Prepared for submission to JCAP

Generating Mock Catalogs for the Baryon Oscillation Spectroscopic Survey: An Approximate N-Body approach

Aaronson Aardvark...e.t.c.

Abstract. We introduce and test an approximate scheme for generating mock catalogs for large-scale structure measurements in galaxy surveys, specializing in this work to the Baryon Oscillation Spectroscopic Survey. things to add later...A brief description of the approximation scheme, tests and the accuracy we reach, and some comments about the timings of the tests and the BOSS sampes.

Contents

1 Introduction

Why do we need large number of large N-body simuations? Discuss both covariance matrix estimation as well as systematic error estimation. Emphasize importance to capture as much of the physics as is possible...e.t.c.

2 HACC

3 Convergence Test 1: Mass Resolution

In this section, we investigate how mass resolution affects on halo bias and mass functions. We prepare two samples with 256^3 and 512^3 particles in the cubic box, whose side length is $256h^{-1}$Mpc. Note that both samples are simulated from the same initial density field with the same number of time steps.

We first show comparison of cumulative mass functions in Figure 1. We calculate mean and its error through the bootstrap method, generating 100 samples by choosing halos randomly from an output of the simulation. In the right panel of Figure 1, the mean and its error for the samples of 512^3 particles are indicated as shaded regions and the mean for 256^3 particles indicated as circles. The discrepancy in the mass functions between 256^3 particles and 512^3 particles becomes larger on large halo masses. This is mainly because there are a few large halos and therefore it is more sensitive to the differences in the number of halos. It is possible that samples with higher mass resolutions resolve more small halos and a multiple halos in the samples with 512^3 particles correspond to one larger halo in the 256^3 particles samples. We also see that agreement between the samples of those two mass resolutions is better at lower redshift (How can I investigate the reason for this?).

In Figure 2, we compare halo-matter cross power spectra at various redshifts. Here, we use the same matter density field generated with 256^3 particles for both cases, and we select halos based on the "soft-mass cut" method using the probability given by

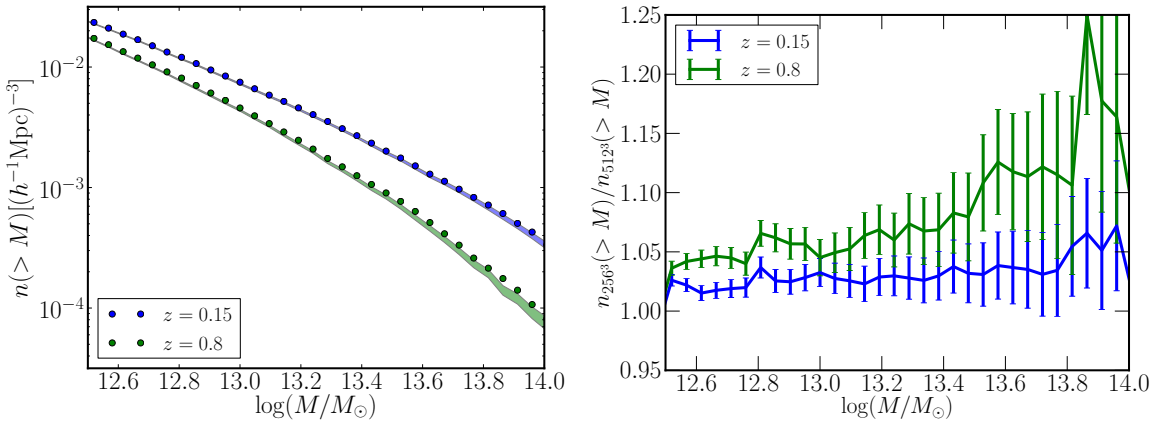

Figure 1. Left: Cumulative mass functions for the simulations of 512^3 and 256^3 particles. Shaded regions are the mean with errors for the samples of 512^3 particles calculated through the bootstrap method. Circles are the mean for 256^3 particles. Right: Ratio of cumulative mass functions between the samples of 512^3 and 256^3 particles. The deviation from one on large halo masses is mainly caused by the fewer number of large halos.

