Resolving Cosmic Structure Formation with the Millennium-II Simulation

Michael Boylan-Kolchin[1]⋆, Volker Springel[1], Simon D. M. White[1], Adrian Jenkins[2], and Gerard Lemson[3,4]

[1] *Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany*
[2] *Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK*
[3] *Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Moenchhofstr. 12-14, 69120 Heidelberg, Germany*
[4] *Max-Planck-Institut für extraterrestrische Physik, Giessenbach-Str. 1, 85748 Garching, Germany*

18 March 2009

ABSTRACT

We present the Millennium-II Simulation (MS-II), a very large N-body simulation of dark matter evolution in the concordance ΛCDM cosmology. The MS-II assumes the same cosmological parameters and uses the same particle number and output data structure as the original Millennium Simulation (MS), but was carried out in a periodic cube one-fifth the size ($100\,h^{-1}\,\mathrm{Mpc}$) with 5 times better spatial resolution (a Plummer equivalent softening of $1.0\,h^{-1}\,\mathrm{kpc}$) and with 125 times better mass resolution (a particle mass of $6.9 \times 10^6\,h^{-1}\,M_\odot$). By comparing results at MS and MS-II resolution, we demonstrate excellent convergence in dark matter statistics such as the halo mass function, the subhalo abundance distribution, the mass dependence of halo formation times, the linear and nonlinear autocorrelations and power spectra, and halo assembly bias. Together, the two simulations provide precise results for such statistics over an unprecedented range of scales, from halos similar to those hosting Local Group dwarf spheroidal galaxies to halos corresponding to the richest galaxy clusters. The "Milky Way" halos of the Aquarius Project were selected from a lower resolution version of the MS-II and were then resimulated at much higher resolution. As a result, they are present in the MS-II along with thousands of other similar mass halos. A comparison of their assembly histories in the MS-II and in resimulations of 1000 times better resolution shows detailed agreement over a growth factor of 100 in mass. We publicly release halo catalogs and assembly trees for the MS-II in the same format within the same archive as those already released for the MS.

Key words: methods: N-body simulations – cosmology: theory – galaxies: halos

1 INTRODUCTION

In order to understand how galaxies form and evolve in their cosmological context, we must understand the properties of dark matter halos over a wide range of physical scales and across virtually all of cosmic history. Numerical simulations provide one of the best methods for approaching this problem and have proven invaluable for studying the growth of cosmological structure and, in particular, of dark matter halos. Increasing computational power and improved algorithms have led to a steady and rapid increase in the ability of N-body simulations to resolve the detailed internal

structure of dark matter halos over substantial cosmological volumes.

Perhaps the most widely-used N-body simulation of cosmological structure formation to date has been the Millennium Simulation (Springel et al. 2005, hereafter MS), which followed more than ten billion particles within a simulation volume of $(500\,h^{-1}\,\mathrm{Mpc})^3$. This provided sufficient mass resolution to see the formation of halos hosting $0.1\,L_\star$ galaxies and sufficient volume to obtain good statistical samples of rare objects such as massive cluster halos and luminous quasars. It also enabled the implementation of physical models for the formation and evolution of galaxy/AGN populations throughout a large and representative cosmological volume (Croton et al. 2006; Bower et al. 2006). Since 2005, when the first results from the MS were published, most new

⋆ e-mail: mrbk@mpa-garching.mpg.de

very large cosmological simulations have focused on larger volumes[1] ($L_{\rm box} \gtrsim 1000\,h^{-1}\,{\rm Mpc}$) in order to study topics such as the statistical detection of baryon acoustic oscillations or weak lensing shear, or to build mock catalogs for the next generation of galaxy surveys (Teyssier et al. 2008; Kim et al. 2008). Moving to larger volume simulations reduces computational cost at fixed particle number both because resolved gravitational perturbations remain linear until later times and because the number of simulation particles in a typical nonlinear structure is smaller.

The opposite regime – smaller volumes with higher mass resolution – is much more computationally demanding but is also of great interest, especially for questions of galaxy formation, where the relevant mass scales are substantially smaller than for large-scale clustering. Understanding the formation and evolution of low-mass galaxies requires adequate resolution of the dark matter halos that host them, and this, in turn, requires *much* smaller particle masses than are currently feasible for Gigaparsec-scale simulations. These objects are important for galaxy formation as a whole because the first galaxies to form, which are low mass, prepare the initial conditions from which more massive systems later form. Another topic of particular interest that can be addressed by high-resolution simulations is the evolution of substructure within dark matter halos. Such simulations show that subhalos can lose considerable mass after being accreted onto a larger halo – sometimes well over 99% – without being completely disrupted (Hayashi et al. 2003; Gao et al. 2004b; Kravtsov et al. 2004). This means that as the resolution of a simulation is increased, so too is the typical time between accretion of a subhalo onto a larger system and its eventual tidal disruption.

Moving to substantially higher resolution in a large-volume simulation is fraught with computational challenges, however. Increasingly small simulation time-steps are required to accurately follow particle orbits in the dense centers of dark matter halos (Power et al. 2003), where the characteristic time-scale $t_{\rm grav} \propto 1/\sqrt{G\rho}$ is significantly shorter than on large scales. While only a small fraction of simulation particles reside in such dense regions, these particles are the limiting factor in how quickly the simulation can be evolved forward. The maximum resolved density contrasts at $z \lesssim 1$ can be one thousand times higher than those at $z \approx 6$; as a result, almost all of the computational time needed for such a simulation is spent at low redshift. Furthermore, the strong clustering of matter within a few very massive clumps can create serious problems with respect to parallelization: it is much more difficult to split such a particle distribution into optimal computational domains than is the case if the matter distribution is more homogeneous.

In spite of these challenges, it is essential to have simulations that probe the structure of galaxy-scale dark matter halos with high mass resolution and over a large enough region to include a sizable and representative sample of objects. In this paper, we present such a calculation, the Millennium-II Simulation (hereafter MS-II). Section 2 gives details of the parameters which define this simulation and

describes some of the post-processing we have carried out on its output, in particular, substructure-finding and merger tree-building. We present results on the evolution of the dark matter power spectrum and the two-point correlation function in Section 3. In Section 4, we investigate the dark matter halo mass function and the clustering bias of dark matter halos. Section 5 focuses on halo formation times, including the dependence of clustering on formation time (so-called "assembly bias"; Gao et al. 2005). A discussion of the relation between the MS-II and the Aquarius Project (Springel et al. 2008), as well as a comparison of the assembly histories of the halos common to the two projects, is presented in Section 6. We summarize our results in Section 7. Throughout this paper, all logarithms without specified bases are natural logarithms.

2   THE MILLENNIUM-II SIMULATION

2.1   Simulation details

The Millennium-II Simulation follows 2160^3 particles within a cubic simulation box of side length $L_{\rm box} = 100\,h^{-1}\,{\rm Mpc}$. This is five times smaller than $L_{\rm box}$ for the Millennium Simulation. The volume sampled by the MS-II is thus 125 times smaller than in the MS but the mass resolution is correspondingly 125 times better: each simulation particle has mass $6.885 \times 10^6\,h^{-1}\,M_\odot$. With this mass resolution, halos similar to those hosting Local Group dwarf spheroidals are resolved at our 20 particle mass limit, while halos of Milky Way-mass galaxies have hundreds of thousands of particles and halos of rich clusters have over fifty million particles. The Plummer-equivalent force softening adopted for the MS-II was $1\,h^{-1}\,{\rm kpc}$ and was kept constant in comoving units; this value corresponds to 0.06% (10%) of the virial radius for the largest (smallest) halos at redshift zero.

The ΛCDM cosmology used for the MS-II is identical to that of the MS and the Aquarius simulations:

$$\Omega_{\rm tot} = 1.0,\ \Omega_m = 0.25,\ \Omega_b = 0.045,\ \Omega_\Lambda = 0.75,$$
$$h = 0.73,\ \sigma_8 = 0.9,\ n_s = 1\,, \tag{1}$$

where h is the Hubble constant at redshift zero in units of $100\,{\rm km\,s^{-1}\,Mpc^{-1}}$, σ_8 is the rms amplitude of linear mass fluctuations in $8\,h^{-1}\,{\rm Mpc}$ spheres at $z = 0$, and n_s is the spectral index of the primordial power spectrum. Retaining the cosmological parameters of the MS allows us to test for convergence by comparing results in the regime where objects are well-resolved in both simulations as well as to extend the range of structures probed by combining, when appropriate, results from the two simulations. In particular, this helps us understand the effects of resolution at low particle number.

The initial conditions for the simulation were created at redshift $z = 127$ using a "glass" initial particle load (White 1996); the initial particle positions and velocities were then computed using the displacement field tabulated on a 4096^3 mesh and the Zeldovich approximation. The transfer function used for calculating the input linear power spectrum was computed with the Boltzmann code `CMBFAST` (Seljak & Zaldarriaga 1996).

The amplitudes and phases of the initial linear fluctuation modes in the MS-II are identical to those in the simu-

---

[1] Recent simulations with $\gtrsim 10^{10}$ particles within smaller volumes ($L_{\rm box} \approx 100\,h^{-1}\,{\rm Mpc}$) have been used primarily for studying cosmic reionization at redshifts $\gtrsim 6$ (e.g., Iliev et al. 2006).

Name	**L_{box}** $[h^{-1}\,\mathrm{Mpc}]$	**N_p**	**ϵ** $[h^{-1}\,\mathrm{kpc}]$	**m_p** $[h^{-1}\,M_\odot]$	**M_{min}** $[h^{-1}\,M_\odot]$	**M_{max}** $[h^{-1}\,M_\odot]$	**f_{group}**
Millennium-II	100	10,077,696,000	1.0	6.89×10^6	1.38×10^8	8.22×10^{14}	0.601
Millennium	500	10,077,696,000	5.0	8.61×10^8	1.72×10^{10}	3.77×10^{15}	0.496
mini-MS-II	100	80,621,568	5.0	8.61×10^8	1.72×10^{10}	8.29×10^{14}	0.502

Table 1. Some basic properties of the new Millennium-II Simulation are compared to those of the MS and to the lower resolution version of MS-II (mini-MS-II). L_{box} is the side length of the simulation box, N_p is the total number of simulation particles used, and ϵ is the Plummer-equivalent force softening of the simulation, in comoving units. m_p gives the mass of each simulation particle, M_{min} gives the mass of the smallest FOF halos (corresponding to our choice of storing all halos with $N_p \geq 20$), and M_{max} gives the maximum FOF halo mass found in the simulation. f_{group} is the fraction of all simulation particles in FOF groups of 20 or more particles at $z = 0$.

lation from which the Aquarius Project halos were chosen. Specifically, all modes with a wave-vector whose maximum component is less than $13.57\,h\,\mathrm{Mpc}^{-1}$ have amplitudes and phases that match those of the Aquarius simulations; all other modes were set at random to have the same underlying power spectrum. The Aquarius halos are thus present in the MS-II. A discussion of the relationship between the Aquarius Project and the MS-II is presented in Section 6.

The MS-II was run with `GADGET-3`, an updated version of the `GADGET` code (Springel et al. 2001b; Springel 2005). `GADGET-3` is a TreePM code: long-range force calculations are performed with a particle-mesh algorithm while short-range forces are calculated via a hierarchical tree. While the original MS was performed with a memory-optimized version of `GADGET-2`, the extremely high level of clustering that occurs in the MS-II results in somewhat different computational requirements; in particular, a more flexible domain decomposition is necessary. `GADGET-3` was developed specifically for this situation.

