ENVIRONMENTAL EFFECTS ON REAL-SPACE AND REDSHIFT-SPACE GALAXY CLUSTERING

YING ZU,1,2 ZHENG ZHENG,3,4,5 GUANGTUN ZHU,1,6 AND Y. P. JING1
Received 2007 December 20; accepted 2008 June 13

ABSTRACT

Galaxy formation inside dark matter halos, as well as the formation of the halo itself, can be affected by the large-scale environment. Evaluating the imprint of environmental effects on galaxy clustering is crucial for constraining precise cosmological constraints from galaxy redshift survey data. We investigate the environmental impact on both real-space and redshift-space galaxy clustering statistics using a semianalytic model (SAM) derived from the Millennium Simulation. We compare clustering statistics from original SAM galaxy samples to shuffled samples with the environmental influences on galaxy properties eliminated. Among the luminosity-threshold samples examined, the one with the lowest threshold luminosity (−0.2L*) is the most affected by environmental effects, having a ~10% decrease in the real-space two-point correlation function (2PCF) after shuffling. By decomposing the 2PCF into five different components based on the source of pairs, we show that the change in the 2PCF can be explained by the age and richness (galaxy occupation number) dependence of halo clustering. The 2PCFs in redshift space are found to change in a similar manner after shuffling. If the environmental effects are neglected, halo occupation distribution modeling of the real-space and redshift-space clustering may have a less than 6.5% systematic uncertainty in constraining σ8Ωm0.6 from the most affected SAM sample, and substantially smaller uncertainties for the other, more luminous samples. We argue that the effect could be even smaller in reality. In the Appendix, we present a method to decompose the 2PCF, which can be applied to measure the two-point autocorrelation functions of galaxy subsamples in a volume-limited galaxy sample and their two-point cross-correlation functions in a single run using only one random catalog.

Subject headings: cosmology: theory — dark matter — galaxies: formation — galaxies: halos —
large-scale structure of universe

Online material: color figures

1. INTRODUCTION

Recently, many authors have identified the environmental impact, which manifests itself as another degree of freedom, on the clustering of halos at fixed mass. Gao et al. (2005) found that low-mass halos (M < M*) which formed earlier are more strongly clustered than their younger counterparts, while for high-mass halos, older halos with M > 10M* turn out to be less clustered than the younger ones (Wechsler et al. 2006; Jing et al. 2007; Wetzel et al. 2007), where M* is the nonlinear mass scale for collapse. This environmental dependence of halo clustering, namely “assembly bias” (Gao & White 2007), contradicts the excursion set theory (EST; Bond et al. 1991; Lacey & Cole 1993; Mo & White 1996), which predicts that an individual halo evolves without awareness of the larger environment except for its own mass when it is observed (White 1996). In this paper, we investigate the effect of halo assembly bias on modeling the real-space and redshift-space galaxy clustering statistics, and discuss the possible consequence for cosmological parameter constraints derived from these clustering statistics.

Several possible explanations are proposed to decode the halo assembly bias, by either studying detailed halo growth within high-resolution N-body simulations, or improving current excursion set theory. Wang et al. (2007) showed that the accretion of low-mass halos in dense regions is severely truncated due to tidal disruption and preheating by their massive companions, and Jing et al. (2007) suggested that the competition for accretion resources also triggers a delayed accretion phase, which results in the inverse age dependence of massive halos, while Keselman & Nusser (2007) argued that highly nonlinear effects such as tidal stripping may not be the main driver for assembly bias. On the other hand, Zentner (2007) implemented a toy model by substituting the sharp-k filter in EST with a localized configuration filter, and Sandvik et al. (2007) integrated EST with an ellipsoidal collapse model and barrier crossing of pancakes and filaments. Both theoretical trials claimed that the assembly bias for massive halos could be naturally recovered, and at least partly offset, by deserting Markovian simplification in EST. Recently, Dalal et al. (2008) showed that the assembly bias of rare massive halos is expected from the statistics of peaks in Gaussian random fields, and they argued that the formation of a nonaccreting subpopulation of low-mass halos is responsible for the assembly bias of low-mass halos (see also Hahn et al. 2008).

As the products of gas physics within dark matter halos, galaxies have no reason to be immune from this environmental effect. Croton et al. (2007) and Zhu et al. (2006) showed that environmental effects are transmitted to the clustering and properties of galaxies in semianalytic model (SAM) and smoothed particle hydrodynamics (SPH) simulations, both of which extract halo merging histories directly from simulations rather than Markovian process. Observationally, Yang et al. (2006) and Berlind et al. (2006) found a residual dependence of galaxy clustering on group properties other than group mass by using group catalogs from the Two Degree Field Galaxy Redshift Survey (2dFGRS;
Colless et al. (2001) and the Sloan Digital Sky Survey (SDSS; York et al. 2000), respectively.

In modeling the galaxy clustering, the halo occupation distribution (HOD), or the closely related conditional luminosity function (CLF), is a powerful method for setting the observed galaxy clustering in an informative form to describe the relation between galaxies and dark matter halos (Jing et al. 1998; Seljak 2000; Peacock & Smith 2000; Scoccimarro et al. 2001; Cooray & Sheth 2002; Berlind & Weinberg 2002; Yang et al. 2003; Zheng et al. 2005). It successfully explains the departure from a power-law in the observed galaxy two-point correlation functions (Zehavi et al. 2005) and bridges the gap between high-resolution N-body simulations of dark matter particles and large-scale surveys of galaxies. HOD modeling also enhances the power of galaxy clustering to constrain cosmological parameters by linking galaxies to dark matter halos and using the clustering data on all scales (van den Bosch et al. 2003; Abazajian et al. 2005; Zheng & Weinberg 2007). However, one key assumption in the current version of the HOD is based on the EST, that the formation and the distribution of galaxies within halos are statistically determined solely by halo mass. Therefore, it is important to quantify the environmental effect on modeling galaxy clustering within the HOD framework in the era of precision cosmology and to provide insights to improve the HOD modeling.

One natural way to study the environmental effect is to extend the current HOD framework by including a second halo variable in addition to the halo mass, and compare the modeling results with previous results. Many candidates for halo variables, such as formation time, concentration, substructure richness, and spin, have been scrutinized, but all proved incapable of capturing the environmental effect neatly and completely (Gao & White 2007; Wechsler et al. 2006). One reason for this is that halo formation history is subject to incidental merging events and uneven accretion phases, both of which produce a large scatter in the relation between any halo property and the environment (Wechsler et al. 2002; Zhao et al. 2003).