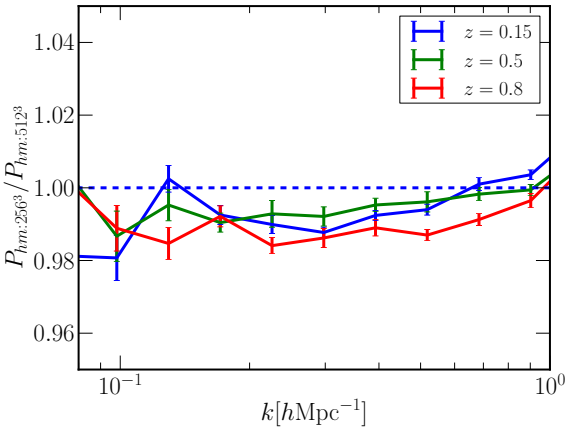

Figure 2. Ratio of halo-matter cross power spectra between the samples of 512^3 and 256^3 particles for various redshifts. The overall agreements are all within 2%. Note that halos are selected based on the "soft-mass cut" method described in the section.

$$< N_{halo}(M) >= \frac{1}{2}\mathrm{erfc}(\frac{\log(\mathrm{M_{cut}}/M)}{\sqrt{2}\sigma}), \tag{3.1}$$

where we set $\mathrm{M_{cut}} = 10^{13.0}[\mathrm{M}_\odot]$ and $\sigma = 0.5$. This probability has a similar form to the halo occupation distribution (hereafter, HOD) technique so that the probability gradually becomes one as increasing halo mass. We use this method to avoid noise from halos scattering across sharp boundaries on halo mass. Using the soft mass-cut method, we generate 10 samples to compute the cross power spectra and their errors. From now on, we use the same method for halo selections to compute power spectra. Note that the ratio of those cross power spectra is equivalent to the ratio of halo bias. Figure 2 shows the halo bias for 256^3 particles is smaller than for 512^3 particles and they agree well within 2%. This also implies the effect on halo bias due to different mass resolusion is less sensitive compared to the effect on mass functions.

4 Convergence Test 2: Selection of the minimal time steps

The goal in this section is to determine the sufficient time-stepping scheme to resolve halo positions and masses reliably for the future galaxy surveys. We are particularly interested in reducing the number of sub-cycles, because most of computational time is spent in the short-range time solver. In order to quantitatively evaluate different time-stepping schemes, we run a set of convergence tests using smaller simulation boxes. We scale down these volumes to $(256h^{-1}\mathrm{Mpc})^3$ with 256^3 particles, which keeps the particle mass unchanged. The number of time steps were chosen as 450/5, 300/3, 300/2, 150/3, 150/2, where the first number indicates the number of long time steps and the second number the number of short time steps for each long time step. Note that all the samples used in this section are at redshift $z = 0.15$. After determining our target time-stepping scheme, we do a more detailed examination at various redshifts in the next section. To evaluate different choices, we compare the properties (masses, positions, and velocities) of the individual halos themselves in addition to the statistical descriptions provided by the mass function and power spectra.

4.1 Matching

Here, we compare halo properties by matching halos in different samples one by one. We first show our algorithm for identifying the corresponding halos in two different samples and then compare halo mass, position, and velocity for those matched halos. From the quantitative comparison, we find that the samples generated from the simulations with 300 global steps have much less scatter for the baseline of the sample of the 450/5 simulation than the samples with 150 global steps. In addition to that, we see that the differences between sub-cycles are almost negligible.

4.1.1 Algorithm

Since our simulations all start with the same initial conditions, we match halos in different simulations by matching their particle content. Given a halo in simulation A, we consider the halos in simulation B with the corresponding particles. Given this list of possible matches, we match to the halo with the largest number of common particles. To avoid spurious matches, we also require that the fraction of common particles (relative to simulation A) exceeds a threshold. As an example to illustlate how this matching algorithm works, we use the samples from the 300/2 simulation and the 450/5 simulation. Figure 3 shows the cumulative fraction of unmatched halos matching the 450/5 simulation to the 300/2 simulation at $z = 0.15$ with various thresholds. As expected, the unmatched fraction increases with increasing threshold and decreasing halo mass. We adopt a threshold of 50% as our default choice.