The MS-II was performed on the IBM Power-6 computer at the Max-Planck Computing Center in Garching, Germany, using 2048 cores and approximately 8 TB of memory. A Fast Fourier Transform with 4096^3 cells was used for the PM calculation. Particles were allowed to have individual, adaptive time-steps. The evolution of the simulation required approximately 1.4 million CPU hours and 2.77×10^{13} force calculations for the 22,142 simulation time-steps. The mid-point of the simulation in terms of computational time was $z = 0.88$; by contrast, evolving the simulation from $z = 127$ to $z = 6$ took only 10% of the total CPU time.

Outputs were saved at sixty-eight epochs: sixty-five snapshots spaced according to

$$\log_{10}(1 + z_N) = \frac{N\,(N + 35)}{4200} \quad (0 \leq N \leq 64) \qquad (2)$$

and three high redshift outputs at $z = 40$, 80, and 127. The spacing scheme in Equation (2) is identical to that used in the MS. We have extended the range of regularly-spaced outputs to $z \approx 31.3$, however, because the increased mass resolution of the MS-II results in earlier-forming first structures.

For comparison purposes, we have also performed a version of the MS-II with identical initial conditions and in the same volume with the same outputs as the main run but at the same mass and force resolution as the original MS (so $N_p = 432^3$). This "mini-MS-II" simulation allows us to test how numerical resolution affects our results. Some basic details of all three simulations are listed in Table 1.

2.2 Halos and Subhalos

Dark matter halos were identified on-the-fly during the simulation for each snapshot using the friends-of-friends (FOF) algorithm (Davis et al. 1985) with a linking length of $b = 0.2$; all groups with at least 20 particles were retained for later analysis. This process resulted in 1.17×10^7 FOF groups at $z = 0$, slightly fewer than the peak value of 1.53×10^7 at $z = 3.06$. Just over 60% of the particles in the full simulation belong to a FOF group at $z = 0$. A catalog with quantities of interest for each FOF halo (e.g., position, velocity, number of particles), as well as a list of the particles composing each halo, was saved at each snapshot.

The largest FOF group at $z = 0$, a cluster-mass dark matter halo, has over 119 million particles. Figure 1 shows images of the dark matter distribution in the MS-II on a number of different physical scales, all centered on this halo[2]. The large panel in the upper left shows a $15\,h^{-1}\,\mathrm{Mpc}$-thick slice through the full simulation volume ($100\,h^{-1}\,\mathrm{Mpc}$ on each side). The well-known cosmic web of filaments and voids can be seen clearly. Starting in upper-right and moving clockwise, the other five panels zoom successively closer into the halo. The bottom-right panel is $5\,h^{-1}\,\mathrm{Mpc}$ on a side, approximately the diameter of the halo. As has been long known (e.g., Moore et al. 1998; Tormen et al. 1998; Ghigna et al. 1998; Klypin et al. 1999a,b; Moore et al. 1999), FOF halos in ΛCDM simulations are not monolithic objects but rather are teeming with substructure; this substructure is clearly evident even at $1/10^{\mathrm{th}}$ the radius of the halo (lower-left panel).

During post-processing, every FOF halo was searched for bound dark matter substructure using the `SUBFIND` algorithm (Springel et al. 2001a). `SUBFIND` identifies substructures within a FOF halo by searching for overdense regions using a local SPH density estimate, identifying substructure candidates as regions bounded by an isodensity surface that traverses a saddle point of the density field, and testing that these potential substructures are physically bound with an iterative unbinding procedure. All self-bound structures with at least 20 particles were deemed to be physical *subhalos* and were stored in subhalo catalogs. Several properties of each subhalo were also tabulated and saved, including velocity dispersion, peak circular velocity V_{max} and the radius R_{max} at which V_{max} is attained, half-mass radius, spin, position, and velocity. The member particles of each subhalo

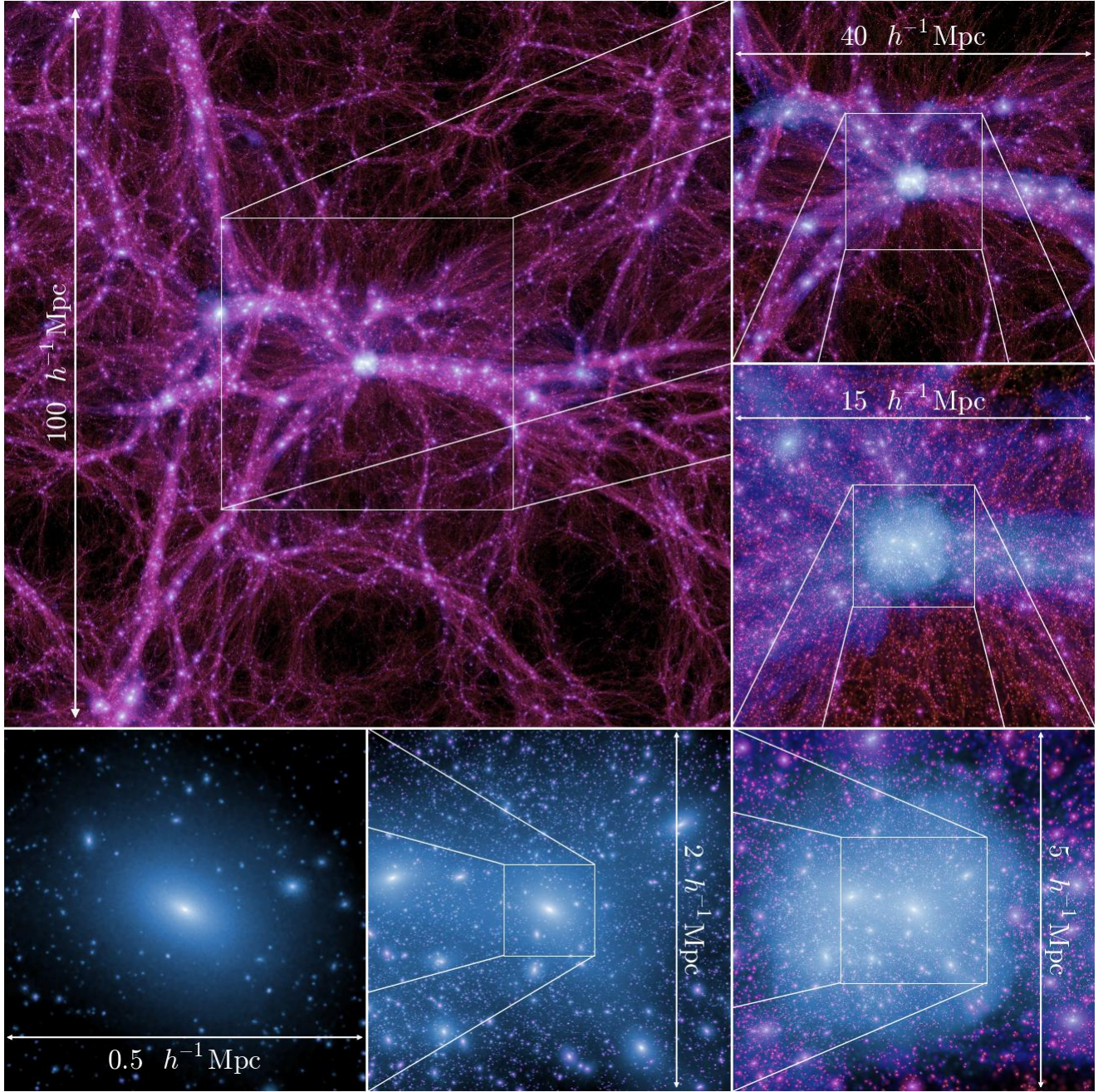[2] Images of individual panels and additional information related to the MS-II are available at
http://www.mpa-garching.mpg.de/galform/millennium-II

Figure 1. A sequential zoom through the Millennium-II Simulation. The large image (upper left) is a $15\,h^{-1}$ Mpc thick slice through the full $100\,h^{-1}$ Mpc simulation box at redshift zero, centered on the most massive halo in the simulation. This FOF halo has $M_{\mathrm{FOF}} = 8.2 \times 10^{14}\,h^{-1}\,M_{\odot}$, similar to the mass of the Coma cluster (Colless & Dunn 1996), is composed of 119.5 million particles, and contains approximately 36,000 resolved subhalos spanning 6.7 decades in mass. Starting from the upper right and moving clockwise, subsequent panels zoom into the cluster region and show slices that are 40, 15, 5, 2, and $0.5\,h^{-1}$ Mpc on a side (with thicknesses of 10, 6, 5, 2, and $0.5\,h^{-1}$ Mpc). Even at $0.5\,h^{-1}$ Mpc, which is approximately $1/10^{\mathrm{th}}$ the diameter of the halo, a rich variety of substructure is visible.
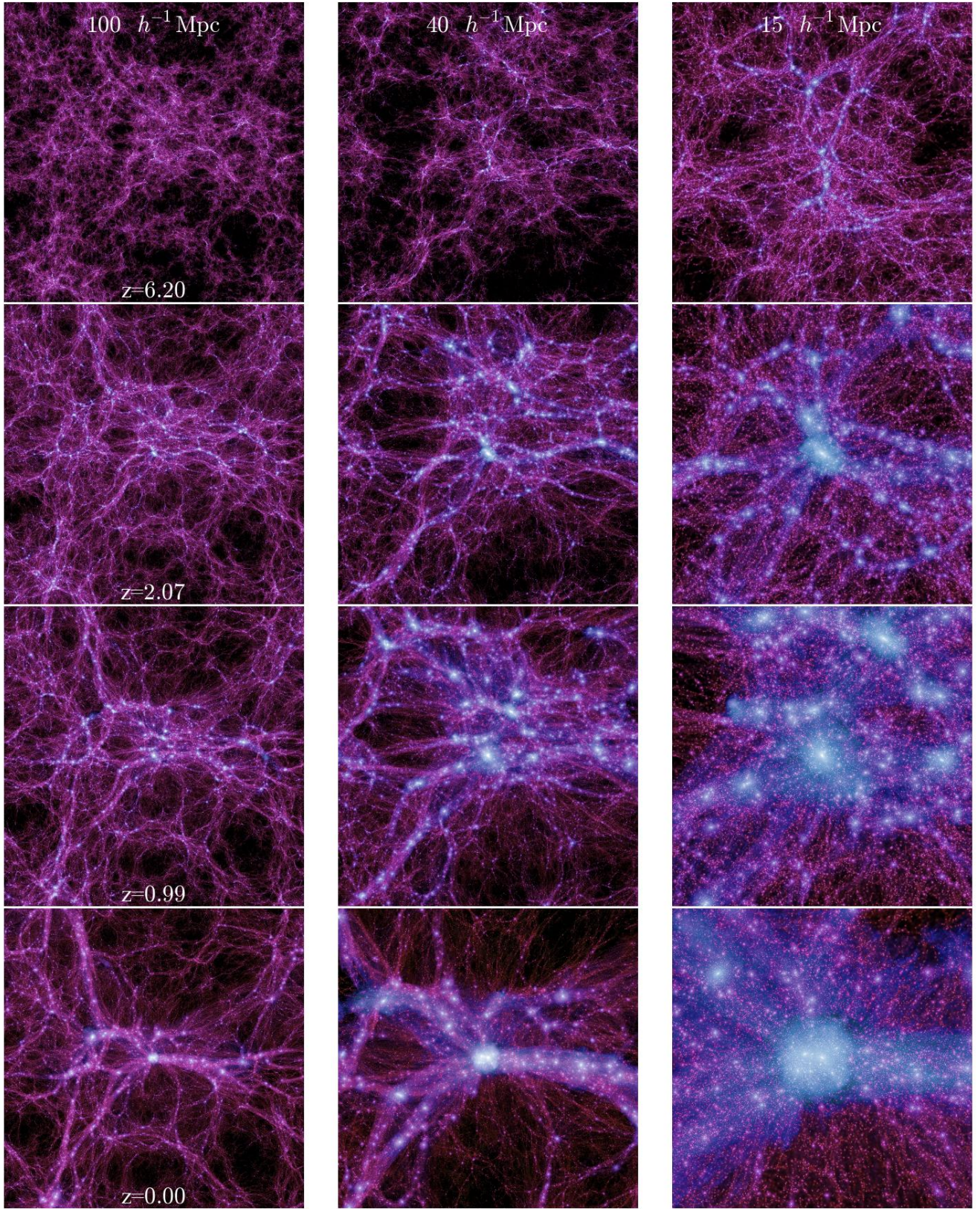
Figure 2. Time evolution of the largest FOF halo at $z = 0$ in the Millennium-II Simulation. The halo is shown at three co-moving scales (from left to right: 100, 40, and 15 h^{-1} Mpc, with thickness 15, 10, and 6 h^{-1} Mpc) and at four different cosmological epochs (from top to bottom: z=6.2, 2.07, 0.99, and 0).

were ranked according to binding energy and stored in that order, which facilitates tracking subhalos across simulation outputs. Note that with these procedures, we have two separate but related sets of dark matter structures: FOF halos and subhalos.

