. The Origin of Halo Shapes

The results of the previous section indicate some correlation between formation history and shape, but also some independence between shape and age. There are various ways in which formation history could influence shape. One simple possibility is that the final shape of a halo is determined by the shape of the initial, uncollapsed matter from which it forms, as implied by previously measured correlations between shape and environment (e.g., Bailin & Steinmetz 2005; Altay et al. 2006; Zhang et al. 2009). This connection would not be apparent in the MAH, since it does not record the shape of the merging material, but only the rate at which it merges. We can test the hypothesis that final shape tracks initial shape by measuring the latter in early timesteps, including all the particles which eventually merge to form a halo.

Figure 7 shows how the elongation of a halo at $z = 0$ correlates with the elongation of the particle distribution from which the halo formed, measured at a higher redshift, for the full (left-hand panel) and relaxed (right-hand panel) samples. Looking back to recent redshifts, there is a strong correlation between initial and final elongation, particularly for relaxed halos; at higher redshift or for less relaxed halos the correlation gradually weakens. Table 5 gives the Spearman rank coefficients for the correlations, showing that they are significant back to $z = 1$.

We note, however, that a significant fraction of the final halo mass is already in place by these redshifts. From the mean MAH in Figure 2, the average factions of the final mass assembled by $z = 0.1, 0.5, 1.0,$ and 2.4 ($a = 0.9, 0.66, 0.5,$ and 0.3) are 0.9, 0.7, 0.4, and 0.15, respectively. Thus, it appears that much of the

---

### Table 3

| Principal Axis | $\lambda_{5}/\lambda_1$ | Contribution | $\log M_{\text{crit}}$ | $\log (c/a)$ | $\log (E)$ | $\log (\lambda_i/\lambda)$ |
|---------------|----------------------|--------------|------------------|---------------|--------------|------------------|
| $S-PC_1$      | 1.00                 | 40%          | 0.29             | $-0.39$       | $-0.34$      | $-0.37$           |
| $S-PC_2$      | 0.34                 | 14%          | 0.11             | 0.29          | $-0.46$      | 0.33              |
| $S-PC_3$      | 0.27                 | 11%          | 0.04             | 0.10          | 0.08         | 0.27              |
| $S-PC_4$      | 0.22                 | 9%           | 0.04             | 0.01          | 0.05         | 0.01              |
| $S-PC_5$      | 0.15                 | 4%           | 0.15             | 0.10          | 0.12         | 0.06              |

### Table 4

| Principal Axis | $\lambda_{5}/\lambda_1$ | Contribution | $\log M_{\text{crit}}$ | $\log (c/a)$ | $\log (E)$ | $\log (\lambda_i/\lambda)$ |
|---------------|----------------------|--------------|------------------|---------------|--------------|------------------|
| $S-PC_1$      | 1.00                 | 36%          | 0.34             | $-0.39$       | $-0.33$      | $-0.40$           |
| $S-PC_2$      | 0.43                 | 16%          | 0.11             | 0.10          | $-0.49$      | 0.32              |
| $S-PC_3$      | 0.29                 | 10%          | 0.19             | 0.06          | 0.20         | 0.06              |
| $S-PC_4$      | 0.24                 | 9%           | 0.11             | 0.17          | 0.20         | 0.17              |
| $S-PC_5$      | 0.16                 | 4%           | 0.04             | 0.17          | 0.43         | 0.36              |

### Table 5

| $\xi$ | $r_{E}$ (All Halos) | $r_{E}$ (Relaxed Halos) |
|-------|---------------------|-------------------------|
| 0.1   | 0.630               | 0.727                   |
| 0.5   | 0.338               | 0.444                   |
| 1.0   | 0.300               | 0.405                   |
| 2.4   | 0.082               | 0.128                   |
correlation between initial and final shapes may arise not from
the geometry of the initial conditions before the halo forms, but
from the conservation of the shape created as the halo forms.
We note that Vera-Ciro et al. (2011) have recently completed
a much more thorough study of the nature and mechanisms
by which shape evolves as a halo grows. Broadly speaking,
they find that galaxy halos are prolate during their early, rapid
growth and become more triaxial or oblate at later times, when
the accretion onto them is more isotropic.