Since the above matching algorithm is unidirectional, multiple halos in the sample A might be matched to a single halo in the sample B; this happens 1 to 2% of the time with a matching threshold of 50%. We refer to these as multiply-booked halos in what follows. Figure 4 compares halo masses matching the 450/5 simulation to the 300/2 simulation for the case of multiply-booked halos and the rest. The top left panel shows a mass scatter for all the matched halos between those two simulations, while the top right panel shows a mass scatter only for the non multiply-booked halos. The bottom panels show a mass scatter for the case of multiply-booked halos. The bottom left panel shows a mass scatter for individual multiply-booked halos, while we plots a summed halo mass for those corresponding halos in the bottom right panel. As shown in the top left panel, there are low-mass halos in the 450/5 simulation corresponding to high-mass halos in the 300/2 simulation. The same trend is observed for the case of multiply-booked halos, but not for the non multiply-booked halos. Furthermore, those disagreement for halo masses between the two simulations are resolved by adding the corresponding halo masses. This implies that there are multiple halos in the 450/5 simulation which are merged into one halo in the 300/2 simulation.

Figure 5 shows the number densities of the unmatched halos in the 450/5 simulation matching to the 300/2 simulation at $z = 0.15$. There are three reasons that halos are considered as unmatched. First, if particles forming a halo in the sample A do not form a halo in the sample B (i.e., halos in the samples do not share common particles), we consider them as unmatched. Second, if the fraction of common particles over the total number of particles in each halo is less than 50%, we eliminate halos

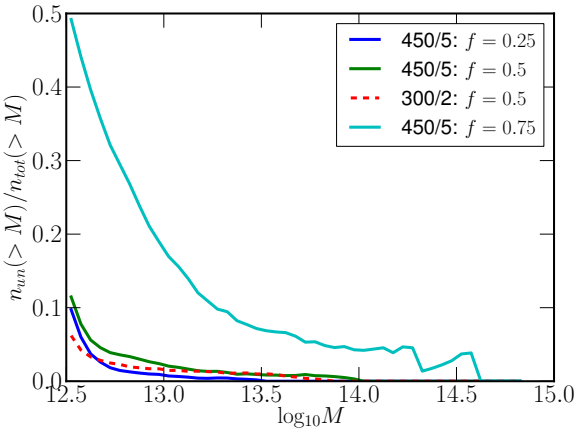

Figure 3. The cumulative fraction of unmatched halos matching the 450/5 simulation to the 300/2 simulation at z=0.15 as a function of halo mass. The solid lines, from top to bottom, correspond to matching thresholds of 75%, 50%, and 25% for the unmatched halos in the 450/5 sample. The dashed line shows the same quantity for the 300/2 sample for a threshold of 50%. As expected, the unmatched fraction increases with decreasing halo mass and increasing threshold. We adopt a threshold of 50% as our default choice. Do we really need the result for the 300/2 simulation?

for the case of spurious matching. At last for the case of multiply-booked halos, we remove all but the one with the largest number of common particles. We showed each unmatched number density as a function of halo mass. We only find unmatched halos on low-mass regions for the reason that the halos do not have any common particles. This is because there are some low-mass halos which are identified in one sample but not in another sample due to the way the FOF algorithm define halos. As shown, most of unmatched halos are due to the threshold criterion. We also checked how the number of matched halos is changed as a function of redshift, and we observed that redshift does not affect to the matching algorithm.

4.1.2 Halo Properties

Here, we compare halo properties (i.e., halo mass, position, and velocity) for halos matched to those in the 450/5 simulation. The comparison of halo mass between the 450/5 simulation and the 300/2 simulation at $z = 0.15$ is shown in Figure 6. Figure 6 shows that the halos in the 300/2 simulation tend to be slightly more massive than the corresponding halos in the 450/5 simulation. The panels in Figure 7, from left to right, show the comparison of halo position and velocity for the matched halos at $z = 0.15$. As shown, the 150 global steps have more scatter in the halo properties and the means differ for the halo mass ratio and the velocity difference. This indicates that the halo structure in these cases is more diffused than the case of the 300 or the 450 global steps. For the 300 global steps, the results are significantly improved and the center position is matched in these cases to better than 200 kpc. As is clear from Figure 7, the difference between 3 and 2 sub-cycles is negligible on halo properties. Note that we observe the same trend in halo properties discussed here at different redshifts.

4.2 Observables

We investigate how the number of time steps affects the observable quantities including mass functions and power spectra.