While each subhalo has a single well-defined mass assigned to it – the sum of the masses of its constituent particles – multiple mass definitions for FOF halos are common in the literature (see White 2001 for a discussion of subtleties associated with assigning masses to halos). The most straightforward definition is M_{FOF}, the total mass of all the member particles. Another possibility is M_{Δ}, defined as the mass contained in a spherical region (centered on the particle in the dominant subhalo with the minimum gravitational potential) with average density a factor Δ larger than the critical density of the universe. For each FOF halo, we calculated M_{Δ} for $\Delta = 200$, $200\,\Omega_m(z)$, and $\Delta_v(z)$, where the last value is taken from the spherical top-hat collapse model (see, e.g., Bryan & Norman 1998). We refer to spherical overdensity masses as M_{200} [$\Delta = 200$], M_{200m} [$\Delta = 200\,\Omega_m(z)$], or M_v [$\Delta = \Delta_v(z)$] and to the corresponding virial radii as R_{200}, R_{200m}, or R_v. At high redshifts, when the matter density is nearly equal to the critical density, all three definitions give similar masses. At lower redshifts, $M_{200m} > M_v > M_{200}$ for a given halo.

2.3   Merger Trees

Merger trees were constructed at the subhalo level by requiring subhalos to have at most one descendant. For many subhalos, this descendant can be found trivially (if it exists): all particles in a subhalo at snapshot S_n may belong to a single subhalo at the subsequent snapshot S_{n+1}, in which case this subhalo is clearly the descendant of the subhalo at the previous snapshot. There is also the possibility that particles belonging to one subhalo at S_n may be distributed over more than one subhalo at S_{n+1}. We still require each subhalo to have at most one descendant for these cases, so a subhalo's unique descendant is identified as follows. First, the binding energy of each particle in the subhalo at S_n is calculated and the particles are ranked by this binding energy. Each potential descendant subhalo – that is, each subhalo at S_{n+1} containing at least one particle j from the subhalo at S_n – is then given a score χ that is based on the binding energy rank $\mathcal{R}$ of these particles: $\chi = \sum_j \mathcal{R}_j^{-2/3}$. The subhalo at S_{n+1} with the largest value of χ is defined to be the descendant subhalo[3]. Note that while descendants are unique, a given subhalo may have many progenitors.

There is one slight complication to this process. Sometimes a subhalo passing through the dense center of a larger system will not be identified by SUBFIND, simply because the density contrast is not high enough. To mitigate this problem, we also search for a descendant at snapshot S_{n+2}. In the vast majority of cases, however, the descendant of a subhalo is found at S_{n+1}.

Once all unique descendants are found, the subhalos are linked across all snapshots to form merger trees. This is done by taking a subhalo at $z = 0$ and linking all subhalos

with descendant pointers to this halo, then repeating with all of those subhalos, and so on, until no more subhalos can be joined. This process results in links between most, though not all, of the subhalos in the simulation: subhalos that are never connected to any $z = 0$ subhalo and that are never connected to any progenitor of any $z = 0$ subhalo are not included in the trees. We save several pointers for each tree subhalo for later use. These include pointers to the dominant subhalo of the subhalo's FOF group, the next most massive subhalo in the FOF group, the progenitor that contains the largest fraction of the subhalo's particles, the subhalo's descendant, and the next most massive subhalo that shares the same descendant[4].

The merger trees for the MS-II contain approximately 590 million subhalos in total (as compared to 760 million subhalos in the MS). While the overall data volume of the MS-II is similar to that of the MS (≈ 25 Terabytes, dominated by the raw particle data), the highly clustered nature of the MS-II means that the trees are markedly less homogeneous. There are only half as many trees in total in the MS-II as in the MS, but the largest tree is *much* larger, with over 90 million subhalos (compared to 500 thousand for the largest tree in the MS).

2.4   An example of subhalo tracking

As an example of our merger trees and subhalo tracking, we consider the main progenitor histories of the most massive halo at $z = 0$ and of two of its subhalos. We use both the MS-II and mini-MS-II in order to highlight the effects of resolution and to probe the convergence of the subhalo identification and merger tree building algorithms. Objects can be matched between the simulations because the initial conditions are identical on all scales that overlap; any differences are due to force and mass resolution and to differing discreteness effects.

The main cluster halo is trivial to find in both simulations. At $z = 0$, the properties agree quite well between the two: the FOF masses are the same to within 1% ($M_{\mathrm{FOF}} = 8.22 \times 10^{15}\,h^{-1}\,M_{\odot}$ for MS-II versus $8.29 \times 10^{15}\,h^{-1}\,M_{\odot}$ for mini-MS-II) and even the position of the halo, as determined by the gravitational potential minimum, agrees to with $0.022\,h^{-1}\,\mathrm{Mpc} = 4\,\epsilon$ (for mini-MS-II). Figure 2 shows the main progenitor of this halo on three co-moving scales (from left to right: 100, 40, and 15 $h^{-1}\,\mathrm{Mpc}$) and at four redshifts (from top to bottom: $z = 6.2$, 2.07, 0.99, and 0.0). Using the merger trees, we track the main progenitor of the most massive cluster back in time for each run until there are no further progenitors. The top two curves in Figure 3 show the mass of the central subhalo of this main progenitor branch[5] for the MS-II (black) and mini-MS-II (magenta). The two are in excellent agreement from $z = 0$ to $z = 9$,

---

[3] This procedure weights the most bound regions of a subhalo most heavily when determining its descendant.

[4] see also figure 5 in the Supplementary Information of Springel et al. (2005)

[5] Note that the while the FOF masses of the halos agree to within 1% between the MS-II and mini-MS-II, the mass of the central subhalo in the MS-II is slightly smaller because more distinct subhalos are identifiable. Many subhalos that are resolvable in the MS-II but not in mini-MS-II show up as extra mass in the central subhalo in mini-MS-II.
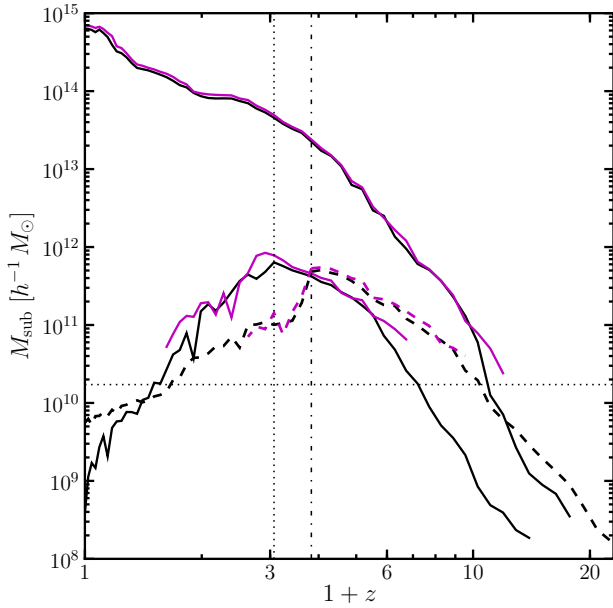
Figure 3. Subhalo assembly histories in the MS-II and mini-MS-II. The upper set of curves shows the mass in the main progenitor branch of the most massive $z = 0$ FOF group for the MS-II (black) and mini-MS-II (magenta), while the lower set of curves shows main progenitor histories for subhalos that are accreted onto the FOF group of the cluster at $z \approx 2$ (solid lines) and $z \approx 3$ (dashed lines). The horizontal dotted line shows the mass resolution of MS and mini-MS-II, while the vertical dotted and dot-dashed lines show the epochs at which the smaller halos joined the FOF group of the main cluster in the MS-II.

Figure 4. The mean cumulative subhalo abundance per parent halo in the MS-II (solid curves) and mini-MS-II (dashed curves) in five parent halo mass bins. The curves for each simulation are plotted down to the 20 particle resolution limit. There is a deficit in subhalos at this limit in mini-MS-II relative to the MS-II due to resolution. At 3-5 times the minimum resolution limit, however, the two simulations agree very well, indicating that subhalos with more than 100 particles are reliably resolved.

at which point the main branch from mini-MS-II falls below 100 particles and resolution effects become relevant (the main progenitor in the MS-II can be traced back all the way to $z \approx 18$). Clearly, the assembly of the main progenitor is very well converged between the two runs at $z \lesssim 9$.

We also consider the evolution of two far less massive subhalos within this FOF group: subhalo A (lower solid lines in Figure 3) and subhalo B (dashed lines) are identified in the MS-II at $z = 0$. They are not massive enough to be identified at $z = 0$ in mini-MS-II, as subhalo A has 57 particles and subhalo B has 812 at the final snapshot (the resolution limit of mini-MS-II corresponds to 2500 particles from the MS-II and is marked by the horizontal dotted line). Nevertheless, Figure 3 shows that by tracking subhalos A and B backward (lower black curves), we find that both were much more massive in the past: each exceeded $5 \times 10^{11} \, h^{-1} \, M_\odot$ at one point in its history (92,431 particles for A and 73,718 for B) and each was the central subhalo of its own FOF group before falling into the main progenitor of the cluster (this accretion happened at $z \approx 2$ for subhalo A, marked by the vertical dotted line, and at $z \approx 3$ for subhalo B, marked by the vertical dot-dashed line)[6].

These maximum masses for A and B are easily resolvable in mini-MS-II, so we can hope to find the equivalent subhalos there and compare their mass histories. It is in general unrealistic to expect subhalos to have identical positions in runs of differing resolution: the gravitational force is softened at different scales and there is a difference in the 'graininess' of the gravitational potential due to finite particle mass, both of which can cause orbital phase offsets that accumulate over time[7]. Nevertheless, we are able to locate A and B in mini-MS-II and we find them to be at almost exactly the same positions as in the MS-II: the absolute positions for subhalo A (subhalo B) differ by only 0.015 (0.010) h^{-1} Mpc at the times marked by the vertical lines, which is only three times the force softening of the low resolution run. Having located A and B in mini-MS-II, we then use the merger trees to track the subhalos both forward and backward in time and we compare to the results from the MS-II.