4.4. Structural Properties versus MAH-PCs

Given the non-parametric decomposition of the MAHs into
their own set of principal components, presented in Section 3.2,
we can examine how structural properties relate to the main
features of the MAH. Figure 8 shows how $z_{0.5}$, concentration $c$, and relaxedness $x_{\text{eff}}$ correlate with the MAH-PC1 and
MAH-PC2 coordinate values, for the full sample (left-hand panels)
and the relaxed sub-sample (right-hand panels). The horizontal
axes are the coordinate value of the individual MAH
projected on to the MAH-PC1 or MAH-PC2 axes, rescaled by
the rms dispersion of each of these variables, respectively (this
is the same scale used for MAH-PC1 on the top axis of Figure 4). Note that in these figures we measured the MAH-PC coordinate
values from smooth ($\beta, \gamma$) fits to the MAHs, eliminating a small
fraction of MAHs where the fit was poor. This selection, together
with the requirement that systems have resolved progenitors at
$z = 4$, accounts for the sharp cutoffs at negative values of the
MAH-PC1 coordinate and positive values of the MAH-PC2 coordinate.

In the first row of panels, we see that the MAH-PC1 co-
dordinate is strongly correlated with $z_{0.5}$, particularly for older
systems. The relationship between $z_{0.5}$ and MAH-PC2 is more
complicated; there is no strong trend in the average value,
but decelerating systems (positive values of the MAH-PC2 coordinate) have less scatter in $z_{0.5}$. This can be understood
by noting that the $+1\sigma$ and $+2\sigma$ curves in the left panel of
Figure 2 converge to very similar values at $M/M(0) = 0.5$, so
$z_{0.5}$ becomes independent of MAH-PC2 at large positive values
of this coordinate.

Concentration (second row of panels) follows a similar
pattern, correlating well with MAH-PC1, particularly for old
and/or relaxed systems. Comparing the full and relaxed samples
(first two and last two columns), we see that young, low-
concentration systems are often unrelaxed as well. Overall, the
distribution suggests a simple relation between concentration and
MAH-PC1, which is slightly disturbed by mergers or rapid
growth in young systems. Analytic models with similar features
relating concentration to the MAH have been proposed by
Wechsler et al. (2002) and Zhao et al. (2009); they distinguish
phases of slow growth, where concentration gradually increases,
from phases of rapid growth, when concentration is reset to a
low value. Thus, concentration should be a good age indicator
for old, relaxed systems, but a less reliable one for low-
concentration and/or unrelaxed systems. We will investigate the
idea of the using of concentration as a quantitative age indicator
in the next section.

Comparing concentration and MAH-PC2, as with $z_{0.5}$ there
is little correlation, but the scatter is greatest in accelerating systems (negative MAH-PC2) and least in decelerating systems
(positive MAH-PC2). Finally, the relaxedness parameter $x_{\text{eff}}$
(bottom row of panels) is anti-correlated with age and slightly
anti-correlated with MAH-PC2. It seems sensible that younger
systems or ones whose growth has accelerated recently should
be less relaxed.

We note that the structural correlations we measure should
depend somewhat on the mass and redshift range of halo sample,
since these will determine the distribution of MAHs sampled,
via cosmology. Figure 9 shows how the range of MAH-PC1
dcoordinate values is limited for the most massive (and thus
youngest) halos, and much broader for lower-mass halos. This
pattern is expected from hierarchical structure formation, since
massive halos have formed more recently on average.

