Nonlinear Bias of Cosmological Halo Formation in the Early Universe

Kyungjin Ahn, Ilian T. Iliev, Paul R. Shapiro and Chaichalit B. Srisawat

1 Department of Earth Sciences, Chosun University, Gwangju 501-759, Korea
2 Astronomy Centre, Department of Physics and Astronomy, Pevensey II Building, University of Sussex, Falmer, Brighton BN1 9QH
3 Department of Astronomy, University of Texas, Austin, TX 78712-1083, USA

ABSTRACT

When galactic haloes form during the gravitational growth of cosmological density fluctuations, the halo density field follows that of the underlying matter density field but with a higher (or sometimes lower) contrast, expressed as the local halo bias – the ratio of the former to the latter. We present estimates of the nonlinear bias of cosmological haloes spanning a wide range in mass, from \( \sim 10^5 \)M\(_\odot\) to \( \sim 10^{12} \)M\(_\odot\), by combining the empirical, average mass function derived from a suite of high-resolution cosmological N-body simulations, and the theoretical nonlinear bias parameter based on the extended Press-Schechter formalism. The halo bias is expressed in terms of the mean bias and stochasticity as a function of local overdensity \( \bar{\delta} \), based on different filtering scales, which is realized as the density of individual cells in uniform grids. The sampled overdensities span a range large enough to include both linear and nonlinear regimes, allowing us to obtain the fully nonlinear bias effect on the formation of haloes. A very strong correlation between \( \bar{\delta} \) and halo population overdensity \( \delta_h \), or nonlinear bias, is found, along with sizable stochasticity. We find that the empirical mean halo bias matches, with good accuracy, the prediction by the peak-background split method based on the excursion set formalism, as long as the empirical, globally-averaged halo mass function is used. Consequently, this bias formalism is insensitive to uncertainties caused by varying halo identification schemes, and can be applied generically. We also find that the probability distribution function of biased halo numbers has wider distribution than the pure Poisson shot noise, which is attributed to the sub-cell scale halo correlation. We explicitly calculate this correlation function and show that both overdense and underdense regions have positive correlation, leading to stochasticity larger than the Poisson shot noise in the range of haloes and halo-collapse epochs we study. Our results can be used to generate mock halo catalogues once a density field is given, such as in cosmological N-body simulations of structure formation. Such an application can extend catalogues of haloes to those which are otherwise unresolved in N-body simulations, as we have done so elsewhere in the study of cosmic reionization to consider a wide dynamic range of radiation sources, or expedite the mock survey of haloes (galaxies) with a large number of coarse-grained simulations of density fields for cosmology.

Key words: cosmology: large-scale structure of Universe – cosmology: theory – galaxies: haloes.

1 INTRODUCTION

In the standard scenario of cosmological structure formation, cosmological haloes are the features of the cosmic web of highest overdensity in approximate virial equilibrium, that result from the nonlinear amplification of initially-linear, Gaussian-random density fluctuations by gravitational instability. Galaxies and haloes, however, are not unbiased tracers of the underlying density distribution, and thus understanding this “bias” effect is crucial to extract cosmological information from the data of galaxy surveys, for example.

The idea that galaxy bias (from this point on, we will sometimes denote dark matter haloes loosely by “galax-
ies” to reflect the original ideas of associating galaxies purely by high-density peaks without resorting to hydrodynamical cooling mechanism) exists and can be calculated from the statistics of Gaussian random initial density fields was pioneered by Kaiser (1984). Bardeen et al. (1986) extended this idea to take a full account of the Gaussian random density field in a cosmological context, to understand how haloes grow out of this random field and cluster spatially. In the meantime, Press & Schechter (1974) extended cosmological haloes (or galaxies) as high-density peaks and estimated halo mass function, and this PS formalism was recounted more rigorously by Bond et al. (1994) through their excursion set formalism (sometimes called the extended PS formalism), where they showed that cloud-in-cloud effect explains the fudge multiplicity factor 2 in the PS mass function. All these ideas form the backbone of the peak-background split scheme for calculating the galaxy bias by Cole & Kaiser (1989), which bears the idea that haloes (peaks) are more typically formed in high density regions. Mo & White (1996, MW hereafter) calculated fully nonlinear bias combining the peak-background split scheme with the spherical top-hat collapse model under the extended Press-Schechter formalism, and also calculated the useful “linear bias parameter” in the linear regime. The peak-background split scheme may not give a perfectly accurate prediction of N-body simulation results (e.g. MW; Manera, Sheth & Scoccimarro 2011), which is usually attributed to the discrepancy between the PS mass function and the N-body halo mass function at low- and high-mass ends (e.g. Sheth & Tormen 1999, ST hereafter; Jenkins et al. 2001). This discrepancy stimulated better-fitting functional forms (e.g. ST; Jenkins et al. 2001; Warren et al. 2006; Reed et al. 2007; Lukić et al. 2008; Lim & Lee 2013; Watson et al. 2014). Barkana & Loeb (2004, BL hereafter) then developed a hybrid scheme of combining ST mass function and the linear bias parameter derived from the extended Press-Schechter formalism and showed that this fitted the linear N-body halo bias better than MW prediction.

Bias can of course have stochasticity, which was formulated theoretically by Dekel & Lahav (1999): haloes sampled inside a suite of Eulerian cells of a given density, or count-in-cell (CIC) haloes, are expected to deviate from purely Poisson distribution, if there is either correlation or anti-correlation of haloes at sub-cell scales which then result in variance of the number of haloes ($\sigma^2(N)$) larger or smaller than Poissonian value, respectively (e.g. Peebles 1993; see also Section 3.3). Somerville et al. (2001) compared the prediction by Dekel & Lahav (1999) to N-body simulation results, and based on the observed $\sigma^2(N)$ they concluded that haloes are usually correlated in overdense regions and anti-correlated in underdense regions (we will however contradict this claim in Section 4.3). Later work found that haloes usually show variance larger than the Poissonian value (e.g. Nervinck et al. 2014) find that haloes of mass $10^{10-11} \, M_\odot$ show this “super-Poissonian” distribution under 2/h Mpc cells), which are well fitted by the functional distributions suggested by Saslaw & Hamilton (1984) and Sheth (1995).

A useful application of the nonlinear halo bias prescription is to create mock halo catalogues in a large scale for either cosmology or astrophysics. While mock galaxy catalogues can be created by schemes based on quasi-linear perturbation theory, such as PINOCCHIO (Monaco et al. 2002, 2013) and PTHALOES (Scoccimarro & Sheth 2002, Manera et al. 2014), they are usually limited to the scales under which density perturbation remains quasi-linear at most. This limitation can be overcome by nonlinear halo bias schemes, as in Kitaura, Yepes & Prada (2014) who prove the concept by generating halo catalogues which are statistically consistent with N-body halo catalogues, suited for probing the baryon acoustic oscillation (BAO) feature by surveys such as the Baryon Oscillation Spectroscopic Survey (BOSS). We intend to achieve a similar goal in the long run, but with a bias scheme that is fully nonlinear to provide a more accurate prescription than e.g. the fitting formula used by Kitaura, Yepes & Prada (2014): even when $\delta = 0$ our prescription correctly predicts that the conditional halo number density may deviate from the globally averaged halo number density (Section 4.3) depending on the filtering scale, while the fitting formula used by Kitaura, Yepes & Prada (2014) does not.

A similar formalism can also be applied to astrophysical problems. Understanding the halo bias is crucial e.g. in the study of cosmic reionization, due to the very large dynamic gap between the very small galaxies believed to be the main drivers of reionization (see e.g. Ciardi & Ferrara 2003, for a review) and the large characteristic scales of the reionization patchiness (Friedrich et al. 2011; Iliev et al. 2014). Barkana & Loeb (2004) used a hybrid halo bias scheme to study the fluctuation of the 21cm background from the fluctuating halo distribution during the epoch of reionization (EoR). Fast semi-numerical simulators of reionization (Zahn et al. 2007; Santos et al. 2008; Alvarez et al. 2009; Mesinger, Purlanetto & Cen 2011), whose basis was formulated by Purlanetto, Zaldarriaga & Hernquist 2004 and Purlanetto & Oh 2004 to replace the time-consuming ray-tracing by a faster excursion set formalism, make use of a similar formalism to seed haloes in a coarse-grained density field.