These subhalo mass histories are shown in the lower magenta lines (solid for subhalo A, dashed for subhalo B) of Figure 3. Comparing the results from the MS-II and mini-MS-II, we see that the subhalos can be tracked remarkably well not only when they are the main FOF subhalo (to the right of the vertical lines) but also when they are non-dominant

---

[6] A and B are two of four subhalos in the main FOF cluster at $z = 0$ that (i) have $N_p(z = 0) < 1000$ and (ii) have over 1500 progenitor subhalos in their sub-tree. This selection picks out halos that were massive at one point in their history but are not at $z = 0$. We have also investigated the other two subhalos from this sample and find similar convergence in the subhalo tracking
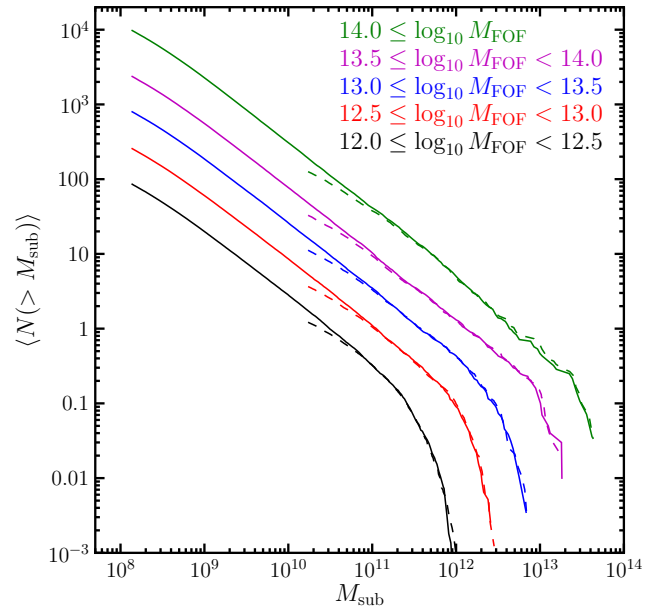
and mass identification between the two simulations; for clarity, the results are not plotted in Figure 3.

[7] See Springel et al. (2008) for a method to match subhalos in simulations based on the positions of the subhalo particles in the initial conditions.

subhalos within a larger FOF group (to the left of the vertical lines). This regime, where the subhalos are subjected to strong tides that vary rapidly in time, can be extremely difficult to capture accurately in simulations of differing resolution. The excellent agreement between the lower magenta and black curves demonstrates that the subhalos have the same dynamical histories, are assigned the same masses, and are linked in the same way by the merger trees in the two runs.

Figure 3 also illustrates resolution limitations. As tides strip mass from the subhalos, they are lost from the mini-MS-II catalog and are considered to have merged with the dominant subhalo, while they persist as independent subhalos to $z = 0$ at the significantly enhanced resolution of the MS-II. The high resolution of the MS-II is required to study the fates of subhalos hosting low mass galaxies within larger structures: note that the maximum masses – approximately $6 \times 10^{11} \, h^{-1} \, M_\odot$ – of A and B are quite large, larger than the halo masses of likely Milky Way progenitors at the redshift of accretion into the massive halo. A and B are therefore likely to host galaxies of stellar mass comparable to that of the Milky Way. At mini-MS-II resolution (i.e. MS resolution) it is not possible to follow the dynamics of these subhalos past $z \approx 1.5$ for subhalo B or $z \approx 0.5$ for subhalo A, by which redshifts their masses have dropped below $10^{11} \, h^{-1} \, M_\odot$. The MS-II captures the later dynamical history of both subhalos, even though subhalo A (subhalo B) retains only 0.06% (1.1%) of its maximum mass at $z = 0$. Note that these final masses are considerably smaller than the likely *stellar* masses of the associated galaxies, so it remains unclear how realistic the late-time dynamical evolution actually is in MS-II.

We can also consider the statistical agreement between the MS-II and mini-MS-II by comparing stacked subhalo abundances as a function of host halo mass. Figure 4 shows the mean number of subhalos per host halo in five host halo mass bins for the two simulations. Resolution effects reduce the number of subhalos at a given subhalo mass in mini-MS-II (dashed curves) relative to the MS-II (solid curves) for subhalos with few particles: at the minimum resolvable mass of 20 particles in mini-MS-II, the abundance of subhalos is reduced by approximately 30% relative to the MS-II. The results from the two simulations are in excellent agreement for more massive subhalos ($M_{\rm sub} > 10^{11} \, h^{-1} \, M_\odot$), showing that subhalos containing at least 50-100 particles are reliably resolved.

3 STATISTICS OF THE DENSITY FIELD

3.1 Power Spectrum

At comoving position $\mathbf{x}$ and time t, the mass density field can be expressed as

$$\rho(\mathbf{x}, t) = \bar{\rho}(t) \left[1 + \delta(\mathbf{x}, t) \right]. \qquad (3)$$

In the standard picture of structure formation in a cold dark matter universe, the initial density fluctuation field $\delta(x, 0)$ is taken to be a Gaussian random field. Its statistical properties are therefore fully specified by its power spectrum $P(k)$ or equivalently its dimensionless power spectrum $\Delta^2(k)$,

$$\Delta^2(k) \equiv \frac{k^3}{2\pi^2} \, P(k). \qquad (4)$$
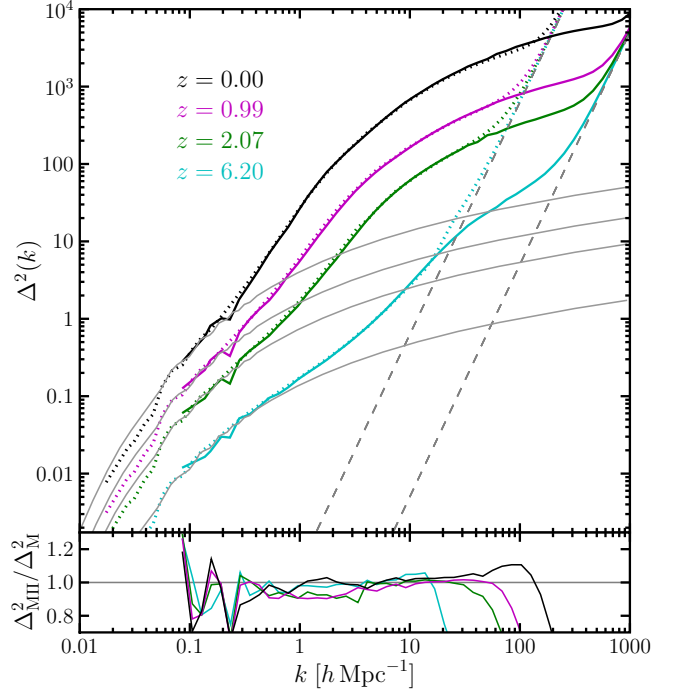

Figure 5. The power spectrum $\Delta^2(k)$ measured from the MS-II at redshifts 0.0 (black curves), 0.99 (magenta), 2.07 (green), and 6.20 (cyan), as well as the linear theory power spectrum at each redshift (gray curves). Power spectra from the MS (dotted curves) at the same redshifts are also shown for comparison. The dashed lines correspond to the shot noise limit for the MS-II (right) and the MS (left); the power spectra have not been corrected for shot noise. The bottom panel shows the ratio of the power spectra.

$\Delta^2(k)$ measures the power per logarithmic interval in wavenumber; $\Delta^2(k) \approx 1$ therefore indicates that fluctuations in density on scales near wavenumber k are of order unity.

The dark matter power spectrum from the MS-II is shown in Figure 5. We plot the results at four redshifts: $z = 0$ (black curves), 0.99 (magenta), 2.07 (green), and 6.20 (cyan). On large physical scales (small wavenumber k), the power spectrum follows the prediction from linear theory (light gray lines). As time progresses, larger and larger physical scales become non-linear and the small-scale power exceeds the linear theory prediction. Results from the MS are also included in Figure 5 for comparison (dotted curves). The agreement between the two simulations is very good, and the MS-II extends the measurement of $\Delta^2(k)$ by a factor of 5 at large k. We have not performed a shot noise correction in this figure, as it not clear that it is appropriate to do so (see, e.g., Baugh et al. 1995; Sirko 2005). The shot noise limit for each run is plotted as a dashed gray line.

The large dynamic range and uniform mass resolution of the MS-II allows us to probe the dark matter power spectrum on a wide range of scales, including scales where existing fitting functions (Peacock & Dodds 1996; Smith et al. 2003) are uncalibrated and untested. Figure 6 compares the power spectrum computed from the MS-II with the halo model predictions of Smith et al. (gray lines). At redshift 2 and below, the Smith et al. model agrees with the calculated power spectrum to within 10% for $k < 5 \, h \, {\rm Mpc}^{-1}$.
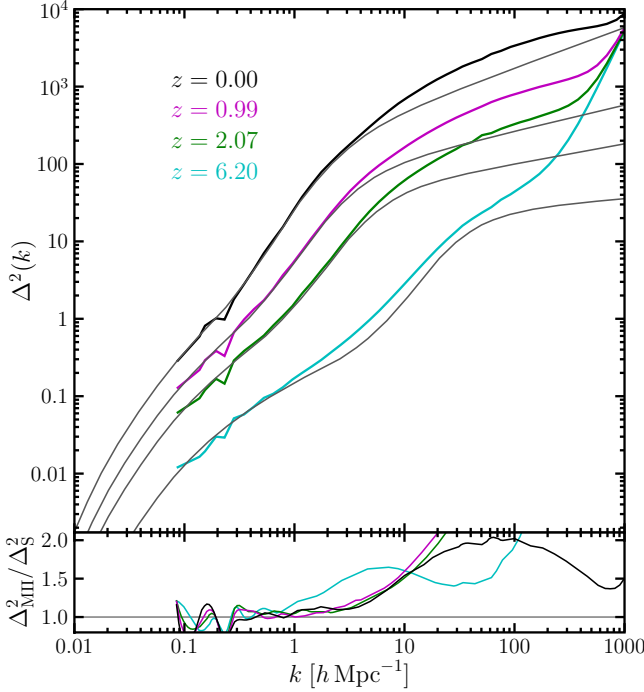
Figure 6. A comparison of the MS-II power spectra at four redshifts to the halo model fit from Smith et al. (2003; gray curves). The fit at $z = 0$ is accurate for $k \lesssim 7\,h\,\mathrm{Mpc}^{-1}$ but underestimates the power spectrum from the MS-II by 50% or more for k between 10 and 100 h Mpc^{-1}. Redshifts one and two show similar results but the agreement at $z = 6$ is poor.
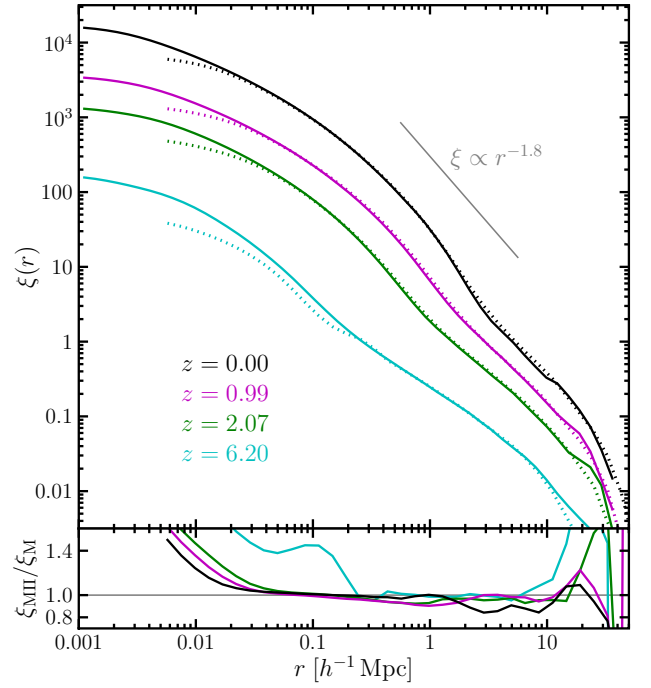
Figure 7. Measurements of the two point correlation function ξ as a function of comoving separation r from the MS-II. We show four redshifts: $z = 0.0$ (black curves), 0.99 (magenta), 2.07 (green), and 6.20 (cyan) and we compare with $\xi(r)$ from the MS (dotted curves) at the same redshifts. On large scales, the correlation functions from the two simulations agree quite well. On small scales ($\lesssim 0.020\,h^{-1}$ Mpc in physical units), the MS-II correlation function amplitude is larger, reflecting structures that are not resolved in the MS. The bottom panel focuses on these differences by plotting the ratio of the correlation function from the MS-II to that from the MS.