5. APPLICATION TO OBSERVATIONS

5.1. What Do Structural Properties Tell Us?

We have shown that the formation history of dark matter
halos, as summarized by their MAH, relates to their final
about an individual MAH? a set of measured structural properties, what do they tell us quantitatively from observable ones. The first question is, given practical application of our results is to infer unobservable quantities from observable ones. Thus, the main structure at a statistical level. Observationally, structure can be measured directly, whereas history cannot. Thus, the main aspect of the formation history does concentration trace most sensitively? Figure 10 shows the correlation between concentration and a number of the age indicators introduced previously. These are \( z_x \), the redshift by which the main progenitor of a halo had acquired a fraction \( x \) of its final mass at \( z = 0 \), for \( x = 0.2, 0.5, 0.8 \), and \( (M/M_0)_x \), the fraction of its final mass a halo had reached by redshift \( z \). They represent the values obtained by intersecting the MAH \( M(z) \) function with a series of horizontal or vertical lines, respectively.

Generally the correlation with \( z_x \) is tighter than with \( (M/M_0)_x \), and of all the age indicators, \( z_{0.2} \) shows the tightest correlation. Thus, concentration is most closely related to the early part of a halo’s formation history, when the dense central core of the halo is established; specifically, it correlates with the redshift at which this happened.

Given that MAH-PC\(_1\) captures much of the variation in MAHs, we might expect a close relationship between the MAH-PC\(_1\) coordinate value and observables such as concentration. Figure 11 shows concentration, relaxedness, and elongation versus the MAH-PC\(_1\) coordinate value, as well as concentration versus \( z_{0.5} \) for comparison. Here, we have also restricted the mass range to 0.6–1.0 \( \times 10^{13} \) \( h^{-1} M_\odot \), to remove any dependence on mass. The results are similar to those with no mass cut applied, however, supporting our earlier statement in Section 4.2 that concentration, relaxedness, and shape are determined by age, not by mass. (Note that the sample is smaller here, so the plots are more noisy than those in Figure 10.) The Spearman rank coefficients for the four relations are \( \rho_{\text{src}} = 0.57, 0.68, -0.31 \), and \(-0.22\). We conclude that concentration is an excellent estimator of the MAH-PC\(_1\) coordinate value \( (\rho_{\text{src}} = 0.68) \), tracing this component of the MAH almost as well as \( z_{0.2} \) \( (\rho_{\text{src}} \sim 0.75) \), as shown in the left panel of Figure 12), and slightly better than \( z_{0.5} \) \( (\rho_{\text{src}} \sim 0.6) \). Relaxedness and shape are (anti-)correlated with age, but not as strongly as concentration. We can generalize
Figure 10. Concentration vs. age indicators for the full sample.
(A color version of this figure is available in the online journal.)

Figure 11. Concentration, and relaxedness, and elongation vs. age indicators for a sample of mass $0.6-1.0 \times 10^{13} \, h^{-1} M_{\odot}$. Concentration is an excellent estimator of age, as parameterized by the MAH-PC1 coordinate value; relaxedness and shape correlate to a lesser degree with age. The Spearman coefficients for the four relationships are 0.57, 0.68, $-0.31$, and 0.22, respectively.
(A color version of this figure is available in the online journal.)