We have indeed applied this formalism to a simulation of cosmic reionization, by which we could span the full dynamic range of halos hosting radiation sources. Cosmic reionization is believed to occur very inhomogeneously with large II regions, whose sizes show a wide distribution peaked at ~20 comoving Mpc before completion if roughly put. Therefore it is necessary to use a large box in order to simulate the reionization process in a statistically reliable way. This requirement, however, limits the ability of the simulation to resolve “minihaloes” which are believed to host Population III stars, and allows the simulation to only resolve the more massive kind, or “atomic-cooling halos”. Indeed most reionization simulations in large boxes used to implement atomic-cooling halos only, while this may underestimate the photon budget in the early stage of reionization. In a large-scale (box size of 114/h Mpc comoving) simulation of cosmic reionization (with ray-tracing method), Ahn et al. (2012) used the conditional halo bias found in Section 3.3 of this paper to include minihaloes, which could not otherwise have been realized due to numerical resolution. This way, they could span the full dynamic range of haloes – both minihaloes and atomic-cooling haloes – responsible for emitting hydrogen-ionizing and H$_2$-dissociation radiation, and observed that the reionization process is extended further in time to comply better with several observational constraints.
On much larger scales (box size of 425/h Mpc comoving) the same technique was used to perform the largest-volume, ray-tracing simulations of cosmic reionization to date, presented in [Hix et al. 2014] and further explored in [Datta et al. 2013], [Park et al. 2013], and [Shapiro et al. 2014]. This used the results in Section 4.2.2 to include the unresolved low-mass atomic cooling haloes ($M = 10^{6} - 10^{9} M_{\odot}$). Another prospective application is in exploring the effects of primordial non-Gaussianity on halo bias, which is an active area of research (e.g. [Dalal et al. 2008] [Adshead et al. 2012] [D’Aloisio et al. 2013]), and which also leads to ionization bias (e.g. [Joudaki et al. 2013] [D’Aloisio et al. 2013]) detectable by 21 cm observations (e.g. [Mao et al. 2013]).

In this paper, we examine and compare the nonlinear halo bias from both our suite of cosmological N-body simulations suited for the study of haloes responsible for EoR and a semi-analytical, fully nonlinear peak-background split scheme. This theoretical scheme is a hybrid scheme similar to the one by BL, but also differ as we combine the empirical (mean) halo mass function to the bias factor and extend it to the fully nonlinear regime in a non-perturbative way. Through this, we investigate whether the bias factor can be purely based upon the excursion set formalism and separated cleanly from the mass function, which bears uncertainty due to its strong dependence on specific halo-identification schemes. We also study the stochasticity of halo bias from these simulations and examine whether they are purely Poissonian or not, which has been investigated recently to conclude that haloes in some mass range indeed have super-Poissonian distribution [Baldauf et al. 2013] [Nevirnck et al. 2014]. Toward this, we calculate the 2-point halo correlation function and quantify its contribution to stochasticity in addition to the Poisson noise. While our paper is focused on the range of haloes responsible for cosmic reionization at $z \geq 6$, and therefore it can be used readily in the study of EoR, our formalism should be applicable in more generic cases.

This paper is organized as follows. In Section 2 we briefly describe our N-body simulation. In Section 3 we describe the theoretical scheme for the nonlinear halo bias, which combines the peak-background split scheme (Sections 3.1 and 3.2) with the empirical N-body halo mass function (Section 3.3), and also describe the stochasticity and various quantities related to halo mass function (Section 3.4). We then describe our results in Section 4 first on the mean halo mass function (Section 4.1), then on the mean bias (Section 4.2) and on the stochasticity (Section 4.3). We further investigate the validity of the usual linear bias approximation in Section 4.4. We conclude our paper in Section 5 together with a schematic layout of our bias prescription toward generating mock halo catalogues.

3 THEORY

Formation of cosmological haloes is strongly correlated with their larger-scale density environment. The excursion set formalism [Bond et al. 1994] gives a quantitative description of this biased halo formation in terms of the conditional halo mass function $dn/dM(M; \delta)$, where $\delta \equiv (\rho - \bar{\rho})/\bar{\rho}$ is the overdensity of the local environment. This description is called the peak-background split, where haloes are considered as the high-density “peaks” that are placed on large-scale density “background”. In the linear regime where $\delta \ll 1$, this yields the linear bias parameter which has been used extensively in cosmology (MW).

In this Section we introduce a formalism which is intended to describe the local nonlinear bias in a non-perturbative way, based mostly on the formalism by MW and the idea of BL. Therefore, we revisit previous theoretical work, and at the same time describe modifications we made in this Section. We will then compare the prediction from this formalism to the N-body data results in Section 4.
3.1 Biased halo mass function in Lagrangian volume

It is shown in the excursion set formalism that distribution of linear overdensity \( \delta \) in the initially Gaussian-random matter density field filtered with a “sharp \( k \)-space filter”,

\[
\rho(r, R_f) = \int d^3r' W_K(r - r'; R_f) \rho(r', 0),
\]

\[
\rho(k, R_f) = W_K(k; R_f) \rho(k, 0), (1)
\]

where the window function \( W_K(r; R_f) \) is the Fourier transform of sharp \( k \)-space filter \( \tilde{W}_K(k; R_f) \equiv \Theta(1 - k R_f) \), still follows Gaussian distribution (Bond et al. 1991). This is obviously true even in the density field linearly extrapolated to the observing redshift with the linear growing factor. This way, one can use the linearly extrapolated density field and the appropriate halo collapse criterion to predict halo population at any filter scale and redshift.

In this formalism the unconditional, globally-averaged differential halo number density (mass function) is given by the Press-Schechter formula (Press & Schechter 1974)

\[
\frac{dn}{dM}_\text{PS} (M) = \left( \frac{dn}{dM}_\text{PS} (\sigma_M^2; \delta_c) \right) = -\frac{1}{2\sqrt{\pi}} \frac{d\sigma_M^2}{dM} \frac{\bar{\rho}_0}{\sigma_M^2} \nu \exp \left[ -\frac{\nu^2}{2} \right], (2)
\]

where \( \sigma_M^2 \) is the variance of Gaussian distribution of the density field (linearly extrapolated to the present) filtered in real space in spheres with radius \( R_f \), \( \bar{\rho}_0 \) is the present matter density, \( \nu \equiv \delta_c / (D(z) \sigma_M) \) (with the linear growth factor \( D(z) \) in \( \Lambda \)CDM universe) is the ratio of critical overdensity \( \delta_c = 1.68 \) to \( \sigma_M(z) = D(z) \sigma_M \), and \( R_f \) is the length scale usually associated with the halo mass \( M \) by

\[
M = M(R_f) = \frac{\bar{\rho}_0 4\pi}{3} R_f^3. (3)
\]

Equation (2) is unconditional in a sense that this represents the average halo distribution in the universe.

The excursion set formalism also predicts halo population inside a region with given mean overdensity and size. Sharp \( k \)-space filtering allows one to write this in terms of conditional probability analytically, because wavevolumes at different filter scales are linearly independent. The barrier crossing and variance under given density environment, which will be denoted by a “cell”, is measured from the new origin \( \delta_{\text{lin}} \) (throughout this paper, unless specified differently, we denote the full, nonlinear overdensity of a cell by \( \delta \) for simplicity) and \( \sigma_{\text{cell}} \), which are linearly extrapolated density of the cell and variance corresponding to the Lagrangian cell size \( R_{\text{cell}} \), respectively.

\[
\rho_{\text{cell}} (1 + \delta) = \frac{4\pi}{3} R_{\text{cell}}^3. (4)
\]

According to the well-known excursion-set formalism (Bond et al. 1991), the differential halo number density (halo mass function) inside a Lagrangian region with \( \delta_{\text{lin}} \) (linearly extrapolated to redshift \( z \)) and \( R_{\text{cell}} \) is then given by a conditional mass function

\[
\left( \frac{dn}{dM} \right)_{\text{PS}, b} (M|\delta_{\text{lin}}) = \left( \frac{dn}{dM} \right)_{\text{PS}} (\sigma_M^2; \delta_c; \delta_{\text{cell}}; \delta_{\text{lin}}) = \left( \frac{dn}{dM} \right)_{\text{PS}} (\sigma_M^2 - \sigma_{\text{cell}}^2; \delta_c - \delta_{\text{lin}}) \frac{1}{2\pi} \frac{d\sigma_M^2}{dM} \frac{\bar{\rho}_0}{\sigma_M^2} \exp \left[ -\frac{(\delta_c - \delta_{\text{lin}})^2}{2 (\sigma_M^2 - \sigma_{\text{cell}}^2)} \right] \times \frac{\bar{\rho}_0}{\sigma_M^2} \frac{d\sigma_M^2}{dM} \frac{\bar{\rho}_0}{\sigma_M^2} \exp \left[ -\frac{(\delta_c - \delta_{\text{lin}})^2}{2 (\sigma_M^2 - \sigma_{\text{cell}}^2)} \right], (5)
\]

which takes the same form as equation (2) but with \( \sigma_M^2 \) and \( \delta_c \) replaced by \( \sigma_M^2 - \sigma_{\text{cell}}^2 \) and \( \delta_c - \delta_{\text{lin}} \), respectively. This defines the Lagrangian over-abundance of haloes of mass \( M \),

\[
\delta_{\text{b}}^L (M|\delta_{\text{lin}}) \equiv \left( \frac{dn}{dM} \right)_{\text{PS}, b} (M|\delta_{\text{lin}}) \left/ \left( \frac{dn}{dM} \right)_{\text{PS}} (M) \right. - 1. (6)
\]

Note that two important factors should be considered in order to generalize equation (5). First, in the nonlinear regime where \( \delta_{\text{lin}} \sim 1 \), one should match the nonlinear \( \delta \) to the linear \( \delta_{\text{lin}} \) to use equation (5), because this is based on the linear theory. Second, \( (dn/dM)^L_{\text{PS}, b} \) and \( \delta_{\text{b}}^L \) should be converted into the corresponding Eulerian mass function and Eulerian halo over-abundance, respectively, because Eulerian quantities are of much more practical use than Lagrangian quantities. This conversion will be described in Section 3.2.