In the present study, we shuffle a semianalytic galaxy sample to produce three sets of artificial samples, which either partly or completely lose their environmental features, and investigate the changes in real-space and redshift-space galaxy clustering statistics. The shuffling enables us to see the consequences of neglecting the environmental dependence in the current version of HOD modeling and gives us an idea of the effect of constraining cosmological parameters using these statistics (e.g., Tinker et al. 2006). The structure of the paper is as follows. In § 2, we introduce the simulation and the SAM model we use and describe our construction of galaxy samples with different threshold luminosities from the SAM. In § 3, we present three methods of shuffling the galaxy samples, with the aim of eliminating the environmental dependence. Then, in § 4, we analyze in detail the effect of environments on the real-space two-point correlation functions (2PCFs) by comparing the results between samples before and after shuffling. In § 5, we study the effect of environments on the redshift-space clustering statistics. We conclude in § 6 with a brief discussion and summary. In the Appendix, we present a method to decompose the 2PCFs into different components based on the properties of galaxy pairs. This method can be generalized to apply to real data to measure the two-point autocorrelation functions of galaxy subsamples in a volume-limited galaxy sample and their two-point cross-correlation functions in a single run using only one random catalog.

### 2. SIMULATION DATA AND SEMIANALYTIC MODEL

In this study, we make use of outputs from a galaxy formation model based on the Millenium Simulation. The Millenium Simulation (Springel 2005) follows the hierarchical growth of dark matter structures from redshift $z = 127$ to the present. The simulation adopts a concordance cosmological model with $\Omega_m = 0.25$, $\Omega_{\Lambda} = 0.75$, $\Omega_b = 0.045$, $\sigma_8 = 0.9$, and $h = 0.73$, and employs 2160$^3$ particles of mass $8.6 \times 10^8$ $h^{-1}$ $M_{\odot}$ in a periodic box with a comoving size 500 $h^{-1}$ Mpc on a side. Friends-of-friends (FOF; Davis et al. 1985) halos are identified in the simulation at each of the 64 snapshots with a linking length 0.2 times the mean particle separation. Substructures are then identified by the SUBFIND algorithm as locally overdense regions in the background FOF halos (Springel et al. 2001). Detailed merger trees starting from all gravitationally self-bound dark matter clumps constructed from this simulation provide a key ingredient for semianalytic models of galaxy formation.

The galaxy catalog we use is from the semianalytic model (SAM) of De Lucia & Blaizot (2007), which is an updated version of that of Croton et al. (2006) and De Lucia et al. (2006). This model explores a variety of physical processes related to galaxy formation. It can reproduce many observed properties of galaxies in the local universe, including the galaxy luminosity function, the bimodal distribution of colors, the Tully-Fisher relation, the morphology distribution, and the 2PCFs for various type and luminosity selected samples. This particular model is of course not guaranteed to be absolutely correct. What is important to our study is that the environment-dependent ingredients inherent in this model, such as the history of dynamical interactions and mergers of halos, should be adequately transmitted to and preserved in the resulting galaxy population. We aim to investigate the likely effects of the environmental dependence in this model on galaxy clustering statistics in real and redshift space, and to explore the implications for cosmological studies conducted with galaxy clustering data.

We construct six luminosity-threshold galaxy samples at $z = 0$ from the SAM catalog according to the rest-frame SDSS $r$-band absolute magnitude $M_r$, with dust extinction included. Table 1 lists the properties of these samples. Our L207 sample has a number density similar to the observed $L > L_*$ sample (see Table 2 of Zehavi et al. 2005). Since more luminous galaxies tend to reside in more massive halos (e.g., Zehavi et al. 2005), these six samples can probe different halo mass ranges (from mass below $M_r$ to above $M_r$). The halo assembly bias has different amplitudes and signs across these mass ranges (e.g., Gao et al. 2005; Gao & White 2007; Wechsler et al. 2006; Jing et al. 2007); we therefore expect different environmental effects from the six samples.

### 3. SHUFFLING SCHEMES

Our purpose in this paper is to study the impact of the environmental dependence on the HOD modeling of real-space and redshift-space clustering statistics. In addition to the six galaxy samples from the SAM, for comparison we also need galaxy
samples with the environmental dependence eliminated. Following Croton et al. (2007), we construct such control samples from the original SAM catalog by shuffling the galaxy content in halos of similar masses. We produce three sets of control samples based on the three shuffling schemes described below.

We first group all the FOF dark matter halos at \( z = 0 \) with \( M_{\text{vir}} \) larger than \( 5.5 \times 10^{10} \, h^{-1} M_{\odot} \) in the catalog into different mass bins of width \( \Delta \log M_{\text{vir}}/(h^{-1} M_{\odot}) = 0.1 \). Then we record the relative positions and velocities of all the satellites to their affiliated central galaxies, whose positions and velocities are set to those of their host halos in the SAM. Finally, we redistribute the galaxies within individual halo mass bins. The three sets of control samples (hereafter CTL1, CTL2 and CTL3, respectively) differ in the way the galaxies are redistributed.

For CTL1, we follow the scheme of Croton et al. (2007) in keeping the original configuration of galaxies inside each halo intact and moving the galaxy content to its new host as a whole. In this way, the one-halo term contribution to the galaxy clustering statistics is almost unchanged. In order to compensate for the nonzero mass bin effect in the shuffling, we scale the recorded relative position and velocity of each galaxy by \( (M_{\text{new}}/M_{\text{old}})^{1/3} \) in order to redistribute the galaxies in the original halo of mass \( M_{\text{old}} \) to the new host halo of mass \( M_{\text{new}} \). This improvement ensures that the position of the shuffled galaxy content is regulated by the virial radius of the new host halos.

For CTL2, we collect the distance \( r \) to the halo center and the velocity \( v \) relative to the halo center for all the satellite galaxies that belong to halos in the same mass bin. Then a pair of \( r \) and \( v \) are randomly drawn from the sets and assigned to a galaxy. This galaxy is put into a randomly selected halo in that mass bin with random orientations for both \( r \) and \( v \) [with the \( (M_{\text{new}}/M_{\text{old}})^{1/3} \) scaling applied]. The central galaxies are randomly assigned to halos of the same mass bin. This shuffling procedure assumes a mean radial galaxy number density profile for all halos in the same mass bin and completely eliminates any environmental features in the galaxy distribution inside halos, including the alignment and segregation of satellites, the nonspherical shape of halos, the infall pattern of the satellite velocity distribution, and any correlation between central and satellite galaxies (e.g., in luminosity). CTL2 should allow us to infer the largest effect that environment may have on galaxy clustering statistics for the given SAM.

In addition to CTL1 and CTL2, we construct another set of samples (CTL3) by isotropizing satellites inside their own halo without shuffling content between different halos. In this way, the radial distribution of galaxies in each individual halo is conserved, but the statistical angular distribution loses its anisotropy. CTL3 allows us to isolate the effect of assuming spherical symmetry for the satellite distribution in modeling 2PCFs. Although CTL2 also isotropizes the satellite distribution inside halos, it effectively uses a radial distribution averaged over halos of similar mass. Therefore, comparing CTL3 and CTL2 would show the effect of the scatter in the distribution of satellites in halos of similar mass.