4.2.1 Time Steps

We first compute mass functions from outputs of different time-stepping schemes, as shown in Figure 8, where we compare simulations with reduced number of time steps to the 450/5 simulation. In Figure 8, we show the ratio $n(> M)/n_{450/5}(> M)$, where $n_{450/5}(> M)$ is a cumulative mass function

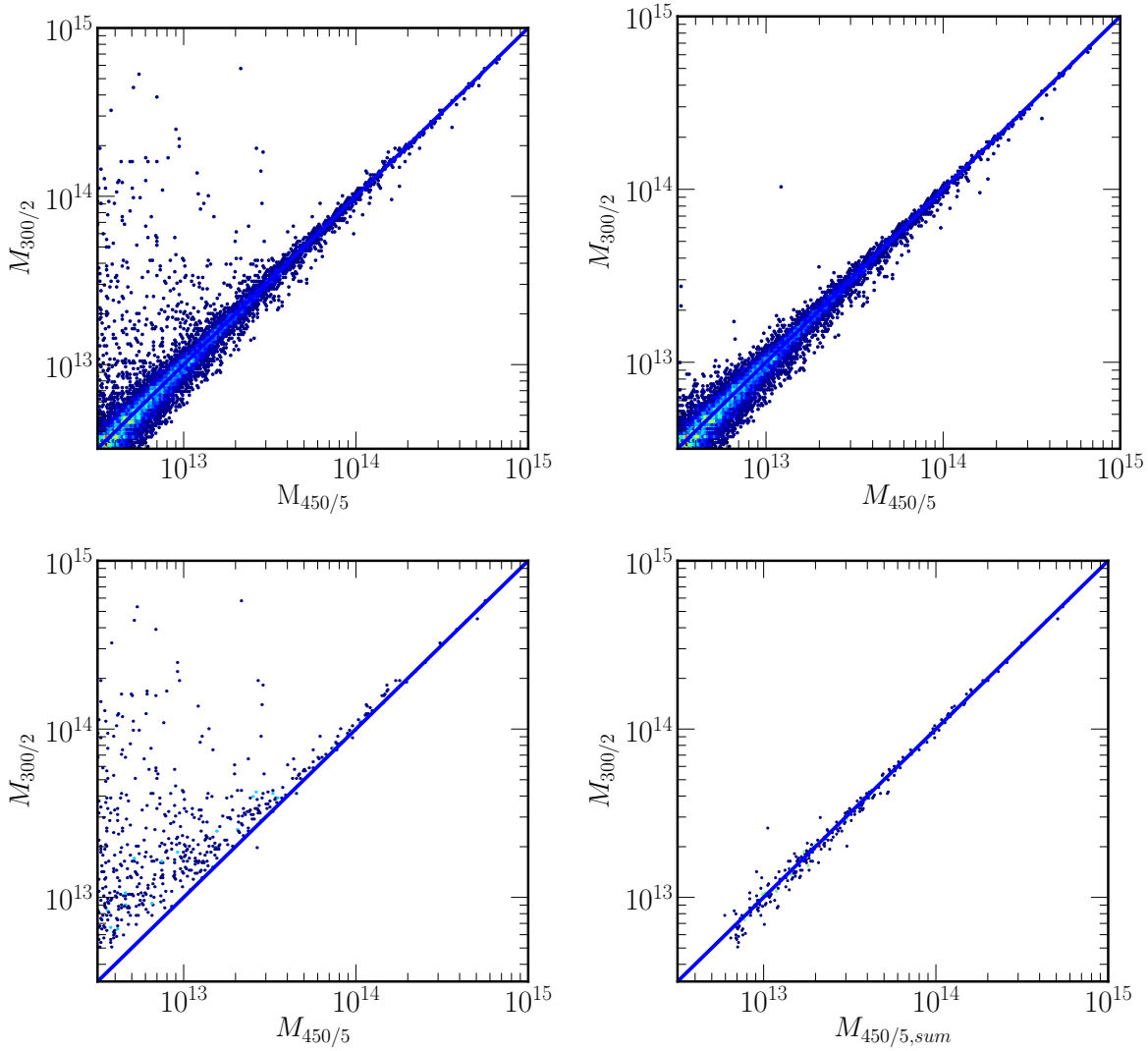

Figure 4. Comparison of halo masses matching the 450/5 simulation (x-axis) to the 300/2 simulation (y-axis) at $z = 0.15$. Panels correspond to halos with different matching criteria: all the matched halos (top left), matched halos having one-to-one correspondence (top right), matched halos not having one-to-one correspondence called "multiply-booked" halos (bottom left), and the "multiply-booked" halos whose corresponding halo masses are added (bottom right). Those panels imply that large mass difference between the 450/5 simulation and the 300/2 simulation shown in the top left panel is mainly because those "multiply-booked" halos in the 450/5 simulation are merged into one halo in the 300/2 simulation due to larger time steps.