3.2 Nonlinear background and biased haloes mass function in Eulerian volume

The quantities \( (dn/dM)^L_{\text{PS}, b} \) and \( \delta_{\text{b}}^L \) in equations (4) and (6) are derived assuming that density grows linearly with the linear growth factor and are defined in the Lagrangian

---

1. We neglect the very weak redshift dependence of \( \delta_c \) in \( \Lambda \)CDM in our study, while for \( z \lesssim 4 \) one should implement its redshift dependence.

2. Rigorously speaking, we cannot associate such a well-defined mass \( M \) with \( R_f \) in the case of sharp-\( k \)-filtering (e.g. Bond et al. 1991). However, we adopt this definition for simplicity.
volume. In reality, growth of density perturbations is nonlinear in general, and this also yields large difference between the Lagrangian and Eulerian volumes.

Therefore, we first need to map nonlinear overdensity $\delta$ to linear overdensity $\delta_{\text{lin}}$. We use the mapping scheme based on the tophat collapse model, which has also been used by MW, where $\delta$, which is nonlinear in general, is linked to $\delta_{\text{lin}}$ in a parametric form of $\theta$ as follows:

$$\delta = \left( \frac{10 \delta_{\text{lin}}}{3(1 - \cos \theta)} \right)^3 - 1, \quad \delta_{\text{lin}} = \frac{3 \times 6^{2/3}}{20} (\theta - \sin \theta)^{2/3},$$

if $\delta > 0$. Similarly, if $\delta < 0$,

$$\delta = \left( \frac{10 \delta_{\text{lin}}}{3(\cosh \theta - 1)} \right)^3 - 1, \quad \delta_{\text{lin}} = \frac{3 \times 6^{2/3}}{20} (\sinh \theta - \theta)^{2/3}.$$  

Note that $\delta$ increases monotonically as $\delta_{\text{lin}}$ increases, such that there exists one-to-one mapping.

We also need to consider the change of Lagrangian volume by multiplying the ratio of Lagrangian volume to the Eulerian volume to obtain the correct Eulerian number density, which yields the final form:

$$\left( \frac{dn}{dM} \right)_{\text{PS}, b} = \left( \frac{dn}{dM} \right)_{\text{PS}, b} L (1 + \delta).$$  

By taking further approximation that $\delta_{b} \gg \delta$ and $\sigma_{M} \gg \sigma_{\text{cell}}$, MW find a useful linear relation between $\delta_{b}$ and $\delta_{\text{cell}}$. This approximation implies that total mass contained in haloes inside a cell is much smaller than the total mass of the cell. However, this approximation is not always valid at high resolution because some cells in our density field, depending on the choice of the cell-size, may achieve very high overdensity $\delta_{\text{cell}}$ such that $\delta_{b} \gtrsim \delta$ and $\sigma_{M} \gtrsim \sigma_{\text{cell}}$. Therefore, we just use equation (9) in its general form, which allows for nonlinear relation between $\delta_{b}$ and $\delta_{\text{cell}}$.

3.3 Nonlinear bias and hybrid conditional mass function

Before proceeding, let us define the mean conditional bias function $b(\delta)$:

$$b(\delta) = \frac{\langle \delta_{b}(M|\delta) \rangle_{\delta_{b}|\delta}}{\delta},$$

where $\delta_{b}(M|\delta)$ is the conditional, Eulerian halo over-abundance, and the seemingly repetitive definition of the average is to clarify the fact that the average is taken only over the cells with the given $\delta$, following the notation from equations (3) and (4) of [Dekel & Lahav 1999, DL henceforth], which is different from the average over all cells regardless of $\delta$, or $(\cdot)$. This average takes the following integral form for any conditional function of $\delta_{b}$ under a given $\delta$, $f(\delta_{b})$:

$$\langle f(\delta_{b}) \rangle \equiv \left( f(\delta_{b}) \right)_{\delta_{b}|\delta} = \int d\delta_{b} P(\delta_{b}|\delta) f(\delta_{b}).$$

where $P(\delta_{b}|\delta)$ is the conditional probability for a cell with $\delta$ to have $\delta_{b}$ as the halo over-abundance inside it (DL), and only those cells with given $\delta$ are included in the integration. To distinguish the conditional averaging from the normal averaging ($\langle f \rangle$), we denote the former by a simple notation, $[f]$, in which the dependence on $\delta$ is assumed implicitly. Equation (11) is equivalent to equation (5) in DL.

Equations (2), (3) and (9) naturally determine by how much the local halo mass function is modified. The Eulerian over-abundance of haloes is then given by

$$[\delta_{h}(M|\delta)] = \left( \frac{dn}{dM} \right)_{\text{PS}, b}(M|\delta) - 1 = \left( \frac{dn}{dM} \right)_{\text{PS}}(\sigma_{M}^{2}; \sigma_{\text{cell}}^{2}; \delta_{\text{lin}}) (1 + \delta) - 1,$$

which is equivalent to equation (19) of MW. The bias function $b$ becomes independent of $\delta$ in the linear regime where $\delta_{b} \gg \delta \simeq |\delta_{b}|$ and $\sigma_{M}^{2} \gg \sigma_{\text{cell}}^{2}$, and is given as a function of $\nu$ alone, at any given $z$:

$$b_{\text{lin}}(\delta) = 1 + \frac{\nu^{2} - 1}{\delta_{c}(z)}.$$  

(MW). $b_{\text{lin}}$ is referred to as the linear bias parameter. We will test the validity of this approximation in § 4.2.

The relation between $\delta_{b}$ and $\delta$ is generally nonlinear, and therefore equation (10) is of limited use for our purposes. Even in the linear regime where $|\delta| < 1$, using equation (13) may be problematic because the other condition $\sigma_{M}^{2} \gg \sigma_{\text{cell}}^{2}$ is not valid in general and then the exponential term in equation (5) cannot be approximated further. For example, for minihaloes of $M \geq 10^{-6} M_{\odot}$, we have $\sigma_{M}^{2} \leq 70.6$, while cells we study here have masses as low as $3.5 \times 10^{6} M_{\odot}$ (in both $6.3 h^{-1}$ Mpc and $20 h^{-1}$ Mpc boxes), which correspond to $\sigma_{\text{cell}}^{2} = 13.2$ and $\sigma_{\text{cell}}^{2} = 24.0$, respectively.

One may naively expect that $(dn/dM)_{\text{PS}, b}$ gives the correct analytical estimate for the biased halo mass function. However, it is well known that the unconditional PS mass function, $(dn/dM)_{\text{PS}}$, is a poor fit to the empirical halo mass function derived from N-body simulations, in general, depending on the range of mass especially so for rare haloes and redshift (e.g. Jenkins et al. 2001). It is thus reasonable to expect that $(dn/dM)_{\text{PS}, b}$ will also become a poor fit to the biased N-body halo mass function.

We therefore adopt a hybrid approach, first introduced by [Barkana & Loeb 2004], to predict the conditional mass function (or bias) by combining $\delta_{b}(\delta)$ (or equivalently $b(\delta)$) as in equation (12), derived from the excursion set formalism, with the unconditional mass function $dn/dM$, which we choose independently. This approach is somewhat advantageous over Sheth & Tormen (2002) and PS, for example, because $\delta_{b}(\delta)$ or $b(\delta)$ is almost independent of how haloes are identified (MW) and thus the unconditional mass function can be found empirically for any arbitrarily identified N-body haloes. We can then expect that when such an empirical mass function $dn/dM$ is combined with equation (12), the resulting mass function may be a better fit to the actual biased halo mass function $(dn/dM)_{b}$.