For each of the CTL1, CTL2, and CTL3 shuffling schemes, we create 10 different galaxy catalogs, varying the random seed. We extract the 6 control luminosity threshold samples from each shuffled catalog in accordance with the L190, L200, L207, L210, L217, and L220 samples of the SAM. To prevent numerical effects from mixing with the physical effects we want to ascertain, we have performed tests by reducing the size of halo mass bins or leaving several of the most massive bins unshuffled, and we find that our choice of bin size does not introduce any noticeable numerical effect.

4. ENVIRONMENTAL EFFECT ON REAL-SPACE 2PCFs

We start by comparing the real-space 2PCFs of the original SAM samples and the shuffled samples. The 2PCFs essentially describe the pair count as a function of pair separation. On small (large) scales, galaxy pairs are dominated by one-halo (two-halo) pairs, i.e., intrahalo (interhalo) pairs.

Our results for the real-space 2PCFs are shown in Figure 1. On small scales \((\lesssim 2 \, h^{-1} \text{Mpc})\), where the one-halo term dominates, 2PCFs from CTL1, CTL2, and CTL3 behave differently. In CTL1, the galaxy contents inside halos are exchanged among halos as a whole, so we do not expect any appreciable change in the one-halo regime of the 2PCFs. Thus, the small-scale clustering in CTL1 remains almost the same as that in the original sample, as seen in Figure 1. The slight differences seen in the plot are a result of the finite mass bin. CTL3 makes the satellite distribution inside halos isotropic, which on average enlarges the separation of intrahalo galaxy pairs. Thus, on small scales, the 2PCFs of CTL3 are always smaller than those of SAM, with a suppression of around \(10\%\). In CTL2, not only is the angular distribution of satellite galaxies inside halos isotropized, but the radial galaxy number density profile is averaged within the same mass bin, which completely erases any environmental memory of galaxies. Figure 1a shows that galaxies in CTL2 exhibit a suppression of up to \(\sim 10\%\) for L190 samples; the suppression becomes weaker for samples with higher threshold luminosity (e.g., sample L210 in Fig. 1b). The 2PCF for the L217 CTL2 sample (Fig. 1c) is too noisy to see the trend, but it is likely to still be a suppression (see below).

On large scales, where the two-halo term dominates, 2PCFs from CTL3 stay the same as those from the original SAM samples, since the CTL3 scheme only shuffles galaxies in each halo. It is not surprising that the large-scale 2PCFs of CTL1 and CTL2 are almost identical, given that they both shuffle halos of similar mass. The difference between the 2PCF in the CTL1/CTL2 sample and the original SAM sample shows a steady trend with the threshold luminosity. For the faint sample \(M_r < -19\), the environmental dependence of the galaxy population and the assembly bias lead to a \(\sim 10\%\) suppression in the 2PCF after shuffling (Fig. 1a). For the intermediate sample \(M_r < -21\), the difference between shuffled and SAM samples is reduced to \(\sim 2\%\) (Fig. 1b). For the bright \(M_r < -21.7\) sample, the 2PCFs of shuffled samples become \(\sim 3\%\) larger than those of the SAM sample (Fig. 1c).

To see the trend more clearly, we show in Figure 2 the ratio \(\xi/SAM\) of large-scale 2PCFs of the shuffled (CTL1 or CTL2) and the SAM samples as a function of the magnitude limit. The ratio \(\xi/SAM\) is calculated by averaging the measurements for each 10 shuffled sample on scales of \(5-25 \, h^{-1} \text{Mpc}\). We note that the trend of the ratio with the threshold luminosity is the same as in Figure 2 of Croton et al. (2007), although we use a different indicator for the large-scale difference.

For a better understanding of the change of clustering strength in the shuffled samples with respect to the original samples, we decompose the galaxy 2PCFs into five components according to the source of galaxy pairs and examine them individually. The five components are denoted as \(1h-\text{cen-sat}, 1h-\text{sat-sat}, 2h-\text{cen-sat}, 2h-\text{sat-sat}\), and \(2h-\text{cen-cen}\), where \(1h\) and \(2h\) refer to one-halo and two-halo pairs, and cen and sat denote the nature (central or satellite galaxies) of the pair of galaxies. That is, we have central galaxy in a halo paired with satellites in the same halo (1h-\text{cen-sat}), satellite galaxy pairs inside halos (1h-\text{sat-sat}), central galaxy in one halo paired with central galaxy in a different halo (2h-\text{cen-cen}), central galaxy in one halo paired with satellites in a different halo (2h-\text{cen-sat}), and
satellites in one halo paired with satellites in a different halo \((2h\text{-sat-sat})\). A detailed description of how we separate these components can be found in the Appendix.

Figure 3 shows the five 2PCF components of the original and shuffled samples for L190, L210, and L217 (left, middle, and right columns for CTL1, CTL2, and CTL3, respectively). Since the spatial distribution of satellites inside halos is conserved in the CTL1 sample [except for the \((M_{\text{new}}/M_{\text{old}})^{1/3}\) scaling], there is almost no change in the one-halo components for this sample with respect to the original sample. In the shuffling schemes of the CTL2 and CTL3 samples, satellites inside halos are angularly redistributed from a nonspherical distribution to an isotropic distribution. The redistribution in either CTL2 or CTL3 does not change the separations of one-halo central-satellite galaxy pairs, so the \(1h\text{-cen-sat}\) component does not change after shuffling, as seen in Figure 3. However, the isotropization statistically increases the separations of one-halo satellite-satellite pairs, and thus dilutes the \(1h\text{-sat-sat}\) clustering signal. This leads to a suppression of the 2PCF on small scales with respect to the original sample (e.g., \(-10\%\) for L190). The CTL2 samples show a smaller suppression in the \(1h\text{-sat-sat}\) component than the CTL3 samples. There may be two reasons for the difference. First, CTL2 effectively uses a mean radial distribution profile of satellites in halos of a given mass, while CTL3 uses the radial distribution in each individual halo. Because of the scatter in the radial profiles at a given halo mass, the distributions of one-halo satellite-satellite pair separations are not identical from the mean and individual profiles. Second, CTL2 ensures that the number of satellites inside halos of a given mass follows a Poisson distribution, while CTL3 follows the distribution in the SAM sample, which can be slightly sub-Poisson in the low-occupation regime (e.g., Zheng et al. 2005).