for the 450/5 simulation and $n(> M)$ is a cumulative mass function for different time steps shown in different colors. We see that the cumulative mass functions from the simulation with 150 global steps have smaller amplitudes overall than those for the samples with 450 and 300 global steps. This is because a smaller number of time steps makes halos more diffuse and the fixed linking length cannot connect some particles in the outputs of 150 global steps whose corresponding particles are connected in the simulations of a larger number of global time steps. In Manera et al. 2012 which uses the second-order Lagrangian perturbation theory to generate dark matter fields, they used a longer linking length and reassigned halo masses in order to solve the same problem of having diffused halos. We find that overall agreement between the 450/5 simulation and the simulations with 300 global

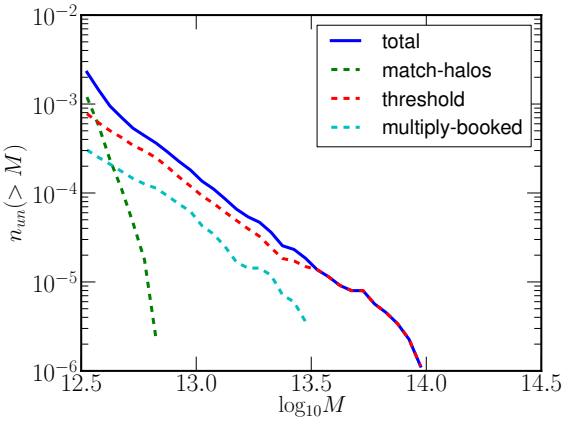

Figure 5. Itemization of unmatched halos shown as cumulative number densities of the unmatched halos from each procedure in the matching algorithm. The solid line is the total and the dashed lines correspond to matching based on particle content (green), elimination due to the matching threshold (red), and elimination of "multiply-booked" halos (cyan). Most large halos being unmatched is due to the threshold.

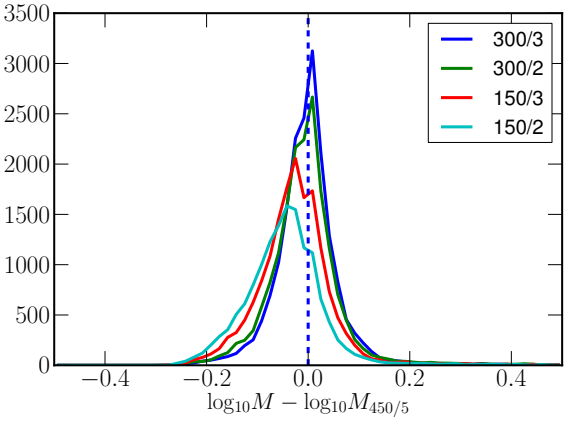

Figure 6. Comparison of halo mass differencecs for matched halos in the different simulations corresponding to the time steps of 300/3 (blue), 300/2 (green), 150/3 (red), and 150/2 (cyan) with respect to the 450/5 simulation. Talk about this plot!!!

steps on mass functions is sufficient.

The next measure of interest is the cross power spectra between halos and matter densities, as shown in Figure 9. Figure 9 shows the ratio $P_{hm}/P_{hm,450/5}$ at $z = 0.15$, where $P_{hm,450/5}$ is the cross power spectrum for the 450/5 simulation and P_{hm} is the cross power spectrum for other time steps labeled in the figure. To calculate the cross power spectra, we use the output of the 450/5 simulation for the matter densities for all the halo samples. In this way, the ratio $P_{hm}/P_{hm,450/5}$ is equivalent to the ratio of halo bias between the 450/5 simulation and the simulations with other time-steps. Note that the errors calculated here are not due to sample variance, because we generate 10 samples from one full sample by using the soft-mass cut method. We see that the agreement between the 450/5 simulation and the simulations with 300 global steps is remarkably good and that the differences are well within 2% on any scales. In addition, we see the deviations on halo biases for the simulations with 150 global steps are also within 5% from the halo bias for the 450/5 simulation on large scales. This implies that halo bias on large scales is not largely affected by reduction of the time steps.