Figure 12. Strength of the correlation between structural properties and the formation redshift $z_f$ for all relaxed halos (left panel) and relaxed halos in the mass range $0.6-1.0 \times 10^{13} \, h^{-1} M_{\odot}$ (right panel). The formation redshift $z_f$ is the redshift by which the halo had built up a fraction $x$ of its final mass at $z = 0$, as a function of $x$. The vertical bars represent 95% confidence intervals.
(A color version of this figure is available in the online journal.)
the specific values of $x$ considered in Figure 10 to a continuous range of values, and test for correlations with other structural parameters. Figure 12 shows how the strength of the correlation between a given structural parameter and $z_x$ varies as a function of $x$ for all relaxed halos (left-hand panel) and relaxed halos in the mass range $0.6\times10^{13} M_\odot$ (right-hand panel). In each case, we use $\rho_{\rm src}$ as our measure of correlation. Looking at the left panel, we see that concentration is most strongly correlated with $z_{0.2}$ (although the decrease in correlation at $z = 0.1$ may be an artifact of our MAH selection and/or limited resolution at early times). Elongation $E$ and sphericity $c/a$ are correlated with formation epoch, but less strongly, and the strength of the correlation is relatively insensitive to $x$. Relaxedness $x_{\text{off}}$ shows a similar pattern of anti-correlation, with a broad peak in strength around $x = 0.4$–0.6. Spin and triaxiality are weakly anti-correlated with $z_x$. We also see that mass $M$ is anti-correlated with $z_x$, particularly for small values of $x$. This is as expected from the conventional picture of hierarchical structure formation, since halos within our sample that formed more recently will be more massive on average. For relaxed halos of fixed mass (right panel), the errors are much larger due to the smaller sample size, but the trends remain the same for all variables with the exception of mass, which is now constrained by the sample selection. Once again, this suggests that assembly history, not mass, has a fundamental role in determining structural properties.

Ignoring the slight differences between samples, we can summarize the results shown in Figures 10, 11, 12, and 14 fairly simply. Concentration is most strongly correlated with $z_{0.2}$ ($\rho_{\text{src}} = 0.75$), but is almost as strongly correlated with MAH-PC$_1$ ($\rho_{\text{src}} = 0.68$). Relaxedness $x_{\text{off}}$ is most (anti-)correlated with $z_{0.5}$ ($\rho_{\text{src}} = -0.50$) and correlates slightly less well with MAH-PC$_1$ ($\rho_{\text{src}} = -0.31$). Elongation $E$ is also most correlated with $z_{0.5}$ ($\rho_{\text{src}} = 0.38$) and correlates less well with MAH-PC$_1$ ($\rho_{\text{src}} = 0.22$). Measurements of concentration, relaxedness, and elongation can therefore provide statistical information about the MAH of an individual system, albeit with large uncertainties.

An alternate approach is to use structural measurements to define samples of halos with different average properties. Given the correlations shown above, these structurally selected samples will then have systematically different mass accretion histories. We consider this approach next.

5.2. Splitting Halo Samples by Structural Properties

Given the considerable scatter in all the correlations seen so far between structure and formation history, the measurement of structural parameters in any single system will give only a rough indication of its past history. A more interesting possibility is to average over large samples of objects. As mentioned in Section 1, structural properties such as shape and concentration have now been measured for hundreds of galaxy cluster halos. In the case of the most massive clusters, it may even be possible to constrain the full three-dimensional shape by combining lensing with other observations (e.g., Morandi et al. 2011). Average shapes and concentrations for samples of galaxy halos may soon be measurable as well, given planned and ongoing weak lensing surveys.

In Figure 13, we show how, by splitting a sample of halos into sub-samples based on structural properties at $z = 0$, we can construct sub-samples of halos with systematically different formation histories, as indicated by the distribution of $z_{0.5}$ values. In the four panels, we consider splits based on concentration, elongation, sphericity, and relaxedness. Consistent with our earlier results, we find that a split based on concentration produces the largest offset in the mean value of $z_{0.5}$ between the two sub-samples, while splits based on shape produce similar but smaller offsets. (These offsets will also be further reduced in practice, since for most systems we can determine only projected shape, not true three-dimensional shape.)