In contrast to Barkana & Loeb (2004), who choose the well-known PS and Sheth-Tormen (ST) mass functions, we choose three mass functions: PS, ST and the empirical fit to our N-body data. The reason for using the empirical (unconditional) mass function is because (1) both PS and ST mass functions are known to be poor-fits to very rare haloes (see discussion in Watson et al. 2014 and references therein) and for the redshift and halo mass range of interest here all haloes are rare and (2) we want a prescription which is independent of the systematic uncertainties of the unconditional mass function due to the vary-
ing halo-identification schemes. The conditional PS bias trivially reduces to \((dn/dM)_{PS,b}\) while in the other two cases, the unconditional ST \((dn/dM)_{ST}\) and the empirical fit \((dn/dM)_{N-body}\) are both simply multiplied by \(1 + \delta_h\) to produce
\[
\left(\frac{dn}{dM}\right)_{ST,b} = \left\{1 + \delta_h(\delta)\right\}\left(\frac{dn}{dM}\right)_{ST}
= \left\{1 + b(\delta)\delta\right\}\left(\frac{dn}{dM}\right)_{ST}
\]
and
\[
\left(\frac{dn}{dM}\right)_{N-body,b} = \left\{1 + \delta_h(\delta)\right\}\left(\frac{dn}{dM}\right)_{N-body}
= \left\{1 + b(\delta)\delta\right\}\left(\frac{dn}{dM}\right)_{N-body},
\]
where \(\delta_h(\delta)\) is given by equation (12). It is important to note that even when \(\delta = 0\), \((1 + \delta_h) \neq 1\) in general. In order to illustrate this, let us consider the limiting case of very rare haloes such that \(\nu \gg 1\). Such haloes will most likely form at very high-density regions – or more explicitly, high-density cells with some fixed Eulerian volume \(-\delta \gg 0\). In this case, \((1 + \delta_h) \rightarrow 0\) or \(b(\delta)\delta \rightarrow -1\) as \(\delta \rightarrow 0\), and thus the linear bias parameter, which yields \((1 + \delta_h) \rightarrow 1\) as \(\delta \rightarrow 0\), inevitably fails in estimating the bias correctly even in the linear regime. More detailed discussion of this aspect is in Section 3.4.

Finally, the fraction of halo-mass to cell-mass, or the collapsed fraction, is given by
\[
f_{c,b}(M_{min}, M_{max}) = f_c(M_{min}, M_{max}|\sigma_c^2; \delta) = \int_{M_{min}}^{M_{max}} \left\{\frac{dn}{dM}\right\}_b M dM
= \int_{M_{min}}^{M_{max}} \left\{\frac{dn}{dM}\right\}_b \frac{\rho_c(1 + \delta)}{\rho_0} M dM
= \int_{M_{min}}^{M_{max}} \left\{\frac{dn}{dM}\right\}_b \frac{dN_{cell}}{dV} M dM
= \int_{M_{min}}^{M_{max}} \left\{\frac{dn}{dM}\right\}_b \frac{dN_{cell}}{dV} \frac{M_{cell}}{V_{cell}}
= \frac{1}{\rho_0} \int_{M_{min}}^{M_{max}} \left\{\frac{dn}{dM}\right\}_b \frac{dN_{cell}}{dV} \frac{M_{cell}}{V_{cell}}
\]
which is naturally expressed in Lagrangian quantities, because both masses inhabit the same Lagrangian region. Here once again, \((dn/dM)_b\) can be based on either the PS mass function, the ST mass function or the empirical fit to simulations.

### 3.4 Expected stochasticity and renormalization

We have so far described the mean conditional mass function. In reality, the observed correlation should exhibit stochasticity as well, because structure forms out of a random density field. In addition, when haloes of our interest are rare, not all the cells with given \(\delta\) will contain such haloes, giving rise to Poisson fluctuations. However, we will soon see that the stochasticity should differ from pure Poissonian distribution. Here we consider only the local stochasticity and postpone the analysis of multi-point correlation and corresponding statistics to a future paper.

Because the conditional mass function has a stochastic element, the total number of haloes inside cells with given overdensity \(\delta\) and Eulerian volume \(V_{cell}\) would show a scatter around the mean value. For the total number of haloes in a mass bin \(M = [M_{min}, M_{max}]\),
\[
N(M_{min}, M_{max}|\delta, V_{cell}) = V_{cell} \int_{M_{min}}^{M_{max}} dM \left\{\frac{dn}{dM}\right\}_{o,cell},
\]
over different cells with the same \(V_{cell}\) and \(\delta\) and where \((dn/dM)_{o,cell}\) is the observed halo mass function inside each cell, one would naively expect that the probability distribution function (PDF) of \(N\) will obey the Poisson statistics:
\[
P_{cell}(N) = P(N|\delta, V_{cell}) = \frac{e^{-[N]} [N]^N}{N!},
\]
where the average is again taken only over the cells with given \(\delta\) such that \([N] = \langle N \rangle_{b,\delta} = \langle N(M_{min}, M_{max}|\delta, V_{cell}) \rangle_{b,\delta}\). If so, both the conditional mean and conditional variance of \(N\) would become identical to \([N]\). However, if correlation of haloes at sub-cell length scale exists, there occurs an additional variance – either positive or negative – in \(N\) (Peebles 1993; DL):
\[
\Delta_{scc}(\delta) = \left\langle \frac{N_{cell}}{V_{cell}} \right\rangle^2 \int V_{cell} dV_{cell} \xi_{12}(\delta),
\]
where “scc” denotes sub-cell correlation such that the integration is taken inside a cell and the conditional sub-cell 2-point correlation function \(\xi_{12}(\delta)\) is defined by
\[
\langle N_1 N_2 \rangle = \left\langle \frac{N_{cell}}{V_{cell}} \right\rangle^2 dV_{cell} \left\{1 + \xi_{12}(\delta)\right\},
\]
where 1 and 2 denote two different sub-cell positions inside the same cell and \(N_1\) and \(N_2\) are number of haloes in each sub-cell. \(\xi_{12}(\delta)\) should not be confused with the global sub-cell correlation function \(\xi_{12}\), defined by
\[
\langle N_1 N_2 \rangle = \left\langle \frac{N_{cell}}{V_{cell}} \right\rangle^2 dV_{cell} \left\{1 + \xi_{12}\right\}.
\]
Note that equation (19) and (20) are restricted only to cells with given \(\delta\), which are direct applications of equations (7.66) and (7.63) in Peebles (1993), respectively. While these equations were originally intended for unconditional quantities in Peebles (1993), applying these to conditional quantities is trivially achieved by replacing the global average \((\cdot)\) with the conditional average \([\cdot]\). This is justified by the fact that when there is no sub-cell correlation in those cells with \(\delta\), or when \(\xi_{12}(\delta) = 0\), the identity \([N_1 N_2] = [N_1] [N_2] = [N] dV_1/V_{cell} [N] dV_2/V_{cell}\) is satisfied by equation (20). The net variance is therefore given as
\[
\sigma^2(\delta) = ([N] - [N])^2 = [N] + \Delta_{scc}(\delta),
\]
which is again an application of equation (7.66) in Peebles (1993) to the conditional cases we consider. This also suggests that the true PDF deviates from the pure Poissonian statistics, and the super-Poissonian PDF suggested by Saslaw & Hamilton (1984), given by
\[
P_{cell}(N) = \frac{[N] e^{-[N](1 - \beta) - [N] \beta (1 - \beta)} ([N] (1 - \beta) + N \beta)^{N-1}},
\]
shows excellent agreement with e.g. the distribution of N-body haloes of \(M = 10^{10-11} \text{ M}_\odot\) (Nerlove et al. 2014). Here \(\beta \equiv 1 - \sqrt{\sigma^2(\delta)/[N]}\) represents the degree of super-Poissonian.