For the shuffled samples CTL1, CTL2, and CTL3, all two-halo components (except \(2h\text{-cen-cen}\)) show enhancements on scales less than \(1\ h^{-1}\) Mpc with respect to the original ones. This is mainly caused by the nonspherical distribution of satellite galaxies inside halos in the SAM sample. The shuffling procedure can cause the satellite populations of two neighboring (nonspherical) halos to become spatially close or even overlap to some extent. Therefore, in the shuffled samples the probability of finding close interhalo galaxy pairs that involve satellites increases. However, such small-scale enhancements in the two-halo components only occur on scales at which the one-halo satellite galaxies inside halos is conserved in the CTL1 sample [except for the \((M_{\text{new}}/M_{\text{old}})^{1/3}\) scaling], there is almost no change in the one-halo components for this sample with respect to the original sample. In the shuffling schemes of the CTL2 and CTL3 samples, satellites inside halos are angularly redistributed from a nonspherical distribution to an isotropic distribution. The redistribution in either CTL2 or CTL3 does not change the separations of one-halo central-satellite galaxy pairs, so the \(1h\text{-cen-sat}\) component does not change after shuffling, as seen in Figure 3. However, the isotropization statistically increases the separations of one-halo satellite-satellite pairs, and thus dilutes the \(1h\text{-sat-sat}\) clustering signal. This leads to a suppression of the 2PCF on small scales with respect to the original sample (e.g., \(-10\%\) for L190). The CTL2 samples show a smaller suppression in the \(1h\text{-sat-sat}\) component than the CTL3 samples. There may be two reasons for the difference. First, CTL2 effectively uses a mean radial distribution profile of satellites in halos of a given mass, while CTL3 uses the radial distribution in each individual halo. Because of the scatter in the radial profiles at a given halo mass, the distributions of one-halo satellite-satellite pair separations are not identical from the mean and individual profiles. Second, CTL2 ensures that the number of satellites inside halos of a given mass follows a Poisson distribution, while CTL3 follows the distribution in the SAM sample, which can be slightly sub-Poisson in the low-occupation regime (e.g., Zheng et al. 2005).

For the shuffled samples CTL1, CTL2, and CTL3, all two-halo components (except \(2h\text{-cen-cen}\)) show enhancements on scales less than \(1\ h^{-1}\) Mpc with respect to the original ones. This is mainly caused by the nonspherical distribution of satellite galaxies inside halos in the SAM sample. The shuffling procedure can cause the satellite populations of two neighboring (nonspherical) halos to become spatially close or even overlap to some extent. Therefore, in the shuffled samples the probability of finding close interhalo galaxy pairs that involve satellites increases. However, such small-scale enhancements in the two-halo components only occur on scales at which the one-halo
term of the 2PCF dominates; thus, they are of no interest in our analysis.

On large scales, where the two-halo term regime dominates, the two-halo components for CTL3 do not change, since it only shuffles galaxies within halos, while every two-halo component changes its amplitude after shuffling with CTL2 and CTL3. There is no doubt that this should be a manifestation of the environmental dependence of the halo clustering and that of the galaxy content inside halos. Let us first consider the $2h$-cen-cen component. The effect of shuffling central galaxies is equivalent to that of shuffling halos. If the galaxy sample were a halo-mass-threshold sample, shuffling would not change the large-scale clustering of central galaxies, as the host halo population remains the same after shuffling. However, the sample we consider is defined by a threshold in luminosity, and it is not a halo-mass-threshold sample because of the scatter between halo mass and central galaxy luminosity. At a fixed mass, older halos tend to host more luminous central galaxies, and the mean central galaxy luminosity is an increasing function of halo mass (Zhu et al. 2006). We thus expect that, at a given luminosity, a central galaxy can reside in a low-mass older halo or in a younger halo of higher mass. That is, for low-mass halos, only a fraction (some older ones) can host the galaxies in our luminosity threshold sample. Since the shuffling is among halos of the same mass, the central galaxies of the sample in these low-mass older halos are moved to younger halos in the same mass bin after shuffling. For the L190 samples, these low-mass halos are in the regime where the clustering of younger halos is weaker, so we see a decrease in the $2h$-cen-cen component of the 2PCF after shuffling (Fig. 3[1] and [2]). However, halos at the low-mass end in L217 samples are in the regime where the clustering of older halos is weaker, leading to an increase in the $2h$-cen-cen component after shuffling (Fig. 3[7] and [8]). For the L210 samples, the low-mass halos are in the regime where the age dependence of halo clustering almost disappear, and as a consequence, the $2h$-cen-cen component does not change much after shuffling (Fig. 3[4] and [5]).

Unlike the $2h$-cen-cen component, for which the effect of shuffling is determined by the halos near the low-mass end for the given sample, the $2h$-cen-sat and $2h$-sat-sat components...
are influenced by all halos above the low-mass end. Zhu et al. (2006) find that, in general, at a fixed halo mass, there are fewer satellite galaxies in older halos. Combining this finding with the age dependence of halo clustering (i.e., older halos being more strongly clustered in the mass range appropriate for our sample), one would infer that shuffling would increase the amplitudes of the 2h-cen-sat and 2h-sat-sat components, since the overall effect of shuffling is to homogenize satellite populations among halos of different ages (i.e., to increase/decrease the number of satellites in older/younger halos). However, this naive expectation is contradictory to what is seen in Figure 3. Then, what is the reason for the suppression in the 2h-cen-sat and 2h-sat-sat components? The answer lies in the richness dependence of halo clustering. Here, the term “richness” refers to the subhalo/substructure/satellite abundance in a halo. Gao & White (2007) show that, in the mass range relevant here, halos with more substructures are always more strongly clustered. Since substructures are the natural dwellings of satellite galaxies, we expect that, at a fixed mass, halos that have more satellites will be more strongly clustered. The effect of shuffling is to move some satellites in strongly clustered halos to weakly clustered halos, and thus lower the amplitude of the 2h-cen-sat and 2h-sat-sat components of the 2PCF.

The above explanation of the two-halo term change still leaves one question. According to the age dependence of halo clustering (older halos are more strongly clustered) and the anticorrelation between age and subhalo abundance (older halos have fewer subhalos; Gao et al. 2004; Zhu et al. 2006), one would expect halos with fewer satellites to be more strongly clustered, in sharp contrast to what is found in simulations (e.g., Gao & White 2007). The solution to the apparent contradiction lies in the scatter in the anticorrelation between age and richness, and the joint dependence of halo clustering on age and richness (Y. Zu et al., in preparation).

Figure 4 summarizes the contributions to the large-scale 2PCFs and the changes caused by CTL1/CTL2 shuffling as a function of threshold luminosity. We plot the contributions from different two-halo components to the large-scale bias factor (squared) for both the original SAM samples (solid lines) and CTL1 samples (dashed lines). Each component contribution to the square bias factor is computed by averaging the ratio of the corresponding two-halo 2PCF component (2h-cen-cen, 2h-cen-sat, or 2h-sat-sat) to the matter 2PCF on scales of 5–15 h−1 Mpc. For galaxy samples with low threshold luminosity, the largest contribution to the large-scale clustering comes from the 2h-cen-sat component. The 2h-cen-cen component then takes over for samples with threshold luminosity around L*, and becomes more and more dominant toward higher luminosity. This trend can be understood by noticing that the satellite fraction decreases with increasing threshold luminosity (e.g., Zheng et al. 2007). The 2h-sat-sat component always has the least contribution to the large-scale clustering.