Relaxedness (which could be determined from detailed mapping of the mass distribution, e.g., in the X-ray) does not produce much of a shift in $z_{0.5}$ if we split the sample evenly, but it does select out systematically younger halos if we cut out the minority of very unrelaxed systems. The cut illustrated in Figure 13 is close to the one we have use throughout the paper to separate relaxed and unrelaxed systems. A similar selection by relaxedness (estimated from substructure in the gravitational lensing potential) has already been applied to a pilot sample of 10 clusters by Smith & Taylor (2008) in order to study the connection between X-ray morphology, offsets from X-ray scaling relations, and formation history; larger samples are needed to reach definitive conclusions. More recently, Krause et al. (2012) have shown that structural parameters determined from SZ measurements of simulated clusters should also be able to identify unrelaxed systems and thus reduce the scatter or flag outliers in scaling relations.

We can get a quantitative indication of the relative power of each different criterion by measuring the overlap in the distributions of systems split by various structural parameters. The exclusively young (blue), exclusively old (red), and overlapping regions of the distributions are 24%/30%/46% for a split based on concentration, 14/16/70 for a split based on elongation $E$, 14/15/71 for a split based on sphericity $c/a$, and 0/62/38 for a split based on relaxedness. In practice, structural splits would define two sub-samples with different mean values of $z_{0.5}$ and different widths; in the case of concentration $z_{0.5} \sim 0.7 \pm 0.3$ for the young sample and $z_{0.5} \sim 1.4 \pm 0.5$ for the old sample. (Note that the horizontal scale in these figures is linear in $z_{0.5}$; thus compared to Figure 10, the scatter in $z_{0.5}$ is larger in the old samples than in the young samples.)

6. SUMMARY

The recent studies of S11 and J11 have clarified how the structural properties of halos are inter-related, and how they relate to a few specific measures of the formation history. Here, we have extended this approach to measure the connection between structure and formation history in more detail. We first examine the intrinsic variation in halo formation histories, as summarized by their MAHs $M(z)$. Applying PCA to the individual steps of the MAHs themselves, we can decompose the range of individual formation histories into a mean MAH and a principal set of variations about the mean.

The strongest principal component, MAH-PC$_1$, accounts for almost 60% of the variance in the MAHs. Since variations along this axis correspond to MAHs that reach a given fraction of the final mass at systematically earlier or later times, it is natural to consider this a parameterization of halo age. In fact, it is the “best” definition of halo age based on the mass assembly history, in the formal sense that it is the definition that captures the most of the variance related to early or late formation. We provide a single-parameter analytic fit to variations along the MAH-PC$_1$ axis that is an excellent match to all but the 5% oldest trajectories. The mean MAH and variations away from it along the first principal axis MAH-PC$_1$ are also well fit by the more elegant two-parameter function suggested by McBride et al. (2009), although fitting two independent parameters $\beta$
Figure 13. Distributions of the formation redshift \( z_{0.5} \) in halo samples split by present-day structural properties.

(A color version of this figure is available in the online journal.)

and \( \gamma \) does not capture the one-dimensional nature of the age sequence.

A second component, MAH-PC2, is also relatively important in the sense that it accounts for 25% of the total variance in the MAHs. This component corresponds to acceleration or deceleration in the accretion rate at late times. It is not particularly well fit by the McBride formula or by our new formula for variation along the MAH-PC1 axis (unsurprisingly, since it describes a component orthogonal to it). Third and higher MAH-PCs correspond roughly to terms in a Fourier-like decomposition of the MAH, with an additional oscillation about the mean in each successive PC, but we have not studied them in detail since they are less important sources of scatter individually. Collectively, however, they account for almost 15% of the total variance. This residual power is presumably due to the large stochastic jumps in individual trajectories produced by major mergers.

Relating formation history and structural parameters, we recover the trends seen in the previous studies. Concentration, in particular, correlates strongly with age indicators, especially for older, relaxed systems. Testing the strength of the correlation against various age-related parameters, we find it is tightest for the formation redshift \( z_{0.5} \), the redshift by which a halo had built up 20% of its final mass in its most massive progenitor, or equivalently the redshift where \( M(z) = 0.2 \). This is close to the limit we can resolve in our merger trees; higher resolution studies might find strong correlations at even earlier times. In any case, clearly concentration is an indication of early assembly history.