At larger k, however, the model significantly underestimates the power, with errors exceeding a factor of 2. The Smith et al. model was not calibrated for this range, so it is not surprising that it fails to reproduce the simulation results. Nevertheless, this is a reminder that extrapolating fitting functions beyond their calibrated range can lead to serious errors. At $z = 6$, the Smith et al. model fails to fit the data over the range $1 < k < 10\,h\,\mathrm{Mpc}^{-1}$, where the MS-II and MS agree very well.

At scales of $k \gtrsim 10\,h\,\mathrm{Mpc}^{-1}$, baryonic physics plays an important role in determining the real dark matter power spectrum (e.g., Rudd et al. 2008). Although a full modeling of baryonic effects will be necessary to get accurate predictions for Δ^2 at these scales, understanding the underlying dark matter-only power spectrum still provides a critical baseline.

3.2 Two Point Correlation Function

The spatial two-point correlation function of the density field is given by

$$\xi(r) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle \,, \tag{5}$$

or equivalently, by a Fourier transform of the power spectrum:

$$\xi(r) = \int \Delta^2(k) \frac{\sin(kr)}{kr} \, d\log k \,. \tag{6}$$

The correlation function is a useful measure of the spatial clustering of dark matter: it gives the excess probability of

finding pairs of particles at a given separation relative to a Poisson distribution. Figure 7 shows $\xi(r)$ at redshifts 0, 0.99, 2.07, and 6.20, with results from the MS at the same redshifts also plotted for comparison. (Note that the scale on the horizontal axis of Figure 7 is in co-moving units.)

The correlation function shows a prominent feature at $r \approx 2\,h^{-1}$ Mpc (for $z = 0$). This is the well-known transition between the 'one-halo' and 'two-halo' contributions: on smaller scales the correlation function is dominated by dark matter particle pairs within the same halo, while on larger scales it is dominated by pairs in separate halos (Peacock & Smith 2000; Seljak 2000; Ma & Fry 2000; Cooray & Sheth 2002). At no redshift is the correlation function even roughly approximated by a single power law. This is in stark contrast to observations of galaxy correlation functions (Zehavi et al. 2002; Hawkins et al. 2003) and the stellar mass autocorrelation function (Li & White 2009) at low redshift, which show a remarkably good power-law behavior $\xi \propto r^{-1.8}$ over $10^{-2} \lesssim r \lesssim 10\,h^{-1}$ Mpc (a gray line with this relation is also plotted in Figure 7 for comparison).

From $z = 2$ to $z = 0$, the agreement between the MS-II and MS results is quite good from 0.03 to $2\,h^{-1}$ Mpc. The MS-II result lies approximately 10% below the MS $\xi(r)$ on scales of 2-10 h^{-1} Mpc; this is the range where the contributions from pairs in separate halos become important and is

likely an indicator of cosmic variance in the two-halo term. On small scales, the MS-II correlation function lies above that of the MS. This is a result of the higher force and mass resolution of the MS-II: low-mass halos that were not resolvable in the MS also boost the clustering on small scales. The MS correlation function is noticeably lower in amplitude than $\xi(r)$ from the MS-II at $z = 6$ for $r \lesssim 0.2\,h^{-1}$ Mpc. This coincides with the mean interparticle spacing, $0.231\,h^{-1}$ Mpc comoving for the MS, and is therefore most likely due to discreteness in the glass-like particle load used for the initial conditions.

4　DARK MATTER HALOS

Understanding how dark matter overdensities grow and ultimately virialize into highly non-linear structures is an extremely difficult problem from a theoretical perspective. No rigorous analytic techniques are available for use in both the linear and non-linear regimes. The most successful model for dark matter halo formation (Press & Schechter 1974; Bond et al. 1991; Lacey & Cole 1993; Sheth et al. 2001; see Zentner 2007 for a recent review of extended Press-Schechter theory) relates the abundance of halos at mass M and redshift z to the initial power spectrum of density fluctuations and to the well-understood regime of linear growth. It considers the linear overdensity field smoothed using a spherical top-hat filter (in real space) containing mass M and extrapolated using linear theory to redshift z. The variance of this smoothed field is:

$$\sigma^2(M, z) = d^2(z) \int \Delta_{\rm lin}^2(k)\, W^2(k; M)\, d\log k \;, \qquad (7)$$

where $d(z)$ is the linear growth factor at redshift z with normalization $d(z=0)=1$, $\Delta_{\rm lin}^2(k)$ is the linear power spectrum extrapolated to $z = 0$, and $W(k; M)$ is the Fourier transform of a real-space spherical top-hat filter. Each dark matter particle is assigned to a halo of mass M at redshift z, where M is taken to be the largest filter scale for which the smoothed linear overdensity at the particle's position (extrapolated to redshift z) exceeds a threshold value δ_c.[8] A characteristic halo mass $M_\star$ can then be defined at each redshift via $\sigma(M_\star, z) = \delta_c$. In this model, the halo multiplicity function $f(\sigma)$, which is related to the comoving number density of halos via

$$f(\sigma) = \frac{M}{\bar{\rho}(z)} \frac{\mathrm{d}\,n(M, z)}{\mathrm{d}\log \sigma^{-1}} \;, \qquad (8)$$

takes on a universal form (for perspectives on universality of the halo mass function, see Jenkins et al. 2001; Reed et al. 2007; Lukić et al. 2007; Cohn & White 2008; Tinker et al. 2008).

4.1　Mass Function

In Figure 8 we plot the halo multiplicity function $f(\sigma)$ from the MS-II at four redshifts: $z = 0.0$, 0.99, 2.07, and 6.20.

---

[8]　$\delta_c = 1.68647$ at all redshifts for an Einstein-de Sitter universe. For our ΛCDM cosmology, δ_c varies slightly, from this standard value at high redshift to 1.6737 at redshift zero due to a weak dependence on $\Omega_m(z)$.
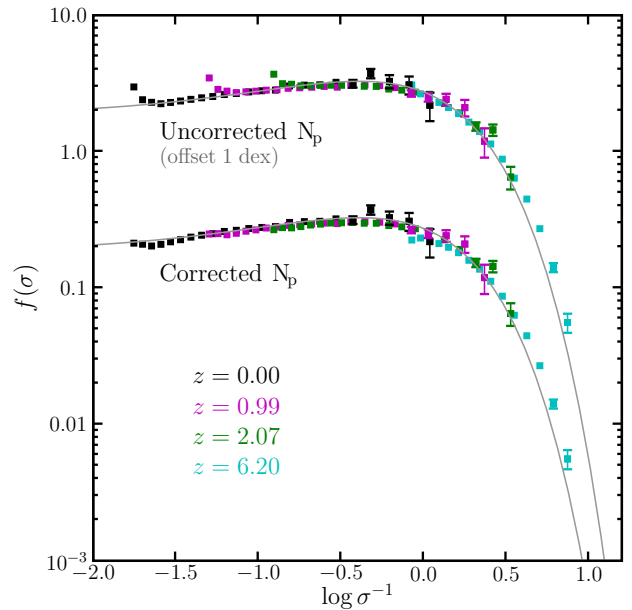

Figure 8. The halo multiplicity function $f(\sigma)$ as a function of $\sigma(M, z)^{-1}$ at four redshifts from the MS-II: $z = 0$ (black squares), 0.99 (magenta), 2.07 (green), and 6.2 (cyan). We compute the multiplicity function both with the Warren et al. N_p correction (lower data points) and without the correction (upper points, offset by 1 dex for clarity). Also overplotted is the Warren et al. (2006) fit to the halo multiplicity function. The MS-II multiplicity function shows universal behavior when scaling with respect to redshift, with deviations at the 10% level. Halo masses here are defined to be $M_{\rm FOF}$.

Here we define halo mass as $M_{\rm FOF}$ and we include Poisson error bars for all bins containing fewer than 400 halos. Results are plotted both using the Warren et al. (2006) correction[9] for sampling bias in N_p (lower set of data points) and without this correction (upper data points, offset upward by 1 dex for clarity). We exclude halos that have $N_p < 20$ when using the corrected N_p. The simulation has fixed mass resolution but $\sigma(M, z)$ evolves significantly with time, so by comparing multiplicity functions at several redshifts we can probe a large range in σ.

The multiplicity function within the MS-II does seem to have a universal form: where the data overlap, the agreement in $f(\sigma)$ is quite good. It thus appears possible to compute the multiplicity function at any redshift simply by combining the linear growth factor $d(z)$ with the rms amplitude of fluctuations as a function of mass at redshift zero. This agreement is at least as good for the uncorrected points, excluding bins containing halos with fewer than 100 particles. (We note, however, that Warren et al. used a minimum of 400 particles per halo in deriving their fitting parameters; in this regime, both the corrected and uncorrected points seem to exhibit 'universality.') The multiplicity function does not agree precisely with the Warren et al. fit (gray line) in either case; however, the volume of the MS-II is not sufficiently large to obtain statistically precise results in the high σ^{-1} regime due to cosmic variance.

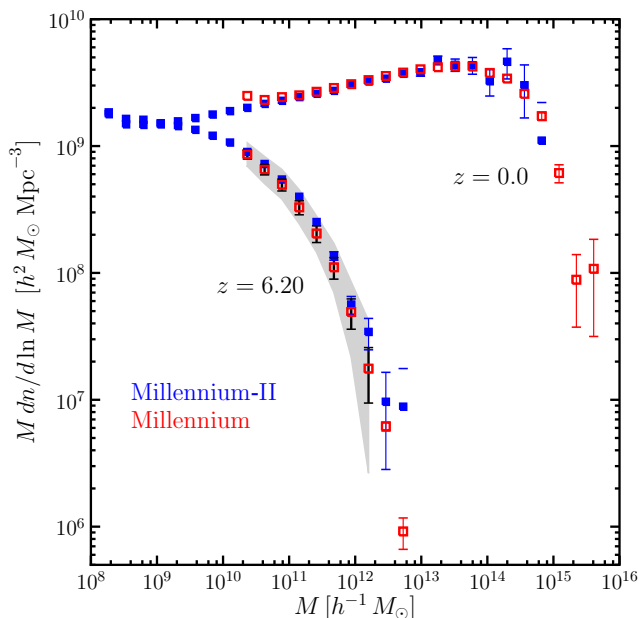[9]　This correction is $N_p \rightarrow N_{\rm corrected} = N_p\,(1 - N_p^{-0.6})$

Figure 9. FOF mass function for the MS-II (solid blue squares) and for the MS (open red squares) compared at redshift 6.2 and 0. The redshift zero mass functions are in excellent agreement over the entire range where the two simulations overlap. At redshift 6.2, the MS-II points lie systematically above those from the MS. The shaded gray region shows the range of mass functions obtained from subdividing the MS into 125 cubes with volume equal to the MS-II and computing a mass function for each subvolume. The MS-II points are well within the scatter, indicating that the difference is likely due to the small volume of the MS-II.