Figure 4 shows that shuffling causes the 2h-cen-cen component to be suppressed slightly for samples with low luminosity thresholds and to be enhanced a little bit for samples with high luminosity thresholds, a trend can be explained by the age dependence of halo clustering, as discussed above. Here “low” and “high” are with respect to L*. We note that the change in the 2h-cen-cen component decreases again at the very high luminosity end (i.e., the L220 sample), similar to the trend seen in the concentration dependence of massive halo clustering (see, e.g., Jing et al. 2007). The 2h-cen-sat component is always suppressed after shuffling, which can be understood by the richness dependence of halo clustering, as mentioned above. The change of the overall large-scale bias factor is dominated by that of the 2h-cen-sat at low luminosity and very high luminosity. The change of the 2h-cen-cen component plays a role in determining that of the overall bias factor for samples with luminosity threshold larger but not much larger than L*, and it nearly compensates for the suppression caused by the change in the 2h-cen-sat component, leading to little change (<2%) in the overall bias factor (see also Fig. 2).

5. ENVIRONMENTAL EFFECT ON REDSHIFT-SPACE 2PCFs

While the 2PCFs in real space are isotropic, the 2PCFs in redshift space are distorted by galaxy peculiar velocities along the line of sight. On small scales, the random virialized motions of galaxies in groups and clusters stretch the redshift distribution of galaxies along the line of sight, producing the so-called “finger-of-god” (FOG) effect. On large scales, the coherent flows of galaxies due to gravity squish the line-of-sight distribution of galaxies (i.e., the Kaiser effect; Kaiser 1987).

In linear theory, the large-scale redshift-space distortion measures a combination of Ωm and the large-scale galaxy bias factor b_g, which is Ωm/Ω_g. Given the measured amplitude of the galaxy 2PCF, a constraint on Ω_m/Ω_g is equivalent to that on σ_g/Ω_m, where σ_g is the rms matter fluctuation on scales of 8 h−1 Mpc. To infer such a constraint on large scales, the small-scale redshift distortion is usually dealt with by simple models, such as the exponential model (Cole et al. 1995). Tinker et al. (2006) demonstrates that by taking advantage of the power of HOD to describe clustering in a fully nonlinear manner, one can consistently model the small-scale and the large-scale clustering (also see Tinker 2007). Furthermore, Tinker et al. (2006) shows that the degeneracy in Ω_m and σ_g from large-scale clustering can be broken by making use of the small- and intermediate-scale clustering in redshift space. The HOD framework used in Tinker et al. (2006) assumes no environmental effects on halo clustering or galaxy content inside halos. In § 4, we have shown that in the SAM we
use, the real-space 2PCFs may suffer a change up to ~10% from environmental effects. It is interesting to perform similar analysis in redshift space and discuss the implications in inferring cosmological parameters from redshift distortions.

In Figure 5, we show the redshift-space 2PCFs $\xi(r_p, r_z)$ measured from the SAM, CTL1, CTL2, and CTL3 catalogs for the three luminosity threshold samples as in Figure 1, where $r_p$ and $r_z$ are the perpendicular and line-of-sight distances in redshift space. The overall effect of the shuffling on the redshift-space 2PCFs is similar to that seen in the real-space 2PCFs. On small scales, where the FOG effect dominates, the clustering amplitudes of CTL1 samples are almost identical to those of the SAM samples, since shuffling does not change the one-halo term in CTL1. For the CTL2 samples, the FOG effect is slightly suppressed, and the small-scale clustering is weaker than that of the SAM for L190, but nearly identical to and stronger than that of the SAM for L210 and L217, respectively. For CTL3 samples, although they exhibit a large difference in real-space 2PCFs, the FOG effect only changes a little, which indicates that the difference caused by galaxy angular distribution is partly masked by the peculiar velocity field on small scales.

On large scales, similar to what is seen in the real-space 2PCFs, at a given large-scale separation ($r_p$, $r_z$), the redshift-space 2PCF of CTL1 or CTL2 sample has a lower amplitude than that of the L190 SAM sample, an almost identical amplitude to the L210 SAM sample, and exceeds that of the L217 SAM sample. The 2PCF amplitudes do not change with CTL3 samples. As a whole, shuffling introduces changes similar to those in the real-space 2PCFs, and these changes can be understood following interpretations in § 4.

To further quantify changes in the redshift-space 2PCFs, we calculate a few statistics derived from the multipoles of the real-space and redshift-space 2PCFs, which were originally proposed by Hamilton (1992). These statistics are also the ones used in the studies of Tinker et al. (2006) and Tinker (2007) for HOD modeling the redshift-space distortion. The multipole moments $\xi(r)$ are given by the coefficients of the Legendre polynomial expansion of $\xi(r_p, r_z)$,

$$\xi(r) = \frac{2l + 1}{2} \int_{-1}^{+1} \xi(r_p, r_z) P_l(\mu) d\mu,$$  \hspace{1cm} (1)

where $r = (r_p^2 + r_z^2)^{1/2}$, $\mu = r_z/r$, and $P_l(\mu)$ is the $l$th order Legendre polynomial. Based on the multipoles of $\xi(r_p, r_z)$, we calculate two statistics. The first is the ratio of the monopole $\xi_0(r)$ to the real-space 2PCF $\xi_b(r)$,

$$\xi_0/b(r) = \xi_0(r) / \xi_b(r).$$  \hspace{1cm} (2)

In linear theory, it is a function of $\beta \equiv \Omega_m^{0.6}/b_g$,

$$\xi_0/b(r) = 1 + \frac{2}{3} \beta + \frac{1}{5} \beta^2.$$  \hspace{1cm} (3)

The second quantity $Q_z(r)$ is related to the quadrupole $\xi_2(r)$,

$$Q_z(r) = \frac{\xi_2(r)}{\xi_0(r) - \xi_0(r)},$$  \hspace{1cm} (4)

7 We use the same symbol $\xi$ for both the real-space and redshift-space 2PCFs. Whenever it introduces a confusion, we add a subscript $R$ for real-space quantities.