Testing for correlations with other parameters, we find that elongation and relaxedness are correlated with the more recent formation history of a halo. Shape parameters such as elongation or sphericity are reasonably well correlated with the shape of the mass distribution from which the halo formed, but only going back to low redshifts, that is to say only considering a halo’s recent past. It seems likely that the formation process, rather than the geometry of the initial conditions, establish these correlations. Finally, irregularities in the density profile, measured using the parameter \( x_{\text{off}} \), also correlate slightly more strongly with the later part of a halo’s growth history.

These theoretical results have several possible applications to real systems. In the case of individual systems, concentration can give us a broad indication of dynamical age and early assembly history, while relaxedness or shape can give us an indication of later assembly history. As mentioned in the introduction, there are many possible ways of measuring these three parameters, particularly in rich clusters, including strong gravitational lensing in cluster cores, weak lensing at larger
radii, maps of X-ray emission, and even the galaxy distribution itself. Perhaps more interesting, however, are the implications for large samples of halos. Splitting cluster catalogs by concentration, shape, or relaxedness should select sub-samples with systematically different formation histories. Figure 13 shows that sub-samples defined by structural properties will always have considerable scatter in their individual histories, but their mean ages should differ significantly for reasonable-sized samples. Comparing these to predicted MAH distributions could offer new cosmological tests, updating an early idea of testing cosmology by measuring the dynamical ages of clusters as indicated by substructure (Richstone et al. 1992). We will consider the practical application of this idea in future work. On galaxy scales, if weak lensing surveys can reach the precision where halo shape is measured routinely, this may also allow a direct empirical test of how the MAH influences galaxy formation. Thus, the structural properties of individual dark matter halos will be incorporated into our broader understanding of cosmology and structure formation.

This work was supported by a Discovery Grant to J.T. and an Undergraduate Summer Research Award to A.W., both from the Natural Sciences and Engineering Research Council (NSERC) of Canada. We thank SHARCNET staff for their technical support throughout the project.

APPENDIX

CORRELATION TABLES FOR HALO PROPERTIES

The correlation analysis of halo structural properties is shown in Figure 14, for the full (relaxed + unrelaxed) sample. The values indicated on the axes are the base—10 log of the parameter in question, with mass measured in units of $10^{10} h^{-1} M_{\odot}$. Numbers in bold on each panel indicate the value of the Spearman rank correlation coefficient $\rho_{\text{src}}$.

REFERENCES

Allgood, B., Flores, R. A., Primack, J. R., et al. 2006, MNRAS, 367, 1781
Altay, G., Colberg, J. M., & Croft, R. A. C. 2006, MNRAS, 370, 1422
Avila-Reese, V., Colin, P., Gottlöber, S., Firmani, C., & Maubach, C. 2005, ApJ, 634, 51
Bailin, J., & Steinmetz, M. 2005, ApJ, 627, 647
Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, ApJ, 717, 379
Berlind, A. A., & Weinberg, D. H. 2002, ApJ, 575, 587
Bertschinger, E. 2001, ApJS, 137, 1
Bett, P., Eke, V., Frenk, C. S., et al. 2007, MNRAS, 376, 215
Broadhurst, T., Benitez, N., Coe, D., et al. 2005, ApJ, 621, 53
Broadhurst, T., Umetsu, K., Medezinski, E., Oguri, M., & Rephaeli, Y. 2008, ApJ, 685, L9
Bullock, J. S., Kolatt, T. S., Sigad, Y., et al. 2001, MNRAS, 321, 559
Cooray, A., & Sheth, R. 2002, Phys. Rep., 372, 1
Corless, V. L., King, L. J., & Clowe, D. 2009, MNRAS, 393, 1235

Figure 14. Correlation table of the parameters discussed in Section 2.6, for the full (relaxed + unrelaxed) sample. (A color version of this figure is available in the online journal.)