As a conclusion through several convergence tests shown in this section, we determine 300/2 as

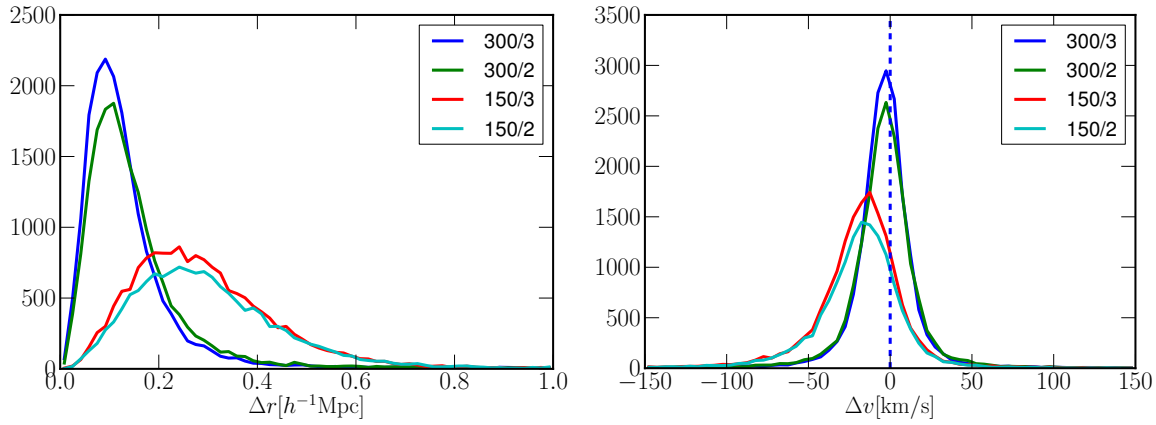

Figure 7. Comparison of matched halos in the different simulations corresponding to the time steps of 300/3 (blue), 300/2 (green), 150/3 (red), and 150/2 (cyan) with respect to the 450/5 simulation. From left to right, we compared halo position and velocity respectively. The agreement between 300 global steps and the 450/5 simulation is considerably good, with little difference from the number of sub-cycles.

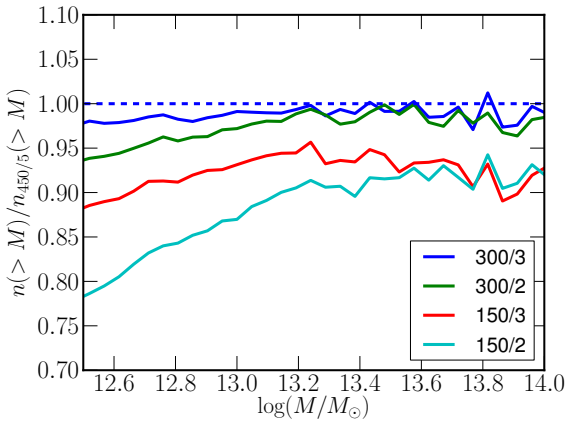

Figure 8. Comparison of cumulative mass functions in different simulations taking the 450/5 simulation as a reference. Lines, from top to bottom, correspond to the simulation with different time steps, 300/3 (blue), 300/2 (green), 150/3 (red), and 150/2 (cyan) respectively. As shown, the agreement between the simulations with 300 global steps and the 450/5 simulation is close to one on large halo mass and sub-cycles make differences only on small halo mass.

our optimal choice.

5 Convergence Test 3: Tuning 300/2

In the previous section, we choose the 300/2 simulation as our final target for the time-stepping scheme. Now, we need to take a close look at observable quantities calculated from the samples of the 300/2 simulation at various redshifts. Here, we first compare cumulative mass functions between the 450/5 simulation and the 300/2 simulation at redshifts $z = 0.15$, $z = 0.5$, and $z = 0.8$ corresponding to different colors shown in Figure 10. As shown, an offset from one increases with higher redshifts, particularly on small halo masses. To make the mass functions for the 300/2 simulation closer to the ones for the 450/5 simulation, we decided to reassign halo masses for halos in the simulations of the 300/2 simulation.