Sometimes, we may only be interested in those cells that contain at least one halo. Quantifying this might be
useful when haloes are rare, such that not all the cells with given \(\delta\) are occupied by these haloes. It is therefore useful to have the conditional probability that there are \(N\) haloes in the cell (with \(\delta\) and \(V_{\text{cell}}\)) once a halo is found in that cell (let us denote these cells by “active cells”). This requires re-normalizing the PDF

\[
P_{\text{cell}}(N|N \geq 1) = \frac{P(N|\delta, V_{\text{cell}}; N \geq 1)}{P(N \geq 1|\delta, V_{\text{cell}})}
\]

\[
= \frac{P(N|\delta, V_{\text{cell}})}{1 - P(N = 0|\delta, V_{\text{cell}})}
\]

\[
= \frac{P_{\text{cell}}(N)}{1 - e^{-|N|(1-\beta)}}, \quad (24)
\]

where in the last equality we used equation (23). The mean value of \(N\) inside “active” cells will then be given by

\[
[N]_a \equiv \sum_{N=1}^{\infty} N P_{\text{cell}}(N|N \geq 1) = \frac{[N]}{1 - e^{-|N|(1-\beta)}}, \quad (25)
\]

which should be used as the estimator of the mean value. Two limiting cases are noteworthy. First, when \([N] \ll 1\), \(P_{\text{cell}}(N|N \geq 1)\) can be approximated as

\[
P_{\text{cell}}(N|N \geq 1) \approx \frac{1}{N!} e^{-N\beta} (N\beta)^{N-1}, \quad (26)
\]

which is no longer dependent on \([N]\). In the other extreme, \([N] \gg 1\), \(P_{\text{cell}}(N|N \geq 1) = P_{\text{cell}}(N)\). Similarly, we use the same renormalization to determine the collapsed fraction inside active cells:

\[
f_{c}(\delta) = \frac{[f_{c}]_a}{1 - e^{-|[N]|(1-\beta)}}. \quad (27)
\]

When \([N] \ll 1\), as the mass function is biased toward the least massive haloes, \([f_{c}]_a \simeq M_{\text{min}}/M_{\text{cell}} = M_{\text{min}}/\nu_0 V_{\text{cell}}^{-1}(1 + \delta)^{-1}\). When \([N] \gg 1\), \([f_{c}]_a\) trivially converges to \([f_c]\).

Note that \([N]\) can be smaller than 1. This does not mean that we will find a fractional, less-than-unity number of haloes on average, which is simply unphysical. This means instead, assuming ergodicity, that

\[
[N] = \frac{\text{total number of haloes found in all cells with } \delta}{\text{total number of cells with } \delta}
\]

\[
\approx \frac{\text{number of active cells with } \delta}{\text{total number of cells with } \delta}, \quad (28)
\]

where the approximation is made possible due to the fact that when \([N] \ll 1\), the PDF \(P_{\text{cell}}(N|N \geq 1)\) is peaked at \(N = 1\).

4 RESULTS

4.1 Mean Unconditional Halo Mass Function

The mean, unconditional halo mass functions at both high and low redshifts were recently discussed in detail in [Watson et al. (2014)], much of it based on the same simulations as the current work. Therefore, we will only summarize a selection of the mean mass function properties that are most relevant here.

In Fig. 1 we show the mass functions in the mass range \(M \geq 10^5\ M_\odot\) at selected redshifts based on the \(L_{\text{box}} = 20/h\ Mpc\) and \(6.3/h\ Mpc\) simulations, together with PS and ST analytical mass functions. The actual quantities plotted are halo number densities \(\Delta n \equiv \int_{M_{\text{min}}}^{M_{\text{max}}} (dn/dM) dM\), integrated over equal-size logarithmic mass bins. The last mass bin includes all haloes with mass \(M \geq 10^6\ M_\odot\). The two simulated mass functions show excellent agreement with each other, except for the high mass end, where the mass function is truncated due to finite volume. This agreement indicates the consistency of the N-body simulation over varying box size.

Compared to the analytical expressions, our N-body mass functions are in better agreement with ST than PS mass functions. The agreement with ST at all redshifts is within \(\sim 25\%\) for \(M = [10^5 - 10^8]\ M_\odot\), the haloes in which range numerically dominate the minihalo population. At very high redshifts, \(z \gtrsim 20\), ST mass function slightly overpredicts halo population at \(M = [10^5 - 10^5.5]\ M_\odot\) and underpredicts halo abundance at \(M \gtrsim 10^6\ M_\odot\), with tendency to deviate increasingly as \(M\) increases, while at relatively low redshifts, over-prediction occurs at \(M = [10^5.5 - 10^6]\ M_\odot\).

As discussed in [Watson et al. (2014)] these differences are partly due to our usage of a halo finder based on spherical overdensity instead of the friends-of-friends one used by ST, and also to the limitations of the ST fit which was based on low-redshift data and relatively small simulations. In contrast, the classical PS mass function gives a poor fit to N-body minihalo data at all redshifts, severely underpredicting the abundance of rare \((\nu = \delta_c/\sigma_M \gg 1)\) haloes and over-predicting the abundance of \(\nu \ll 1\) haloes. Only for the most common \((\nu \approx 1)\) haloes PS is a more reasonable approximation (and also agrees with ST).

Assuming that the prescription for the conditional mass function (linking eq. 12 with unconditional mass function) provides a correct theoretical framework, one may expect that a good fit to unconditional mass function will also provide a good fit to conditional mass function when combined with eq. 12. Therefore, we can expect that \((dn/dM)_{\text{N-body}, b}\) will be the best fit to the mean conditional mass function from the simulations, and \((dn/dM)_{\text{ST}, b}\) will also be a good fit, while \((dn/dM)_{\text{PS}, b}\) will be a poor fit. We will test this expectation in Section 4.2.

4.2 Mean biased halo mass function

We now show how the mean, conditional mass functions of N-body haloes behave in terms of \(\delta\), and compare this to the modelling predictions based on the different mass functions, \((dn/dM)_{\text{PS}, b}\), \((dn/dM)_{\text{ST}, b}\), and \((dn/dM)_{\text{N-body}, b}\). We also compare these to the model based on the linear bias. The stochasticity in this relation will be treated in § 4.3.

4.2.1 Minihaloes

Minihaloes are usually defined by their hydrodynamical properties. Their minimum mass is the cosmic Jeans mass determined by the mean IGM temperature, and their maximum mass is the mass of haloes whose virial temperature is about \(10^4\ K\). While this is the general definition, the uncertainty of the mean IGM temperature at high redshift makes the definition of the minimum mass somewhat uncertain. In this work we instead take their mass to be in a fixed range \(M = [10^5 - 10^8]\ M_\odot\), which is of more direct use to N-body
data at fixed mass resolution. Both 6.3/h Mpc and 20/h Mpc boxes resolve haloes down to $M = 10^7 M_\odot$. The latter simulation thus provides a better statistics by encompassing a volume 32 times as large as that of the former one.

We first examine how well the models based on the analytical mass function fits match the N-body data. Figs 2-3 show the analytical estimates and N-body data on the total number of minihaloes $[N_{5.8}]_a$ and the corresponding total collapsed fraction $[f_{c, 5.8}]_a$. We find that the numerical data from the two simulation volumes is in excellent agreement and that $(dn/dM)_{ST, b}$ and $(dn/dM)_{N-body, b}$ fit the N-body data well over almost the entire range of $\delta$ and $z$, while $(dn/dM)_{PS, b}$ and the linear bias parameter both provide poor fits to the data in general. Even though $[N_{5.8}]_a$ and $[f_{c, 5.8}]_a$ are integral quantities, given that smallest-mass haloes numerically dominate the halo population, both the data and semi-analytical estimates reflect predominantly the low-mass end. Note that as seen in Fig. 4, $(dn/dM)_{ST}$ agrees well with $(dn/dM)_{N-body}$ in the low-mass end, and this is the reason why $(dn/dM)_{ST, b}$ provides a good fit. If we focused on the high-mass end only, $(dn/dM)_{ST, b}$ would be a very poor fit to the observed bias, because the average ST mass function $(dn/dM)_{ST}$ has large discrepancy from the actual N-body data for e.g. $M \geq 10^7 M_\odot$. In contrast, the collapsed fraction in haloes $[f_{c, 5.8}]_a$ (Figs 3 and 9) is a mass-weighted quantity and thus reflects the high mass end better than does $[N_{5.8}]_a$, but the rapid exponential cutoff of the mean halo mass function $dn/dM$ at increasing $M$ still moderates the contribution from the high-mass haloes. The similarity between $(dn/dM)_{ST, b}$ and $(dn/dM)_{N-body, b}$ reflect the simple fact that the unconditional mass functions, $(dn/dM)_{ST}$ and $(dn/dM)_{N-body}$, are similar around the low-mass end.

As the cell size shrinks, however, some discrepancy appears at high $\delta$ regime. Both $(dn/dM)_{ST, b}$ and $(dn/dM)_{N-body, b}$ predictions overestimate the N-body data substantially at $\delta \gtrsim 1.5$, when the volume of the cell has shrunk from $0.45/h$ Mpc$^3$ to $(0.15/h$ Mpc$^3$); see Figs 4 and 9. At this point, where $\delta$ approaches the overdensity criterion for halo identification, we suspect that this could be a symptom of extreme nonlinearity: the mean mass of the cell, $M_{cell} \approx 3.8 \times 10^8 M_\odot$, is small enough to be comparable to the high-mass end of minihaloes, or $10^8 M_\odot$.