![Figure 5. Comparison of redshift-space correlation functions between the SAM and shuffled samples. Panels (a), (b), (c) show the L190, L210, and L217 samples, respectively. In each panel, three quadrants show the comparison between the SAM sample (solid contours) and the shuffled samples CTL1, CTL2, and CTL3 (dotted contours). Contour levels are set as $2^n$, with $n$ from −5 to 2. (See the electronic edition of the Journal for a color version of this figure.)](image-url)
Fig. 6.— Comparison of redshift-space clustering statistics between the SAM and shuffled samples for the L190 galaxy sample. Top: $\xi_{0/R}$, the ratio of the monopole of the redshift-space 2PCF to the real-space 2PCF; middle: $Q_t$, which is related to the quadrupole of the redshift-space 2PCF; bottom: $r_{1/2}$, which is the value of $r_p$ at which the redshift-space 2PCF at a give $r_p$ decreases by a factor of 2 relative to its value at $r_p = 0$. In each panel, solid curves show the SAM sample, and dashed, dot-dashed, and dotted curves show the CTL1, CTL2, and CTL3 shuffled samples, respectively. The horizontal axes are shown in linear (logarithmic) scales in the left (right) panels to highlight the large (small) scale behavior. Error bars are plotted only for the SAM sample to avoid crowding; those for the shuffled samples are comparable. The abnormal error bars around $10^{1} h^{-1}$ Mpc in the middle panels of (c) and (d) are because $Q_t$ here approaches zero. [See the electronic edition of the Journal for a color version of this figure.]
where $\xi_0(r)$ is the volume-averaged monopole,
\[
\xi_0(r) = \frac{3}{r^3} \int_0^r \xi_0(s) s^2 ds.
\] (5)

In linear theory, $Q_\ell(r)$ is also a function of $\beta$,
\[
Q_\ell(r) = \frac{\beta + \frac{3}{2} \beta^2}{1 + \frac{1}{2} \beta + \frac{1}{2} \beta^2}.
\] (6)

Tinker et al. (2006) also introduce a quantity $r_{g2}$, which is the value of $r_\sigma$ at which the redshift-space 2PCF at the given $r_\sigma$ decreases by a factor of 2 with respect to the value of 2PCF at $r_\sigma = 0$. We also compute this quantity.

Figure 6 plots $\xi_{0R}$ and $Q_\ell$ as a function of $r$ and $r_{g2}$ as a function of $r_p$ for the L190 sample using both linear and logarithmic axes to highlight large and small scales separately. According to the above results, this sample, among the six luminosity threshold samples, is expected to show the largest environmental effect (we also check the luminosity threshold sample with magnitude limit $M_r = -18.0$, and find that the large-scale suppression is still at the 10% level, as it is in L190).

Compared to the SAM sample, the monopole term $\xi_{0R}$ is only $\sim 2\%$ higher on large scales in CTL1 and CTL2, while it stays the same in CTL3. On small scales (0.1–1 $h^{-1}$ Mpc), the difference is at a level of 5% in CTL1 and CTL2. Only on extremely small scales ($\leq 0.1$ $h^{-1}$ Mpc) does $\xi_{0R}$ of the CTL2 sample show a 10% drop, which is not important, since the error bars are large and these scales are likely to be excluded in HOD modeling. In the CTL3 sample, since the suppression of the spherically averaged redshift-space 2PCFs and that of the real-space 2PCFs cancel each other on small scales, $\xi_{0R}$ stays at the same level as in the SAM sample.

For the quadrupole term $Q_\ell$, the difference between the results of the original and shuffled samples is well within 5.5% in CTL1 and CTL2 for most scales. Note that the difference extends all the way to the largest scales in CTL3 at a 0.5% level, which means that the nonlinearity still affects clustering behavior on linear scales in redshift space. Also note that the large fractional differences around 10 $h^{-1}$ Mpc are simply because $Q_\ell$ is crossing zero. The fractional changes in $\xi_{0R}$ and $Q_\ell$ caused by shuffling are much less than those in the real-space 2PCFs, which are at a level of 10%. For the quantity $r_{g2}$, the global enhancement in CTL3 is caused only by the angular isotropizing of galaxies, which disrupts the original compact configuration of SAM halos. This makes it more difficult for the redshift-space 2PCFs to decrease to the half value of $\xi_{1r=0}$ at a given $r_\sigma$, especially at $r_\sigma < 0.2$ $h^{-1}$ Mpc, where the enhancement becomes much more prominent. In CTL2, $r_{g2}$ shows behavior similar to that in CTL3 at small $r_\sigma$, and for the same reason; then it becomes smaller than in CTL3 at $r_\sigma > 0.2$ $h^{-1}$ Mpc, while $r_{g2}$ from the CTL1 shuffled sample is consistent with that from the original sample within the error bars.

How would the above changes in the redshift distortion statistics induced by the environmental effect affect the inference of cosmological parameters from HOD modeling? For a complete answer to this question, one needs to perform the analysis presented in Tinker (2007) with the 2PCF measurements in the original and shuffled samples and compare the results in the inferred cosmological parameters. However, even without the full analysis, we can still figure out the likely magnitude by using the linear theory results (eqs. [3] and [6]). For the L190 sample presented in Figure 6, a 2% increase in the large-scale $\xi_{0R}$ only leads to a $\sim 6.5\%$ increase in the inferred $\beta$. For $Q_\ell$, shuffling gives rise to a 5.5% increase on large scales, which also translates to a $\sim 6.5\%$ increase in $\beta$. For $r_{g2}$, it is not straightforward to see the consequence. Based on Figure 16 of Tinker et al. (2006), it is likely that the difference in $r_{g2}$ on large scales between the SAM and shuffle samples can at most lead to a $\sim$5% uncertainty in constraining $\sigma_8$. We note that the effects of the shuffling on both the real-space and redshift-space clustering statistics are likely from the same cause, in that the $\sim 12\%$ decrease in the real-space 2PCF $\xi_0(r)$ leads to a $6\%$ decrease in galaxy bias $b_g$, which in turn corresponds to a $\sim 6.5\%$ increase in $\beta$, about the number we infer from redshift-space clustering statistics. Since the shuffling-induced changes in the real-space 2PCFs for other brighter samples are smaller, the environmental effect on $\beta$ is expected to be smaller for them. Therefore, in the SAM galaxy catalog we use, neglecting any environmental dependence of halo clustering and galaxy formation is likely to cause a less than 7% systematic uncertainty in constraining $\sigma_8 \Omega_m^{0.6}$.

6. SUMMARY AND DISCUSSION

In this work, we investigate the effect of the environmental dependence of halo clustering and galaxy formation on real-space and redshift-space clustering of galaxies. Our study makes use of the galaxy catalog from the SAM of De Lucia & Blaizot (2007), which is based on the Millennium Simulation (Springel 2005). The inherent dependence of galaxy properties on environment in the SAM catalog is eliminated by shuffling galaxies among halos of similar mass.

The real-space 2PCFs in the original sample and those in the shuffled samples have a difference of $\sim 10\%$, with some dependencies on scales for samples with low threshold luminosities. The difference becomes much smaller for samples with threshold luminosities approaching or exceeding $L_*$. We decompose the 2PCFs into five components by accounting for the nature of galaxy pairs (e.g., one-halo or two-halo, central galaxies or satellites) and study the effect of environment on each. In general, on large scales, the changes in the 2h–cen–cen component of the 2PCF caused by shuffling can be well understood by noting the dependence of halo bias and central galaxy luminosity on halo formation time, while those in the 2h–cen–sat and 2h–sat–sat components are determined by the richness (substructure) dependence of halo bias. The 2h–cen–sat component appears to dominate the change in the overall 2PCF for samples with low or very high threshold luminosity, while the change in the 2h–cen–cen component nearly compensates for that in the 2h–cen–sat component for threshold luminosity $L_*$, where the amplitude of the environment effect is small. These results imply that we could use high-resolution N-body simulations to accurately model the 2PCFs by associating satellites with substructures identified in halos and neglecting the environmental dependence of central galaxies at fixed halo mass. On small scales, the assumption of spherical symmetry in galaxy distribution may lead to an uncertainty as large as 10% in 2PCFs, but this effect could be absorbed into a free parameter describing the halo concentration.