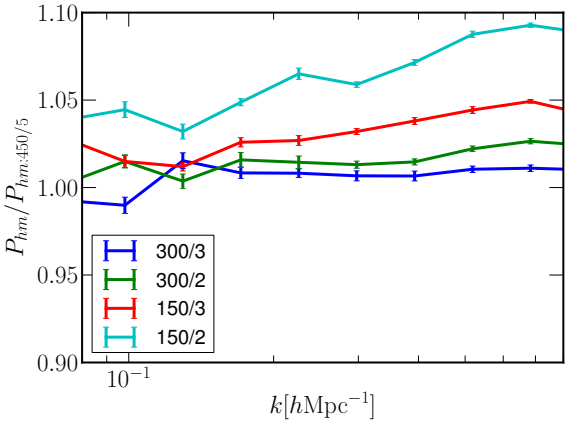

Figure 9. Ratio of halo-matter cross power spectra as a function of time steps with respect to the 450/5. The agreements with the 450/5 are all within 5% on large scales. For 300 global steps, both agree well even on small scales with little difference by sub-cycles. Note that the halos are selected based on the soft mass-cut method with $M_{cut} = 13.0$ and $\sigma = 0.5$.

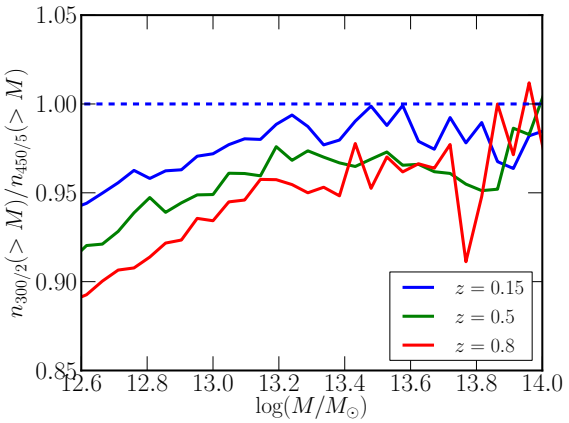

Figure 10. Comparison of cumulative mass functions between the 450/5 simulation and the 300/2 simulation at $z = 0.15$, $z = 0.5$, and $z = 0.8$. For the larger redshifts, the deviation gets larger particularly on small halo masses.

5.1 method

In order to compare halo mass differences between the 300/2 simulation and the 450/5 simulation, we first match halos in the 300/2 simulation to halos in the 450/5 simulation by using our matching algorithm. Next, we take mean of the halo mass differences for those matched halos as a function of halo mass of 300/2 and fit the mean to a functional shown below,

$$M_{re} = M_{300/2}(1.0 + \alpha(M_{300/2}/10^{12.0}[\mathrm{M}_\odot])^\beta, \tag{5.1}$$

where M_{re} is a reassigned halo mass for the samples of 300/2, $M_{300/2}$ is an original halo mass of 300/2, α and β are free parameters which are functions of the redshift. By fitting to the samples of the 300/2 simulation to the 450/5 simulation by using the above functional, we find best fit parameters shown below:

$$\alpha(z) = 0.123z + 0.052, \tag{5.2}$$

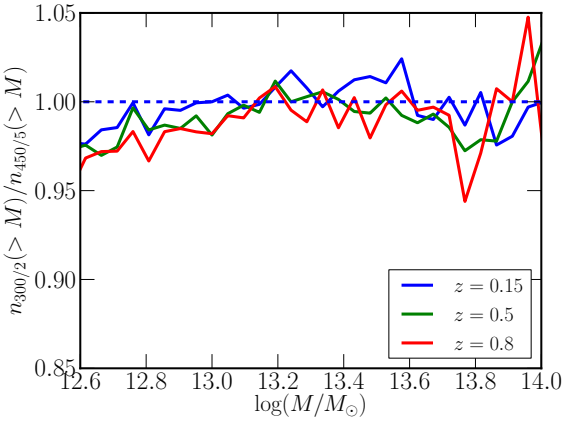

Figure 11. Comparison of cumulative mass functions after reassigning halo masses for the 300/2 simulation. The agreement between the 300/2 simulation and the 450/5 simulation is significantly improved especially on halo masses greater than $10^{13.0}\mathrm{M}_\odot$.

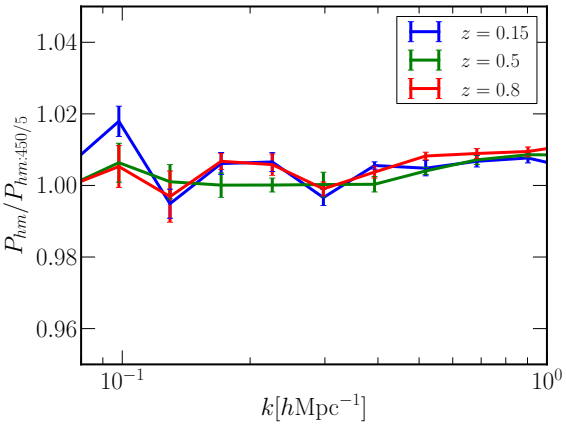

Figure 12. Ratio of halo-matter cross power spectra after reassigning halo masses for the sample from the 300/2 simulations. We select halos based on the soft mass-cut method with $M_{cut} = 13.0$ and $\sigma = 0.5$. The overall agreements are well within 2%.