In summary, unless the cell is too small, and thus potentially quite nonlinear, the mean nonlinear bias of N-body minihaloes at high redshifts can be explained well by the simple hybrid prescriptions $(dn/dM)_{N-body, b}$ and $(dn/dM)_{ST, b}$. In contrast, at high redshifts, the linear bias approximation deviate too much from the N-body minihalo data to be of much practical use at least under the filtering scales of $\lesssim$ Mpc. The disagreement of $(dn/dM)_{PS, b}$ with the N-body data is just as severe, and we expect that $(dn/dM)_{PS, b}$ will be useless regardless of the filtering scale, because the disagreement is caused by the poorness of the mean PS mass function. It is notable that the rarity of haloes at high redshifts make the linear bias approximation fail even when $|\delta| \ll 1$, which will be discussed in much detail in Section 4.3.

4.2.2 Atomically-cooling haloes

Atomically-cooling haloes (ACH hereafter) are named after the dominant cooling mechanism of baryonic gas inside. Atomic line radiation can cool primordial-composition gas to $T \approx 10^4 K$ from its initially higher virial temperature. Star formation are believed to occur inside these haloes as pre-existing metals or newly-formed $H_2$ can further cool the gas down to $T \sim 100 K$. Therefore, ACHs are usually defined by their virial temperature: haloes with $T \gtrsim 10^4 K$. As this threshold virial temperature roughly coincides with $M \approx 10^8 M_\odot$, we here define ACHs as those haloes with $M \geq 10^8 M_\odot$. The ACHs can be grouped further into low-mass ACHs (LMACH), for which the gas pressure of the photoheated IGM in an ionized patch prevented the halo from capturing the gas it needs to form stars, and high-mass ACHs (HMACH), for which gravity was strong enough to overcome this “Jeans-mass filter” and form stars even in the ionized patches. The dividing line between LMACHs and HMACHs occurred roughly at $\sim 10^9 M_\odot$ (although the precise boundary value is still uncertain).

As our 114/h Mpc box simulation resolves haloes of $M \geq 10^8 M_\odot$, ACHs defined as above are fully identified. Even though the inner structure of low-mass end haloes is not resolved near the resolution limit (see Section 2), for our considerations only the number count of haloes matters, both for the mean halo bias and stochasticity and therefore our results are not affected by this.

We choose two filtering scales, 114/h and 114/h/32 = 3.56/h Mpc. While these choices are somewhat arbitrary, we increased the filtering scales for ACHs from those for minihaloes, due to the increased rarity of ACHs. The halo collapsed fraction is plotted in Figs 6-9. While LMACHs have a finite range in mass, because HMACHs are defined to have a loose end, we assign their maximum mass as the one somewhat smaller than the mass of the average-density cell: $M_{max} = 10^{11.5} M_\odot$ and $M_{max} = 10^{12.5} M_\odot$ for cells with $V_{cell} = (1.78/h$ Mpc$^3$) and $V_{cell} = (3.56/h$ Mpc$^3$), respectively. Otherwise, the bias formalism breaks down (equation 5). Overall, the mean values of both the LMACH collapsed fraction $\langle f_{c, 8.91}\rangle$ and the HMACH collapsed fraction $\langle f_{c, 9.11.5}\rangle$ and $\langle f_{c, 9.12.5}\rangle$ are well predicted by equation 10 when we adopt $(dn/dM)_{PS, b}$ (equation 10). For LMACHs, $(dn/dM)_{ST, b}$ provides as good a fit as $(dn/dM)_{N-body, b}$, except at $z = 6$ where ST prescription somewhat overestimates the mean. For HMACHs, the biggest discrepancy between $(dn/dM)_{ST, b}$ and $(dn/dM)_{N-body, b}$ exist at higher redshifts (e.g. $z = 15.6$) at log(1 + $\delta$) $\gtrsim 0$: here the small number of sampled cells at high cell-density makes it difficult to conclude which prescription provides a better estimator for the mean bias. PS prescription provides a very poor fit at all redshifts.

The linear bias parameter, for both LMACHs and HMACHs, fails in predicting the mean bias in general. This is noteworthy because even in the linear regime, including

3 As to be seen in Fig 4 and Fig 7 the conditional halo correlation function determines the stochasticity. The halo correlation function is composed of the 1-halo term and the 2-halo term, and the dominant contribution to stochasticity comes from the 2-halo term. Therefore, it is not required to fully resolve the halo structure in estimating the stochasticity.
the point \( \delta = 0 \), the linear bias parameter predicts the bias to be off from the observed values, which was also the case for minihaloes. We discuss this issue in detail in Section 4.4.

In summary, even though LMACs and HMACHs are very rare in the regime we study, the nonlinear bias prescription combined with the mean N-body halo mass function fits the observed mean halo bias very well throughout the ranges of redshift and cell density we observe. Therefore, this hybrid bias prescription can be applied for astrophysical and cosmological applications in general. We have indeed applied the bias prescription from this work in simulating cosmic reionization by ACHs in a very large box, 425/h Mpc, in order to populate Eulerian cells with size 425/h/504 = 0.843 Mpc [Hirv et al. 2014]. Because the halo mass resolution of the corresponding N-body simulation was only \( 10^9 M_\odot \), we assigned each cell the missing LMACs using the mean conditional mass function \( (dn/dM)_{N-body, b} \) where the LMACH mean mass function from our 114/h Mpc simulation was used to generate \( (dn/dM)_{N-body, b} \).

4.3 Stochasticity

The average behaviour of conditional mass function is well understood in terms of the biased mass function \( (dn/dM)_{N-body, b} \). Now, how does the scatter of correlation around the mean compare to the expected stochasticity? We showed in Section 4.3 that the variance of the number of haloes \( N \) and given \( \delta \) deviate from the simple Poisson value \( \langle N \rangle \) by the amount \( \Delta_{scc} (\delta) \). We now show the result of simulation and compare this to the Poisson statistics (equation 24) and \( \Delta_{scc} (\delta) \) (equation 14) by explicitly calculating the sub-cell-scale correlation function (equation 24).

In Figs 10-15 we show the actual PDF and compare it to the expected Poisson distribution. We find that the empirical PDF does not follow pure Poisson distribution in general: the observed PDFs usually show large outliers compared to the Poisson statistics (and truncation at some order) is useful in observing the discrepancy comes from the erroneous definition of \( \Delta_{scc} \) in [Somerville et al. 2001], where they subtracted the global mean number of haloes \( \langle N \rangle \) averaged over all cells of \( \delta \) such that \( \Delta_{scc} = \sigma^2(\delta) - \langle N \rangle \), while one should indeed define this as in equation 22 to reflect the effect of the sub-cell correlation function. The observed anti-correlation of \( \xi_{12}(\delta) \), or negative values of \( \xi_{12}(\delta) \) when \( r \) becomes comparable to the cell size as seen in Figs 10-15 is due to the finite cell size, because any correlation existing inside a cell should be counter-balanced by anti-correlation in order to conserve the halo number.

We explicitly calculate \( \xi_{12}(\delta) \) defined by equation 20 and \( \Delta_{scc} (\delta) \) from equation 14. Toward this, we place a uniform grid with \( 25^3 \) sub-cells on each cell with \( \delta \), such that \( dV_1 = dV_2 = V_{cell}/25^3 \). We then sample all sub-cell pairs with given distance \( r_{12} \) – discretized as the distance between centres of sub-cells – and calculate \( \xi_{12}(\delta) \) using equation 20 and \( \Delta_{scc} (\delta) \) using equation 14. We compare this value to the observed, residual variance \( \Delta_{scc} = \sigma^2(\delta) - \langle N \rangle \), which are shown in the bottom panels of Figs 10-15 denoted by \( \Delta_{scc} (\text{obs}) \) respectively. The agreement between the two quantities are excellent, and thus proves the fact that \( \xi_{12}(\delta) \) is the sole origin for the super-Poissonian (or sometimes sub-Poissonian) variance in \( N(\delta) \). Due to halo-number conservation, the correlation function is composed of positive (correlation) and negative (anti-correlation) parts as seen in Figs 10-15.

4.4 Validity of Linear Bias Approximation

Linear bias approximation is widely used in literature and in practical applications such as galaxy surveys for cosmology. The linear bias parameter \( b_{lin} \) is useful when the mass of haloes is fixed, because then \( b_{lin} \) is a simple constant coefficient for varying \( \delta \), or \( \delta_{lin} = b_{lin} \delta \), and the same relation applies to k-space bias such that \( \delta_{lin}(k) = b_{lin} \delta(k) \). Its limitation, however, has already been pointed out by MW themselves, by expanding the nonlinear relation (equation 12) to second order in \( \delta \) and first order in \( \sigma_{2coll}^2/\sigma_M^2 \). Such expansion (and truncation at some order) is useful in observing the halo bias in k-space because algebraic connection between real-space parameters and k-space parameters is possible, and also in understanding the generic behaviour of nonlinear bias. We therefore examine the Taylor-expanded form of equation 12. The main difference from MW is that we expand the nonlinear relation to second order in \( \delta \) but keeping \( \sigma_{2coll}^2/\sigma_M^2 \)-dependence accurate, because we sometimes reach \( \sigma_{2coll}^2/\sigma_M^2 \ll 1 \). This will enable us to examine the dependence of nonlinear bias on the filtering scale more accurately.