Figure 9 compares the FOF mass function at redshifts 0 and 6.2 determined from the MS-II (solid blue squares) with the MS mass function (open red squares). Poisson error bars are included for all bins with fewer than 400 halos and the data points do not include the Warren et al. correction for the sampling bias in N_p. At $z = 0$, the agreement between the two simulations is excellent for all halo masses (excluding bins containing halos with fewer than 100 particles). Combining the two allows for a consistent measurement of the halo mass function over seven decades in halo mass. At $z = 6.2$, the MS-II mass function lies systematically above that of the MS. The most likely explanation of this difference is cosmic variance: the halos probed by either simulation at $z = 6.2$ are inherently rare objects, as the characteristic mass $M_\star$ is $4.5 \times 10^5\,h^{-1}\,M_\odot$ at that time[10]. Furthermore, the MS-II probes only $1/125^{\rm th}$ the volume of the MS, making statistical fluctuations much more likely.

In order to estimate the effects of cosmic variance on these mass functions, we divided the MS into 125 disjoint sub-cubes, each with the same volume as the MS-II, and we measured the scatter in mass functions and in the mean matter densities $\bar{\rho}_m$ computed from these sub-volumes at $z = 6.2$. The full range of these mass functions is plotted as a gray shaded region in Figure 9, while the rms values at each
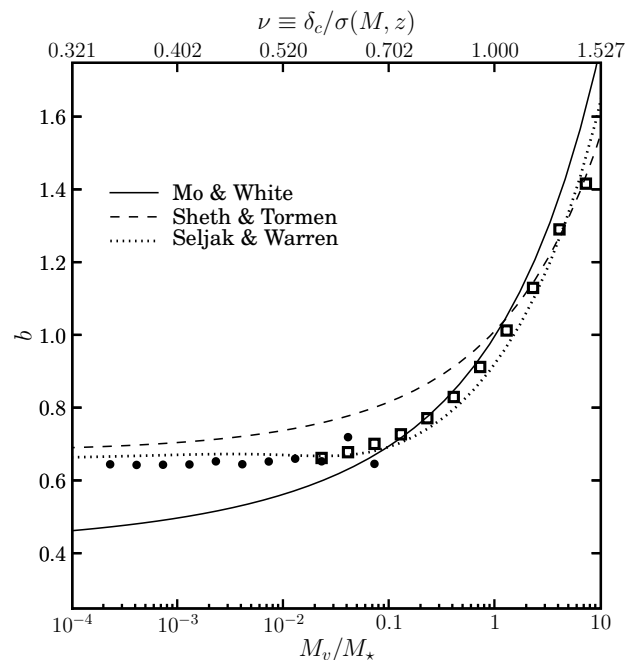

Figure 10. Halo bias at redshift zero. We combine results from the MS-II (filled circles) and the MS (open squares) to explore bias from 10^{-4} to $10\,M_\star$. As expected, the bias decreases as the halo mass decreases, reaching $b(M_\star) \approx 1$. At very low masses ($M_v/M_\star \lesssim 2 \times 10^{-2}$ or $\nu \lesssim 0.55$), the bias reaches an asymptotic value of 0.65.

mass are shown as black error bars on the MS data points. The MS-II points typically lie slightly outside of the rms region but well within the full distribution of mass functions, indicating that they are fully consistent with the MS when the volume of the MS-II is taken into account. We emphasize that the variation in the mass functions between the 125 MS sub-cubes is *not* due to differences in the mean matter density, as the rms scatter in $\bar{\rho}_m$ is only 2% while the rms scatter in the mass function exceeds 8% (the full range of the scatter exceeds $\pm 20\%$) for all of the data points.

4.2 Bias

Dark matter halos do not cluster in the same way as the underlying mass density field but rather exhibit a bias relative to the dark matter. Mo & White (1996), building on the earlier work of Efstathiou et al. (1988) and Cole & Kaiser (1989), showed that the two-point correlation function of halos should be simply related to that of the mass density field. According to the excursion set model, on large scales one should find that

$$\xi_{hh}(M, z; r) = b^2(M, z)\, \xi_{mm}(z; r)\,, \qquad (9)$$

where the bias factor b is given by

$$b(M, z) = 1 + \delta_c^{-1}\left(\nu^2 - 1\right)\,. \qquad (10)$$

Massive halos ($M \gtrsim M_\star$) are therefore predicted to cluster more strongly than the underlying mass density field while low-mass halos should cluster less strongly. This basic picture has been extensively validated, with newer models making improved quantitative predictions for bias (Jing 1998;

---

[10] The minimum halo mass in the MS, $10^{10}\,h^{-1}\,M_\odot$, corresponds to a peak height $\nu \equiv \delta_c/\sigma(M, z)$ of 1.5 at $z = 6.2$, which is equivalent to a mass of $7 \times 10^{13}\,h^{-1}\,M_\odot$ at $z = 0$.

Sheth & Tormen 1999; Sheth et al. 2001; Seljak & Warren 2004).

The bias of low-mass halos ($M \ll M_\star$) remains an unresolved issue. In the Mo & White model, $b \to 1 - \delta_c^{-1} \approx 0.4$ for $M \ll M_\star$. Sheth & Tormen (1999) and Seljak & Warren (2004), on the other hand, find $b \approx 0.7$ in this same regime. The mass resolution of the MS-II allows us to study bias for $M \ll M_\star$: halos with 200 particles, whose number densities and spatial distributions are certainly well-resolved, correspond to $2 \times 10^{-4} M_\star$ at $z = 0$.

Halo bias at redshift zero is shown in Figure 10 for masses down to $2 \times 10^{-4} M_\star$. The bias for $M_v \ll M_\star$ is constant at $b \approx 0.65$ over approximately two decades in mass, from $M_v/M_\star = 2 \times 10^{-4}$ to 2×10^{-2}. In terms of peak height ν (shown on the upper horizontal axis of Figure 10), the bias is constant for $\nu < 0.55$. Above $0.1 M_\star$, the bias rises rapidly with mass, reaching $b = 1$ at M slightly greater than $M_\star$ and $b \approx 1.5$ for $10 M_\star$.

Figure 10 also shows the predictions for $b(M)$ from three fitting formulae: the original Mo & White prediction (solid curve), the Sheth & Tormen model (dashed), and the fit from Seljak & Warren (their eq. 5; dotted). The Seljak & Warren fit clearly agrees best with the data over the range plotted, though it slightly underpredicts the bias at $M > M_\star$ and slightly overpredicts it at $M < M_\star$. These differences are only at the 5% level, however.

5   HALO FORMATION

5.1   Formation Times

The hierarchical nature of ΛCDM models means that the typical formation redshift z_f of halos with mass M is a decreasing function of M. The form of the relation between z_f and M and its intrinsic scatter are important for a number of applications, such as understanding how well galaxy properties can be predicted by halo mass alone, and how well a halo's history can be predicted from its present-day properties. Such characterizations are complicated by the fact that the most useful definition of formation time depends on the question one is asking[11]. For example, the innermost region of a halo – where the galaxy resides – typically assembles much earlier than the outer regions which contain most of the mass (Zhao et al. 2003b; Gao et al. 2004a).

One of the simplest definitions of formation redshift is the time at which a halo's main progenitor reaches a fixed fraction of its present-day mass. We use the merger trees described in Section 2.3 to trace each FOF halo back in time and define its formation redshift z_f as the first redshift at which one of the halo's progenitors reached half of the halo's redshift zero mass (we interpolate between snapshot redshifts to obtain z_f). This 'half-mass' formation time is the most common choice of formation time in the literature. We use M_v as the definition of halo and progenitor mass when estimating such formation times (we have checked that the following results are insensitive to halo mass definition).

In Figure 11, we show the mean relation between halo mass M_v and formation time z_f. In order to determine what mass is required for converged results, we compute

[11] See, e.g., Li et al. 2008 for several possible definitions of z_f
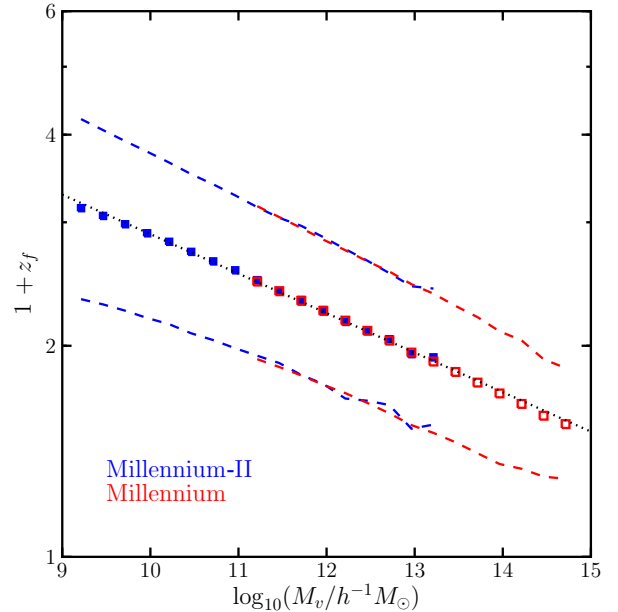

Figure 11. Mean half-mass formation redshift as a function of mass for halos from the MS-II (filled blue squares) and the MS (open red squares). We also show the relation for the 16% earliest and latest forming halos (dashed lines) and the best-fitting linear relation between $\log(1 + z_f)$ and $\log M_v$ (black dotted line; see Equation (11) for the fitting parameters). This fit deviates from the mean relation by less than 2% over the entire range of masses plotted.

z_f from the MS and compare with the MS-II. We find that the two simulations are in excellent agreement above a redshift zero mass of $10^{11} h^{-1} M_\odot$, corresponding to approximately 150 particles in the MS. This is the convergence limit we adopt, so we consider all halos with masses greater than $10^9 h^{-1} M_\odot$ in the MS-II. We only include halos that SUBFIND determines to be bound, although in practice, this restriction makes almost no difference as the fraction of subhalos with $N_p > 100$ that are unbound is very small. Over the entire range where halo formation can be resolved ($9 \leq \log_{10}(M_v/h^{-1} M_\odot) \leq 14.7$), a simple linear fit in $\log(1 + z_f)$ versus $\log(M_v)$ provides an excellent description of the data:

$$1 + z_f = 2.89 \left(\frac{M_v}{10^{10} h^{-1} M_\odot} \right)^{-0.0563}. \quad (11)$$

The maximum deviation between the binned data and Equation (11) is 1.8% over the entire region where: (1) there are at least 100 halos per bin; and (2) halos have at least 125 particles. The 1-σ scatter in the relation, defined as the logarithmic difference between the 84th and 16th percentiles of the data (which are shown in dashed lines in Figure 11), decreases gradually from $\sigma_{\log 1+z} = 0.6$ to 0.4 as M_v increases from 10^9 to $5 \times 10^{14} h^{-1} M_\odot$.