The effects of environmental dependence on redshift-space 2PCFs are similar to what are seen in the real-space 2PCFs. On large scales, the effects can be attributed solely to the change in the large-scale bias factor. For inferring cosmological parameters ($\sigma_8$ and $\Omega_m$) through HOD modeling of the redshift-space distortion (Tinker et al. 2006; Tinker 2007), the systematic effect caused by neglecting the environmental dependence of halo clustering and galaxy formation is likely to be at the level of $< 6.5\%$ for the worst case (the $M_r < -19.0$ or $L > 0.2 L_*$ sample) and can be much smaller for brighter samples, especially for samples...
with threshold luminosities near $L_\star$. The underlying assumption of this statement is that the environmental effect on galaxy formation in reality is as large as that seen in the SAM we use in this paper.

Our results are based on one particular galaxy formation model, i.e., the SAM of De Lucia & Blaizot (2007). Although this model can reproduce many observed properties of galaxies, it is not guaranteed to be absolutely correct. In this model, the environmental effect on the formation and evolution of galaxies is mostly linked to the formation and merger history of dark matter halos. Compared to observations, it overproduces faint red galaxies (Croton et al. 2006). Although the model predicts the correct trend of color-dependent galaxy clustering, it predicts too large a difference between the amplitudes of the 2PCFs of blue and red galaxies (Springel 2005). By comparison with galaxies in SDSS groups, Weinmann et al. (2006) find that SAM produces too many faint satellites in massive halos and incorrect blue fractions of central and satellites galaxies.

The discrepancies between observations and the SAM model suggest that the effect of environment on galaxy formation and evolution may be exaggerated in this particular model. Such discrepancies provide opportunities for enhancing our understanding of galaxy formation and evolution. There are also tests with void statistics (Tinker et al. 2008), galaxy-galaxy lensing (Yoo et al. 2006), and marked galaxy correlation function (Skibba et al. 2006), which show that the observed properties of galaxies are mainly driven by host halo mass rather than the environment in which halos form. Therefore, in reality, it is quite possible that the environmental effect on modeling galaxy clustering statistics is much smaller than what we obtain in this paper, and that the systematic effect on cosmological parameter constraints from HOD modeling is not larger than a few percent or even better.

We thank Jeremy Tinker for helpful discussions and David Weinberg for useful comments. Y. Z. and Y. P. J. are supported by NSF (10533030), by the Knowledge Innovation Program of CAS (No. KJCX2-YW-T05), by the 973 Program (No.2007CB815402), and by Shanghai Key Projects in Basic Research (05XD14019). At an earlier stage of this work, Z. Z. was supported by NASA through Hubble Fellowship grant HF-01181.01-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. Z. Z. gratefully acknowledges support from the Institute for Advanced Study through a John Bahcall Fellowship. The Millennium Simulation databases used in this paper and the web application providing online access to them were constructed as part of the activities of the German Astrophysical Virtual Observatory.

APPENDIX

DECOMPOSITION OF THE TWO-POINT CORRELATION FUNCTION

In the HOD framework, the 2PCF $\xi(r)$ is usually decomposed into two components (e.g., Zheng 2004),

$$\xi(r) = [1 + \xi_{1h}(r)] + \xi_{2h}(r),$$  \hspace{1cm} (A1)

where the one-halo term $\xi_{1h}(r)$ and the two-halo term $\xi_{2h}(r)$ represent contributions from intrahalo and interhalo pairs, respectively. To separate such components from measurements in a mock catalog, one only needs to weight the total correlation function appropriately. That is, $1 + \xi(r)$ weighted by the fraction of intrahalo (interhalo) pairs at a separation $r$ gives $1 + \xi_{1h}(r) [1 + \xi_{2h}(r)]$. However, the way to decompose $\xi(r)$ into more components on the basis of pair counts, as we do in this paper (central/satellite, one-halo, two-halo pairs), is not immediately clear. In this Appendix, we develop a method for such a decomposition. The method can be generalized to apply to real data: for example, one is able to measure the two-point autocorrelation functions of red and blue galaxies and their two-point cross-correlation functions in a single run with only one random catalog.

We first provide a general consideration of the component separation, and then describe the decomposition used in this paper in more details.

A1. GENERAL CONSIDERATION

Let us start from the definition of the two-point correlation function,

$$\xi(r) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle,$$  \hspace{1cm} (A2)

where $\langle \ldots \rangle$ represents an ensemble average. The overdensity field $\delta$ is defined as

$$\delta(\mathbf{x}) = \frac{n(\mathbf{x}) - \bar{n}}{\bar{n}},$$  \hspace{1cm} (A3)

where $n(\mathbf{x})$ is the galaxy density at $\mathbf{x}$ and $\bar{n}$ is the mean. Let us assume that the galaxy sample is composed of several subsamples, $n(\mathbf{x}) = \sum_i n_i(\mathbf{x})$. Now we decompose the overdensity into different components based on subsamples

$$\delta = \sum_i \delta_i,$$  \hspace{1cm} (A4)

where $\delta_i$ is the overdensity contributed by the $i$th component (subsample),

$$\delta_i = \frac{n_i - \bar{n}_i}{\bar{n}} = \frac{n_i - \bar{n}_i}{\bar{n}_i} \times \frac{\bar{n}_i}{\bar{n}} = \delta_i \frac{\bar{n}_i}{\bar{n}}.$$  \hspace{1cm} (A5)
Note that in the above equation, $\delta_i$ is the $i$th component’s own overdensity field (i.e., fractional fluctuation with respect to $\bar{n}_i$ instead of $\bar{n}$).