We thus Taylor-expand \( \delta_{lin}(M) \) to the second order in \( \delta \) while keeping the dependency on \( R_{cell} \) accurate (as in...
It is important to note that if the bias parameter (equation 10) is expanded instead of $\delta_h$, one should include the singular term $B_0/\delta$ such that $b = B_0/\delta + B_1$, because otherwise the approximated linear bias parameter cannot explain the offset. In this sense, $B$ should not be taken as a physical quantity but merely as a mathematical entity. $\delta_h (= \hat{b})$ is a physical quantity which does not become singular when $\delta \to 0$.

$B_1$ is a good indicator of the overall trend of bias. The sign of $B_1$, which is always positive, guarantees that haloes are not anti-biased but biased for higher cell-densities regardless of $\nu$ or the filtering scale, as long as $\delta$ is in the linear regime. $B_1$ depends on both $\nu$ and $\sigma_{R_{\text{cell},m}}/\sigma_M$. At fixed halo-mass and filtering scale, $B_1$ increases as $\nu$ increases when $\nu < \nu_{\text{crit,lin}} \equiv p\sqrt{1 - \delta_c + 2(1 - p)}$ and decreases when $\nu > \nu_{\text{crit,lin}}$ (figure 22). Such non-monotonic trend in $B_1$ would not be observed when filtering scale is large enough, because then $\nu_{\text{crit,lin}} \to \infty$. At fixed halo mass (and thus fixed $\nu$ and $\sigma_{R_{\text{cell}}}$ at some $z$), the effect of filtering scale or $p$ on $B_1$ is also a mixed bag depending on rarity of haloes (or $\nu$) as seen in figure 22.

5 SUMMARY AND DISCUSSION

We investigated the local bias of cosmological halo formation in the fully nonlinear regime, using both halo data from N-body simulations sampled on uniform grids and theoretical estimates for Eulerian halo bias. Over the wide dynamic range of halo mass, from $10^5 M_\odot$ to $\sim 10^{12} M_\odot$, we find that the observed biased population of haloes $(dn/dM)_b$ inside a cell with density $\delta$ can be matched well by the convolution of the mean N-body mass function $(dn/dM)_{N-\text{body}}$ with the nonlinear bias parameter derived from the extended Press-Schechter formalism. Convolution with the PS mass function provides very poor fits in general, and convolution with the ST mass function provides fits slightly poorer than $(dn/dM)_{N-\text{body}}$. Nevertheless, as the ST mass function is known to break down for very rare haloes (see e.g. the large discrepancy of the ST mass function for haloes of $M \geq 10^8 M_\odot$ at $z \geq 20$ in figure 1), it is best to avoid both PS and ST, and instead use $(dn/dM)_{N-\text{body}}$ in convolving the mean mass function to the bias factor given by equation 12.

Based on the fact that the observed bias in halo population is well matched by the hybrid estimate $(dn/dM)_{N-\text{body},b}$ which combines two physical quantities with different origins (the average mass function $(dn/dM)_{N-\text{body}}$ is determined by a specific halo-identification scheme and the nonlinear bias parameter is based on the extended Press-Schechter theory), this prescription should be applicable in general to cases under other halo-identification schemes.

We also find that the variance of halo numbers inside grid cells with given overdensity is not purely Poissonian, but has additional variance. This variance originates from the sub-cell scale halo-halo correlation, which we proved quantitatively by explicitly calculating the conditional correlation functions. In the regime we studied ($z \geq 6$ and uni-grid filtering with cell size of $\sim [0.2 - 3.6] h^{-1}$ Mpc), we find that the additional variance is always positive except for some negative values sporadically observed for haloes with $M > 10^8 M_\odot$.

The nonlinear bias prescription described in our paper...
Nonlinear Bias of Cosmological Halo Formation

(i) Generate or adopt a mean mass function of haloes $(dn/dM)_{N\text{-body}}$. It is advised not to use the PS mass function, due to the large discrepancy from the usual N-body halo catalogues practically over the full mass range.

(ii) Generate a density field at a redshift of interest: if N-body data is available, adopt a proper smoothing scheme to generate a density field from the distribution of particles. Depending on the size of cells, cell-density can become non-linear, and therefore N-body simulation is recommended.

(iii) Place a uniform grid on the density field from step (ii), and identify the comoving volume of the cell as $V_{\text{cell}}$.

(iv) Visit a cell, and identify the cell overdensity $\delta$. Use equation (1) to deduce $R_{\text{cell}}$. Take $R_{\text{cell}}$ as the spatial filtering scale of the linearly extrapolated density field to $z = 0$, and calculate the corresponding variance $\sigma^2_{R_{\text{cell}}}$. Use equations (7) and (8) (or the numerical fit given by equation 18 of MW) to find matching $\delta_{\text{in}}$ of $\delta$.

(v) To populate a cell with a halo of mass $M$, use equation (2) to obtain $R_f$, and take this as the filtering scale of the linearized density field at $z = 0$ and calculate the corresponding variance $\sigma^2_{R_f}$.

(vi) Plug quantities from steps (iv) and (v) in equation (9), then use equation (10), then finally use equation (11) to calculate the biased halo mass function $(dn/dM)_{N\text{-body}, k'}$. Multiplying the infinitesimal mass bin $dM$ and $V_{\text{cell}}$ to $(dn/dM)_{N\text{-body}, k'}$, one obtains the mean number of haloes $[N]$ of $M = [M, M + dM]$ in the cell.

(vii) Iterate steps (iv) - (vi) over all cells in the box.

(viii) If one wants to implement stochasticity, which should indeed affect the power spectrum of halo density field, use equation (23) with $[N]$ from step (vi) and an empirically found $\sigma^2(\delta)$ to include super-Poisson stochasticity and sample halos by the Monte-Carlo method. For a selected range of halo masses and cell sizes as described in Sections 4.2 and 4.3 a reader may contact us for these values.

The perturbative approach to the nonlinear bias is found very limited. First, the linear bias parameter fails to match the observed bias even when $\delta = 0$, which is a clear indication that the bias parameter is actually singular, or $b \rightarrow B_0/\delta + B_1(\delta \rightarrow B_0 + B_1\delta$ correspondingly) with nonzero $B_0$ as $\delta \rightarrow 0$ (equations 30, 33, and 34). This 0-point offset $(\langle dn/dM\rangle_0(\delta = 0) \neq \langle dn/dM\rangle)$ occurs in general when (1) haloes are rare and/or (2) the cell size is small, and any linear bias parameter without this singular term (e.g. linear bias parameters by MW) is bound to produce this unwanted offset, which MW has already recognized and we have confirmed from our data analysis. The second-order perturbation, which we calculated without the approximation taken by MW, also provides a very poor fit in general. We thus claim that the local nonlinear bias parameter should be used unless perturbative approach is unavoidable.

The nonlinear bias parameter can be applied to both cosmological and astrophysical problems. For example, we already used the mean bias prescription in this paper as a sub-grid treatment to populate simulation boxes with haloes which are not resolved otherwise, for simulating cosmic reionization process: see Ahn et al. (2012) for populating $114 h^{-1} \text{ Mpc}$ box with minihaloes, and Iliev et al. (2013) for populating $425 h^{-1} \text{ Mpc}$ box with LMACHs. Similar approach has been attempted by de la Torre & Peacock (2013), where they test their bias-based sub-grid treatment against resolved N-body haloes in terms of two-point statistics. Their bias prescription, however, still has a form that is not able to reproduce the 0-point offset. Even though we have just studied cosmological haloes at $z \gtrsim 6$, the agreement between data and theoretical prediction in such wide range of halo mass, cell size, cell density, and redshift suggest that this prescription is valid in general.

The nonlinear bias scheme studied here is valid when the primordial density field is Gaussian, and thus may not be directly used to study non-Gaussianity. It is also preferred that further study of the super-Poissonian (or sometimes sub-Poissonian) stochasticity, which we quantified here with $20 h^{-1} \text{ Mpc}$ box for minihaloes and $114 h^{-1} \text{ Mpc}$ box for LMACHs and HMACHs, is devised with higher-resolution, larger-box simulations to increase statistical reliability. Stochasticity is likely to have temporal correlation as well as spatial correlation, which should be further studied for a more self-contained bias prescription.

ACKNOWLEDGMENTS

This work was supported by a research grant from Chosun University (2010). All simulations in this work were undertaken at the Texas Advanced Computing Center (TACC) at The University of Texas at Austin under TeraGrid allocations.