and

$$\beta(z) = -0.154z - 0.447. \tag{5.3}$$

Applying the functional to the samples from the 300/2 simulation, we obtain mass functions shown in Figure 11. The match of reassigned halo mass functions for the 300/2 simulation to the ones for the 450/5 simulation is significantly improved that now the difference is well within 5% on any halo masses at any redshifts.

5.2 Power Spectra

We calculate halo-matter cross power spectra after reassigning halo masses for the 300/2 simulation, as shown in Figure 12. We use halo mass thresholds of $10^{12.5}\mathrm{M}_\odot$ for the left panel and $10^{13.0}\mathrm{M}_\odot$ for the right panel. As expected from the mass functions, the match between the 300/2 simulation and the 450/5 simulation with the threshold of $10^{13.0}\mathrm{M}_\odot$ is better than the one with $10^{12.5}\mathrm{M}_\odot$. The agreement between the 300/2 simulation and the 450/5 simulation is, however, well within 2% for both cases, which is a significant improvement compared to the results shown in the previous section.

6 THE BOSS SIMULATIONS

6.1 Simulation Parameters

6.2 Building the Galaxy Catalog

6.3 An Application

Table: $\log_{10}(M/M_{450/5})$

z=0.15	median	$\Delta_{65\%}$	$\Delta_{95\%}$	z=0.5	median	$\Delta_{65\%}$	$\Delta_{95\%}$	z=0.8	median	$\Delta_{65\%}$	$\Delta_{95\%}$
300/3	-0.0026	0.0877	0.2864	300/3	-0.0062	0.0905	0.2869	300/3	-0.0067	0.0932	0.2891
300/2	-0.0078	0.0972	0.3458	300/2	-0.0139	0.1029	0.3345	300/2	-0.0189	0.1048	0.3347
150/3	-0.0266	0.1107	0.3101	150/3	-0.0372	0.1148	0.3083	150/3	-0.0519	0.1172	0.2926
150/2	-0.0454	0.1207	0.3315	150/2	-0.0595	0.1252	0.3093	150/2	-0.0782	0.1274	0.28878

Table 1. Comparison of halo mass ratios $\log_{10}M/M_{450/5}$ in log-based, comparing various time steps to the 450/5 simulation at $z = 0.15$, $z = 0.5$, and $z = 0.8$. For each redshift, we report median, $\Delta_{95\%}$, and $\Delta_{65\%}$ (how should I explain about $\Delta_{95\%}$?). As shown in Figure 7, the results indicate that mass ratio distributions for the 300 global steps have less scatter than for the 150 global steps.

z=0.15	median	$\Delta_{65\%}$	$\Delta_{95\%}$	z=0.5	median	$\Delta_{65\%}$	$\Delta_{95\%}$	z=0.8	median	$\Delta_{65\%}$	$\Delta_{95\%}$
300/3	-3.36	24.3872	82.20	300/3	-3.78	36.87	118.48	300/3	-6.32	47.41	149.72
300/2	-3.26	27.5063	99.40	300/2	-3.94	40.77	137.62	300/2	-6.47	53.48	169.1556
150/3	-16.61	38.4702	116.93	150/3	-23.25	53.41	151.75	150/3	-25.23	67.62	186.29
150/2	-17.05	39.9572	121.45	150/2	-24.26	56.75	160.15	150/2	-24.76	71.97	189.26

Table 4. Comparison of halo velocity for various time steps to the 450/5 simulation at $z = 0.15$, $z = 0.5$, and $z = 0.8$. For each redshift, we report median, $\Delta_{95\%}$, and $\Delta_{65\%}$. As shown in Figure 7, the results indicate that mass ratio distributions for the 300 global steps have less scatter than for the 150 global steps.

z=0.15	fraction within 10 degree
300/3	0.951
300/2	0.938
150/3	0.928
150/2	0.923

Table 5. Fractions of halos whose velocity direction matches with the 450/5 within 10 degree at $z = 0.15$. More than 90% f halos for any time steps agree with the 450/5 simulation.