Neto et al. (2007) previously studied the relation between median z_f and halo mass for a set of relaxed halos from the MS. Their fit is similar to ours although the parameters differ slightly (their exponent is -0.046 and their normalization is approximately 2.74 using our form of the fitting formula) due presumably to the selection criteria used for the halos they studied, their use of M_{200} rather than
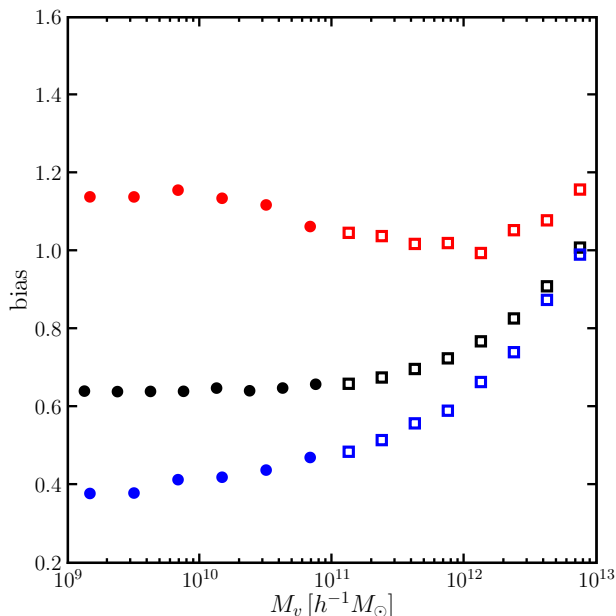
Figure 12. Assembly bias from the MS-II (open symbols) and the MS (filled symbols). We plot the bias of the oldest 20% of halos (red points) and the youngest 20% of halos (blue points), as well as the bias of the full halo sample (black points), as a function of halo mass. The oldest halos cluster much more strongly than the youngest, a ratio of 2.85 in the bias (which is over a factor of 8 in correlation amplitude). At low masses, the bias for each subset reaches an asymptotic value (≈ 0.4 for the young halos, ≈ 1.15 for the old halos).

M_v, and the difference between the median and the mean relation. McBride et al. (2009) have also recently computed the formation times of massive halos from the MS and fitted to the relation $z_f = a \log_{10}(M/10^{12}\,M_\odot) + b$. This gives very similar results to ours over the range of the data they used ($M \gtrsim 10^{12}\,h^{-1}\,M_\odot$), with differences at the 5% level after an empirical normalization correction due to a different mass definition. Their formula underestimates z_f from the MS-II at lower masses: at $10^9\,h^{-1}\,M_\odot$, the difference is approximately 20%.

5.2 Clustering and Formation Times

In the simple version of the excursion set model of structure formation, which is based on top-hat k-space filtering, halo formation is governed by Markov random walks of the (linearly-extrapolated) mass density field $\delta(\mathbf{x}; M)$. A direct consequence of this aspect of excursion set theory is the prediction that properties such as the clustering of halos depend on halo mass alone and not on halo assembly history. Recent N-body simulations have produced results that contradict this prediction, however. Gao et al. (2005) used the MS to show that clustering depends strongly on formation time at masses $M \lesssim M_\star$: they found that early-forming halos cluster much more strongly than late-forming halos, indicating an "assembly bias". Subsequent work confirmed these findings and extended the results to the $M \gtrsim M_\star$ regime and to halo properties other than formation time, including concentration, substructure content, and spin (e.g., Harker et al.

2006; Wechsler et al. 2006; Wetzel et al. 2007; Jing et al. 2007; Bett et al. 2007; Gao & White 2007; Dalal et al. 2008)

With the MS-II, we are able to investigate assembly bias at much lower masses than was previously possible: $M \approx 10^{-4}\,M_\star$ or equivalently $\nu \approx 0.32$. We split each mass bin into the oldest and youngest 20% of halos and compute the bias factor $b(M)$ in the same manner described in Section 4.2. Our results for the dependence of clustering on formation time are presented in Figure 12. We plot the bias of halos as a function of mass at redshift zero for the entire sample of halos (black symbols) as well as for the oldest 20% (red symbols) and youngest 20% (blue symbols) of halos at each mass. We show results for the MS-II (filled circles) for $M \lesssim 2 \times 10^{11}\,h^{-1}\,M_\odot$ and for the MS (open squares) at higher masses in order to maximize statistical significance.

Over virtually the full range of masses probed by the MS-II, the biases of the oldest and youngest halos are approximately constant. The bias of young halos is substantially lower than that of old halos, however, in agreement with previous work. The oldest 20% of halos at a given mass have a slight positive bias with respect to the dark matter distribution. The youngest 20% have a bias that is approximately $b = 0.4$; interestingly, and perhaps coincidentally, this is the value of $1 - \delta_c^{-1} \approx 0.4$ predicted in the $\nu \ll 1$ regime by Press-Schechter and excursion set models (Cole & Kaiser 1989; Mo & White 1996; see also Dalal et al. 2008).

Li et al. (2008) have suggested that the magnitude of assembly bias depends on the adopted definition of formation time. To investigate whether our definition of formation redshift, $M(z_f) = \frac{1}{2}\,M(z = 0)$, influences our results, we have repeated our analysis with a new definition: $M(z_f) = \frac{1}{4}\,M(z = 0)$. We have also tested whether computing the half-mass formation time relative to $M_{\rm FOF}$ rather than to M_v affects our results. Neither of these changes has any detectable influence, so the results we obtain using the standard half-mass formation time appear robust.

6 'MILKY WAY' HALOS IN MILLENNIUM-II

The Milky Way can provide us with a unique variety of insights into galaxy formation, so it is extremely interesting to study the formation of Milky Way-mass halos for comparison with available and forthcoming observations of the detailed structure of our Galaxy. Cosmological simulations of representative volumes of the universe can provide large, statistically complete samples of Milky Way-mass halos, but they cannot resolve the full range of observable structures within the Milky Way. Even at the mass resolution of the MS-II, for example, the halos of dwarf spheroidal galaxies are just barely resolvable. An alternative tack is to focus all of one's computational resources on the formation of a single galaxy-mass halo, thus allowing substantially enhanced mass resolution (Diemand et al. 2007, 2008; Springel et al. 2008; Stadel et al. 2008). With this method, one sacrifices statistical understanding of a representative sample for detailed analysis of one or a few objects. Here we discuss how these two approaches may be combined to extract maximum information about the formation and evolution of galaxy-scale dark matter halos.

Halo	M_{200m} $[h^{-1}\,M_\odot]$	V_{max} [km/s]	R_{max} $[h^{-1}\,kpc]$	$V_{max,\,sub}$ [km/s]
Aq-A	1.842×10^{12}	208.49	20.54	60.42
MS-II-A	1.826×10^{12}	210.20	21.87	61.68
Aq-B	7.629×10^{11}	157.68	29.31	48.31
MS-II-B	7.470×10^{11}	156.21	27.98	45.87
Aq-C	1.641×10^{12}	222.40	23.70	87.07
MS-II-C	1.682×10^{12}	222.03	23.88	89.70
Aq-D	1.839×10^{12}	203.20	39.48	90.63
MS-II-D	1.944×10^{12}	204.39	42.28	91.43
Aq-E	1.130×10^{12}	179.00	40.52	40.87
MS-II-E	1.187×10^{12}	184.15	43.00	45.36
Aq-F	1.107×10^{12}	169.08	31.15	78.33
MS-II-F	1.152×10^{12}	167.18	34.47	81.53

Table 2. Comparison of Aquarius level 2 halos (Aq-X) and Aquarius halos in the MS-II (MS-II-X). M_{200m} is the mass of the dark matter halo, V_{max} is the maximum value of the circular velocity curve, R_{max} is the radius at which the circular velocity curve attains its maximum, and $V_{max,\,sub}$ is the circular velocity curve maximum for the largest subhalo in each halo; all values are quoted at $z = 0$.

6.1 The Aquarius Project and Millennium-II

The Aquarius project (Springel et al. 2008) is a suite of cosmological simulations of the formation of Milky Way-mass dark matter halos. Six halos (denoted 'Aq-A' through 'Aq-F') were simulated at up to five different levels of mass resolution (levels 5 through 1). The highest resolution simulation, Aq-A-1, used a particle mass of $1.25 \times 10^3\,h^{-1}\,M_\odot$, resulting in approximately 1.5 billion particles within R_{200m} at $z = 0$. The halos resimulated in the Aquarius project were selected from the cosmological simulation 'hMS' (Gao et al. 2008), which followed 900^3 particles in a $100\,h^{-1}$ Mpc box. Both the cosmological parameters of hMS and the amplitudes and phases of modes used to generate its initial conditions are identical to those in the MS-II. As a result, all the structures present in hMS are also present in the MS-II, but with a mass resolution that is a factor of 13.8 times better. Since the Aquarius halos were selected from hMS, they are also present in the MS-II and we can compare their properties in the MS-II and in the much higher resolution Aquarius resimulations.

6.2 Comparing Aquarius and Millennium-II halos

As a first test, we compare some basic properties of the Aquarius halos in the Aquarius level 2 simulations – where the particle mass is $(0.5-1) \times 10^4\,h^{-1}\,M_\odot$ – and in the MS-II, where the mass resolution is approximately 1000 times lower. Table 2 contains such a comparison, with halos from the Aquarius simulations in the upper of each set of two rows (labeled 'Aq-X') and halos from the MS-II in the lower of each set of rows (labeled 'MS-II-X'). The first data column compares M_{200m} values for the halos, showing that their masses agree very well: each M_{200m} measured from the MS-II agrees with the corresponding Aquarius resimulation to better than 6% and, for 3 of the 6, to better than 3%. The
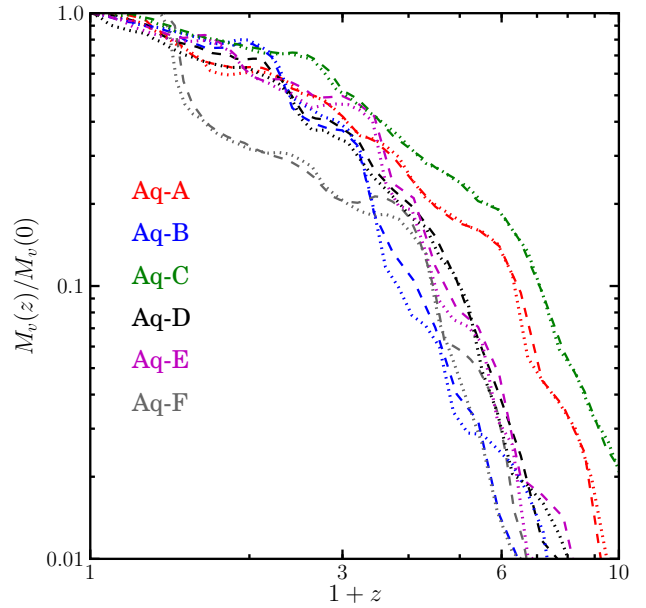

Figure 13. Mass accretion histories for the six Aquarius level 2 resolution halos (dashed lines) and for the corresponding MS-II halos (dotted lines). Even though the mass resolution of the Aquarius simulations is one thousand times better than that of the MS-II, the individual mass accretion histories of the Aquarius halos are captured quite well in the MS-II.

measured V_{max} values all agree within 3% and, for 5 of the 6, to better than 1.5%. The radius at which the circular velocity curve peaks, R_{max}, typically agrees within 5%, with a maximum deviation of 10%. Table 2 also lists the maximum circular velocity of the largest subhalo in each simulation. In all cases this is the *same* subhalo in the MS-II and the Aquarius resimulation, and the circular velocities typically agree to within a few percent.

A more stringent test is to consider the mass accretion history[12] for each Aquarius halo and to compare results from the Aquarius resimulations and from the MS-II. For the MS-II Aquarius halos, we use the merger trees described in Section 2.3 to determine the main progenitor of each halo at each redshift, and we define the mass accretion history of a halo as the mass of its main progenitor $M(z)$ at each redshift z. Merger trees have also been built for the halos in the Aquarius resimulations, and we use these in an identical manner to define the corresponding mass accretion histories.