Substituting equations (A4) and (A5) into equation (A2), we obtain

$$\xi(r) = \sum_i \langle \delta_i(x)\delta_i(x + r) \rangle \frac{\bar{n}_i^2}{\bar{n}^2} + \sum_{i < j} \langle \delta_i(x)\delta_j(x + r) \rangle \frac{2\bar{n}_i\bar{n}_j}{\bar{n}^2}.$$  

(A6)

That is, the total correlation function is a weighted sum of the auto- and cross-correlation functions of all components, where the weight is the pair fraction. In terms of measurement from pair counts in a galaxy catalog and an auxiliary random catalog, this converts to

$$\xi(r) = \sum_{i,j} \frac{dd_{ij}(r) - rrr_{ij}(r)}{rr_{ij}(r)} f_{ij},$$  

(A7)

where $dd_{ij}$ and $rr_{ij}$ are the $ij$ data-data and random-random pairs. The quantity $f_{ij}$ is the overall $ij$ pair fraction (the ratio of the total number of $ij$ pairs in the volume to that of all pairs in the volume). Note that, for random pairs, the ratio of the number of random $ij$ pairs to that of all random pairs is independent of separation, i.e., $f_{ij} = rrr_{ij}(r)/RR(r)$, where $RR$ is the number of random pairs for all galaxies. Therefore, we have the contribution from the $ij$ component as

$$\xi_{ij}(r) = \frac{dd_{ij}(r) - rrr_{ij}(r)}{RR(r)}.$$  

(A8)

where $rr_{ij}/RR$ is a known quantity given the number density of each component, and one only needs to measure $dd_{ij}(r)$ and $RR(r)$.

An interesting application of the above results to real data is to measure 2PCFs (either projected or redshift-space) of subsamples of galaxies of a volume-limited sample (e.g., a sample of galaxies divided into blue, green, and red galaxy subsamples). To measure all the two-point autocorrelation functions of galaxies in the subsamples and their two-point cross-correlation functions, we do not need to construct random catalogs for each subsample. We only need one random catalog for the whole sample, and we measure all the correlation functions in a single run based on equation (A8). The two-point autocorrelation function for all the galaxies in the whole sample, as the weighted sum of the two-point auto- and cross-correlation functions of subsamples (eq. [A6]), is obtained for free. To generalize equation (A8) in the spirit of the widely used Landy-Szalay estimator (Landy & Szalay 1993), one can replace $-rr_{ij}(r)$ with $-2dr_{ij}(r) + rrr_{ij}(r)$. To count $dr_{ij}$ data-random pairs, one may randomly tag the points in the random catalog with component indices according to the fraction of the subsample galaxy spatial density in the overall sample.

A2. DETAILS ON THE DECOMPOSITION OF THE 2PCF INTO CENTRAL/SATELLITE AND ONE-HALO/TWO-HALO TERMS

Following similar reasoning as in § A.1, now let us tag galaxies with two subscripts and decompose the overdensity field as

$$\delta = \sum_{i,\alpha} \delta_{i\alpha},$$  

(A9)

where $i$ denotes the ID of the host halo, and $\alpha$ is either $c$ (central) or $s$ (satellite). That is, the overdensity is decomposed into contributions from central and satellite galaxies from each halo. The random catalog can be obtained by randomly redistributing all the galaxies in the volume with their tags untouched.

In a similar way as before, we can write $\delta_{i\alpha}$ as

$$\delta_{i\alpha} = \frac{n_{i\alpha} - \bar{n}_{i\alpha}}{\bar{n}} \times \frac{\bar{n}_{i\alpha}}{\bar{n}} = \delta_{i\alpha} \frac{n_{i\alpha}}{\bar{n}}.$$  

(A10)

It is straightforward to show that $\xi(r)$ can be formally decomposed as

$$\xi(r) = \sum_i \langle \delta_{i\alpha}(x)\delta_{i\alpha}(x + r) \rangle \frac{n_{i\alpha}^2}{\bar{n}^2} + \sum_i \langle \delta_{i\alpha}(x)\delta_{i\beta}(x + r) \rangle \frac{2n_{i\alpha}n_{i\beta}}{\bar{n}^2} + \sum_i \langle \delta_{i\beta}(x)\delta_{i\alpha}(x + r) \rangle \frac{n_{i\beta}^2}{\bar{n}^2}$$

$$+ \sum_{i < j} \langle \delta_{i\alpha}(x)\delta_{j\alpha}(x + r) \rangle \frac{2n_{i\alpha}n_{j\alpha}}{\bar{n}^2} + \sum_{i < j} \langle \delta_{i\alpha}(x)\delta_{j\beta}(x + r) \rangle \frac{n_{i\alpha}n_{j\beta}}{\bar{n}^2} + \sum_{i < j} \langle \delta_{i\beta}(x)\delta_{j\alpha}(x + r) \rangle \frac{2n_{i\beta}n_{j\alpha}}{\bar{n}^2}.$$  

(A11)

It is easy to identify the six terms on the right-hand side as contributions from the one-halo cen–cen (which is a Dirac $\delta_{i\alpha}$ function that we are not interested in), the one-halo cen–sat, the one-halo sat–sat, the two-halo cen–cen, the two-halo cen–sat, and the two-halo sat–sat pairs, respectively (“cen” for central galaxy and “sat” for satellite galaxy).
In terms of measurement, each component can be reduced to the \( (dd - rr)/RR \) form. As an example, consider the case for the two-halo cen-sat term. Note that \( n_{ic} = 1/V, n_{js} = N_{ic}/V, \) and \( n = N/V, \) where \( N_{ic} \) is the number of satellites in the halo of ID \( i \) and \( N \) is the total number of all galaxies in the volume \( V. \) Therefore, the two-halo cen-sat contribution is

\[
\xi_{2h,cs}(r) = \sum_{i \neq j} \langle \delta_{ic}(x) \delta_{js}(x + r) \rangle \frac{\bar{n}_{ic} \bar{n}_{js}}{n^2} = \sum_{i \neq j} \frac{dd_{ic,js}(r) - rr_{ic,js}(r)}{rr_{ic,js}(r)} N_{js} \frac{N_{ic} N_{js}}{N^2},
\]

where \( dd_{ic,js} \) and \( rr_{ic,js} \) are numbers of data-data and random-random pairs between galaxies tagged as \( i/c \) and \( js. \) One thing to notice is that \( 2rr_{ic,js} N_{js} \) does not depend on \( i \) and \( j, \) but is equal to \( rr_{cs}(r)/N_{pair,cs}, \) where \( rr_{cs}(r) = \sum_{i \neq j} rr_{ic,js}(r) \) is the count of all the random two-halo cen-sat pairs with separation around \( r, \) and \( N_{pair,cs} \) is the total number of two-halo cen-sat pairs in the volume (where \( cs' \) denotes a two-halo pair). Also noting that \( N^2/2 = N_{pair,total} (N \gg 1), \) we then have

\[
\xi_{2h,cs}(r) = \frac{dd_{cs}(r) - rr_{cs}(r)}{rr_{cs}(r)}/N_{pair,cs}/N_{pair,total}.
\]

where \( dd_{cs} \) and \( rr_{cs} \) are all the data-data and random-random two-halo cen-sat pairs with separation around \( r. \) We find the following relation between \( rr_{cs} \) and the total number of all random pairs around separation \( r, \) \( RR(r), \)

\[
\frac{rr_{cs}(r)}{N_{pair,cs}} = \frac{RR(r)}{N_{pair,total}}.
\]

Therefore, we end up with

\[
\xi_{2h,cs}(r) = \frac{dd_{cs}(r) - rr_{cs}(r)}{RR(r)}.
\]