The results are compared in Figure 13, where we plot mass accretion histories over the redshift range $0 \leq z \leq 9$. We use dashed lines for the Aquarius resimulations and dotted lines for the MS-II. The assembly histories of all the halos are reproduced remarkably well at MS-II resolution. This is a highly non-trivial test, as the mass resolution of the Aquarius level 2 simulations is one thousand times better than that of the MS-II: at $z = 0$ the Aquarius level 2 halos have on the order of 1.80×10^8 particles within R_{200m}

---

[12] for a discussion of the statistical properties of mass accretion histories, see Lacey & Cole (1993); Wechsler et al. (2002); van den Bosch (2002); Zhao et al. (2003b,a, 2008); McBride et al. (2009).

while their MS-II counterparts have approximately 1.80×10^5 particles within this radius.

This agreement is a testament to the integration accuracy of the `GADGET-3` code and shows that the properties of Milky Way-mass halos and their most massive subhalos are well converged in the MS-II. As a result, the MS-II will be very useful for understanding the statistical properties of Milky Way-mass halos since it contains over six thousand halos with $11.5 < \log_{10}(M_v/h^{-1}\,M_\odot) < 12.5$. These halos can be used for a detailed statistical investigation of the growth, internal structure and subhalo populations of halos similar in mass to that of our own Galaxy (Boylan-Kolchin 2009, in preparation). This will test the extent to which the six Aquarius halos are representative of the full halo population at this mass, thereby allowing results obtained from the ultra-high resolution Aquarius resimulations to be interpreted in their full cosmological context.

7 DISCUSSION AND SUMMARY

Understanding galaxy formation in a ΛCDM universe requires knowledge of physical processes over a wide range of scales, from sub-galactic to cosmological. Simulations with volumes large enough to probe the statistics of large-scale structure and resolution high enough to resolve subhalo dynamics within galaxy halos are thus critical for this quest. We have presented initial results from a new simulation, the Millennium-II Simulation, that can resolve all halos down to mass scales comparable of the halos of Local Group dwarf spheroidal galaxies.

Furthermore, the Millennium-II Simulation is closely connected to two other very large computational endeavors, the Millennium Simulation and the Aquarius Project. Throughout this paper we have shown that MS-II results agree extremely well with those from the MS, so we can combine the two simulations to cover an even broader range of physical scales. We have also shown that the properties of the resimulated Aquarius halos agree precisely and in considerable detail with those of their counterparts in the MS-II, even though the resimulations have 1000 times better mass resolution. This gives us confidence not only that the assembly histories of Milky Way-mass halos are well resolved in the MS-II, but also that the properties of their more massive subhalos are converged. As a result, we will be able to use the MS-II to make statistical statements about an ensemble of galaxy-scale halos where the Aquarius Project provides much higher resolution results for a limited but well understood subset.

There is much to do with the MS-II, both from a dark matter perspective and from the point of view of galaxy formation. There are four cluster-size halos in the MS-II with over 60 million particles at $z = 0$; investigating density profiles and substructure abundances for these objects, and comparing with state-of-the-art galaxy-scale simulations such as Aquarius, will shed light on whether dark matter structures are self-similar as a function of scale. Subhalo survival times and merger rates, which are crucial ingredients in galaxy formation models, can be checked in untested regimes. Furthermore, directly resolving much smaller halos means that semi-analytic models initially developed for the Millennium Simulation (Springel et al. 2005; Croton et al.

2006; Bower et al. 2006; De Lucia & Blaizot 2007) can now be updated and extended to much lower galaxy masses (Guo et al. 2009, in preparation). With this paper, we publicly release the FOF halo and subhalo catalogs and the subhalo merger trees in a searchable database structured in the same way as has already been done for the Millennium Simulation[13] (Lemson & The Virgo Consortium 2006). As work progresses, we plan to make MS-II semi-analytic galaxy catalogs available as well. This will allow others to use the MS-II data for their own research purposes.

REFERENCES

Baugh, C. M., Gaztanaga, E., & Efstathiou, G. 1995, MN-RAS, 274, 1049

Bett, P., Eke, V., Frenk, C. S., Jenkins, A., Helly, J., & Navarro, J. 2007, MNRAS, 376, 215

Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, 379, 440

Bower, R. G., Benson, A. J., Malbon, R., Helly, J. C., Frenk, C. S., Baugh, C. M., Cole, S., & Lacey, C. G. 2006, MNRAS, 370, 645

Bryan, G. L., & Norman, M. L. 1998, ApJ, 495, 80

Cohn, J. D., & White, M. 2008, MNRAS, 385, 2025

Cole, S., & Kaiser, N. 1989, MNRAS, 237, 1127

Colless, M., & Dunn, A. M. 1996, ApJ, 458, 435

Cooray, A., & Sheth, R. 2002, Phys. Rep., 372, 1

Croton, D. J. et al. 2006, MNRAS, 365, 11

Dalal, N., White, M., Bond, J. R., & Shirokov, A. 2008, ApJ, 687, 12

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371

De Lucia, G., & Blaizot, J. 2007, MNRAS, 375, 2

Diemand, J., Kuhlen, M., & Madau, P. 2007, ApJ, 657, 262

Diemand, J., Kuhlen, M., Madau, P., Zemp, M., Moore, B., Potter, D., & Stadel, J. 2008, Nature, 454, 735

Efstathiou, G., Frenk, C. S., White, S. D. M., & Davis, M. 1988, MNRAS, 235, 715

Gao, L., Loeb, A., Peebles, P. J. E., White, S. D. M., & Jenkins, A. 2004a, ApJ, 614, 17

---

[13] see http://www.mpa-garching.mpg.de/millennium

Gao, L., Navarro, J. F., Cole, S., Frenk, C. S., White, S. D. M., Springel, V., Jenkins, A., & Neto, A. F. 2008, MNRAS, 387, 536

Gao, L., Springel, V., & White, S. D. M. 2005, MNRAS, 363, L66

Gao, L., & White, S. D. M. 2007, MNRAS, 377, L5

Gao, L., White, S. D. M., Jenkins, A., Stoehr, F., & Springel, V. 2004b, MNRAS, 355, 819

Ghigna, S., Moore, B., Governato, F., Lake, G., Quinn, T., & Stadel, J. 1998, MNRAS, 300, 146

Harker, G., Cole, S., Helly, J., Frenk, C., & Jenkins, A. 2006, MNRAS, 367, 1039

Hawkins, E. et al. 2003, MNRAS, 346, 78

Hayashi, E., Navarro, J. F., Taylor, J. E., Stadel, J., & Quinn, T. 2003, ApJ, 584, 541

Iliev, I. T., Mellema, G., Pen, U.-L., Merz, H., Shapiro, P. R., & Alvarez, M. A. 2006, MNRAS, 369, 1625

Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., & Yoshida, N. 2001, MNRAS, 321, 372

Jing, Y. P. 1998, ApJ, 503, L9

Jing, Y. P., Suto, Y., & Mo, H. J. 2007, ApJ, 657, 664

Kim, J., Park, C., Gott, J. R. I., & Dubinski, J. 2008, arXiv:0812.1392 [astro-ph]

Klypin, A., Gottlöber, S., Kravtsov, A. V., & Khokhlov, A. M. 1999a, ApJ, 516, 530

Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999b, ApJ, 522, 82

Kravtsov, A. V., Gnedin, O. Y., & Klypin, A. A. 2004, ApJ, 609, 482

Lacey, C., & Cole, S. 1993, MNRAS, 262, 627

Lemson, G., & The Virgo Consortium. 2006, arXiv:astro-ph/0608019

Li, C., & White, S. D. M. 2009, arXiv:0901.0706 [astro-ph]

Li, Y., Mo, H. J., & Gao, L. 2008, MNRAS, 389, 1419

Lukić, Z., Heitmann, K., Habib, S., Bashinsky, S., & Ricker, P. M. 2007, ApJ, 671, 1160

Ma, C.-P., & Fry, J. N. 2000, ApJ, 543, 503

McBride, J., Fakhouri, O., & Ma, C.-P. 2009, arXiv:0902.3659 [astro-ph]

Mo, H. J., & White, S. D. M. 1996, MNRAS, 282, 347

Moore, B., Ghigna, S., Governato, F., Lake, G., Quinn, T., Stadel, J., & Tozzi, P. 1999, ApJ, 524, L19

Moore, B., Lake, G., & Katz, N. 1998, ApJ, 495, 139

Neto, A. F. et al. 2007, MNRAS, 381, 1450

Peacock, J. A., & Dodds, S. J. 1996, MNRAS, 280, L19

Peacock, J. A., & Smith, R. E. 2000, MNRAS, 318, 1144

Power, C., Navarro, J. F., Jenkins, A., Frenk, C. S., White, S. D. M., Springel, V., Stadel, J., & Quinn, T. 2003, MNRAS, 338, 14

Press, W. H., & Schechter, P. 1974, ApJ, 187, 425

Reed, D. S., Bower, R., Frenk, C. S., Jenkins, A., & Theuns, T. 2007, MNRAS, 374, 2

Rudd, D. H., Zentner, A. R., & Kravtsov, A. V. 2008, ApJ, 672, 19

Seljak, U. 2000, MNRAS, 318, 203

Seljak, U., & Warren, M. S. 2004, MNRAS, 355, 129

Seljak, U., & Zaldarriaga, M. 1996, ApJ, 469, 437

Sheth, R. K., Mo, H. J., & Tormen, G. 2001, MNRAS, 323, 1

Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119

Sirko, E. 2005, ApJ, 634, 728

Smith, R. E. et al. 2003, MNRAS, 341, 1311

Springel, V. 2005, MNRAS, 364, 1105

Springel, V. et al. 2008, MNRAS, 391, 1685

—. 2005, Nature, 435, 629

Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001a, MNRAS, 328, 726

Springel, V., Yoshida, N., & White, S. D. M. 2001b, New Astronomy, 6, 79

Stadel, J., Potter, D., Moore, B., Diemand, J., Madau, P., Zemp, M., Kuhlen, M., & Quilis, V. 2008, arXiv:0808.2981 [astro-ph]

Teyssier, R. et al. 2008, arXiv:0807.3651 [astro-ph]

Tinker, J., Kravtsov, A. V., Klypin, A., Abazajian, K., Warren, M., Yepes, G., Gottlöber, S., & Holz, D. E. 2008, ApJ, 688, 709

Tormen, G., Diaferio, A., & Syer, D. 1998, MNRAS, 299, 728

van den Bosch, F. C. 2002, MNRAS, 331, 98

Warren, M. S., Abazajian, K., Holz, D. E., & Teodoro, L. 2006, ApJ, 646, 881

Wechsler, R. H., Bullock, J. S., Primack, J. R., Kravtsov, A. V., & Dekel, A. 2002, ApJ, 568, 52

Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., & Allgood, B. 2006, ApJ, 652, 71

Wetzel, A. R., Cohn, J. D., White, M., Holz, D. E., & Warren, M. S. 2007, ApJ, 656, 139

White, M. 2001, A&A, 367, 27

White, S. D. M. 1996, in Cosmology and Large Scale Structure, ed. R. Schaeffer, J. Silk, M. Spiro, & J. Zinn-Justin, 349

Zehavi, I. et al. 2002, ApJ, 571, 172

Zentner, A. R. 2007, International Journal of Modern Physics D, 16, 763

Zhao, D. H., Jing, Y. P., Mo, H. J., & Börner, G. 2003a, ApJ, 597, L9

—. 2008, arXiv:0811.0828 [astro-ph]

Zhao, D. H., Mo, H. J., Jing, Y. P., & Börner, G. 2003b, MNRAS, 339, 12