REFERENCES

Adshhead P., Baxter E. J., Dodelson S., Lidz A., 2012, Phys. Rev. D , 86, 063526
Ahn K., Iliev I. T., Shapiro P. R., Mellema G., Koda J., Mao Y., 2012, ApJL, 756, L16
Alvarez M. A., Busha M., Abel T., Wechsler R. H., 2009, ApJL, 703, L167
Balkaaufl T., Seljak U., Smith R. E., Hamaus N., Desjacques V., 2013, Phys. Rev. D , 88, 083507
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Barkana R., Loeb A., 2004, ApJ, 609, 474
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Ciardi B., Ferrara A., 2005, Space Science Reviews, 116, 625
Cole S., Kaiser N., 1989, MNRAS, 237, 1127
Dalal N., Doré O., Huterer D., Shirokov A., 2008, Phys. Rev. D , 77, 123514
D’Aloisio A., Zhang J., Jeong D., Shapiro P. R., 2013, MNRAS, 428, 2765
Datta K. K., Friedrich M. M., Mellema G., Iliev I. T., Shapiro P. R., 2012, MNRAS, 424, 762
de la Torre S., Peacock J. A., 2013, MNRAS, 435, 743
Dekel A., Lahav O., 1999, ApJ, 520, 24
Friedrich M. M., Mellema G., Alvarez M. A., Shapiro P. R., Iliev I. T., 2011, MNRAS, 413, 1353
Furlanetto S. R., Oh S. P., 2005, MNRAS, 363, 1031
Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, ApJ, 613, 1

...
LIST OF FIGURES

1 Unconditional, mean mass functions of MHs found in small-box N-body simulations. Plotted are the mass functions (each top panel) inside 20/h Mpc (black, solid) and 6/h Mpc (purple, dashed) boxes, along with the Sheth-Tormen (blue, dotted) and Press-Schechter (red, dot-dashed) mass functions, all integrated over equal-size logarithmic mass bins, and the ratios of these mass functions (each bottom panel with the same line types) to the mass function inside the 20/h Mpc box. The error bars represent 1σ standard deviation in each mass bin.

2 Correlation between the number of minihaloes $N_{\text{scc}}$ and the cell overdensity $\delta$, when the Eulerian volume of the cell is chosen to be $(0.45/h \text{ Mpc})^3$. Data points are from N-body simulations in the 6.3/h Mpc box (red, larger dot) and 20/h Mpc box (slate blue, smaller dot), which are sampled by $14^3$ and $45^3$ cells, respectively. Theoretical predictions for active cells (eq. 25) based on $(dn/dM)_{\text{SH}}$ (solid, black; eq. 15), $(dn/dM)_{\text{PS}, b}$ (short-dashed, blue; eq. 9), the one by the linear bias approximation (long-dashed, blue; eq. 13) combined with $(dn/dM)_{\text{SH}}$, and the one by the 2nd order approximation with a zero-point offset $B_0$ (dot-dashed, blue; equations 20, 24) also combined with $(dn/dM)_{\text{SH}}$ are plotted for comparison.

3 Correlation between the fraction of mass collapsed into minihaloes ($f_{\text{coll}, 5.8}$) and the cell overdensity $\delta$ in the 6.3/h Mpc box, where the box is sampled by $14^3$ grid-cells. Conventions for plotting follow those of Fig. 2.

4 Same as Fig. 2 except that the volume of the cell is now $(0.15/h \text{ Mpc})^3$. The 6.3/h Mpc box (red circle) and 20/h Mpc box (green dot) are sampled by $44^3$ and $135^3$ cells, respectively.

5 Correlation between the fraction of mass collapsed into minihaloes ($f_{\text{coll}, 5.8}$) and the cell overdensity $\delta$ in the 6.3/h Mpc box and 20/h Mpc box, sampled by $44^3$ and $135^3$ cells, respectively. Conventions for plotting follow those of fig.2.

6 Correlation between the fraction of mass collapsed ($f_{\text{coll}, 8.9}$) into LMACHs ($M = 10^8 - 10^9 M_{\odot}$) and the cell overdensity $\delta$ in the 114/h Mpc box, where the box is sampled by $32^3$ grid-cells. Conventions for plotting follow those of Fig. 2 except for the data points (red point).

7 Correlation between the fraction of mass collapsed ($f_{\text{coll}, 8.9}$) into LMACHs ($M = 10^8 - 10^9 M_{\odot}$) and the cell overdensity $\delta$ in the 114/h Mpc box, where the box is sampled by $64^3$ grid-cells. Conventions for plotting follow those of Fig. 2 except for the data points (red point).

8 Correlation between the fraction of mass collapsed ($f_{\text{coll}, 9.12.5}$) into HMACHs ($M = 10^9 - 10^{12.5} M_{\odot}$), where the maximum mass is roughly the mass of a cell and the cell overdensity $\delta$ in the 114/h Mpc box, where the box is sampled by $32^3$ grid-cells. Conventions for plotting follow those of Fig. 2 except for the data points (red point).

9 Correlation between the fraction of mass collapsed ($f_{\text{coll}, 8.9}$) into HMACHs ($M = 10^8 - 10^{11.5} M_{\odot}$), where the maximum mass is roughly the mass of a cell and the cell overdensity $\delta$ in the 114/h Mpc box, where the box is sampled by $64^3$ grid-cells. Conventions for plotting follow those of Fig. 2 except for the data points (red point).

10 (A) PDFs of the number of minihaloes $N_{\text{scc}}$ at given overdensity $\delta$ (denoted on top of each subpanel) of cells with Eulerian volume $(0.45/h \text{ Mpc})^3$ observed in 20/h Mpc box. The horizontal and vertical axes represent $N_{\text{scc}}$ and $\log P_{\text{cell}}(N_{\text{scc}}|\delta)$, respectively. The data from simulation (histogram) is compared to the pure Poisson PDF (black, dotted) and the super-Poissonian PDF (equation 23, red, dashed). (B) Variances. Plotted are the observed total variance $(N - [N])^2$ (black, solid), the purely Poissonian variance $[N]$ (blue, dashed), the observed excess $\Delta_{\text{acc}}(\text{obs}) = (N - [N])^2 - [N]$ (red, dotted; eq. 20), and a value calculated from the sub-cell correlation function, $\Delta_{\text{acc}}(f)$ (cyan, dot-dashed; eq. 19). The range of $\delta$ and $z$ are selected such that the number of cells with given $\delta$ (binned properly as described in Section 4.3) at given $z$ exceeds 200 for the statistical calculation of the calculated variances. The ratio $\Delta_{\text{acc}}/[N]$, plotted in the bottom panels, quantifies the excess of variance over the purely Poissonian one, $[N]$.

11 (A) Same as Fig. 10(A), but of cells with Eulerian volume $(0.15/h \text{ Mpc})^3$ in 20/h Mpc box. (B) Same as Fig. 10(B). Here we also find that $\Delta_{\text{acc}} > 0$ with this cell size at any $\delta$ at $z \leq 28$. Note the hugely nonlinear regime of $\delta$ observed at $z \leq 12$.

12 (A) PDFs of LMACHs at given overdensity $\delta$ in cells with Eulerian volume $(3.56/h \text{ Mpc})^3$ in 114/h Mpc box. (B) Same as Fig. 10(A). Here we also find that $\Delta_{\text{acc}} > 0$ with this cell size at any $\delta$ at $z \leq 20$.

13 (A) PDFs of LMACHs at given overdensity $\delta$ in cells with Eulerian volume $(1.78/h \text{ Mpc})^3$ in 114/h Mpc box. (B) Same as Fig. 10(A). Here we also find that $\Delta_{\text{acc}} > 0$ with this cell size at any $\delta$ at $z \leq 20$.

14 (A) PDFs of HMACHs at given overdensity $\delta$ in cells with Eulerian volume $(3.56/h \text{ Mpc})^3$ in 114/h Mpc box. (B) Same as Fig. 10(A). Here we also find that $\Delta_{\text{acc}} > 0$ with this cell size at any $\delta$ at $z \leq 16$ except at $\delta = -0.1$ at $z = 15.6$ where $\Delta_{\text{acc}} = 0$. Overall, at $z = 15.6$, PDFs of HMACHs at any $\delta$ may be approximated well by pure Poisson distribution, as the ratio $\Delta_{\text{acc}}/[N]$ is small.

15 (A) PDFs of HMACHs at given overdensity $\delta$ in cells with Eulerian volume $(1.78/h \text{ Mpc})^3$ in 114/h Mpc box. (B) Same as Fig. 10(A). The outliers compared to pure Poisson distribution is overall small, albeit $\Delta_{\text{acc}}$ is comparable to $[N]$ at $\delta \gtrsim 0.5$. In some cases $\Delta_{\text{acc}} < 0$, which may be an indication of overall anti-correlation in the sub-cell scale such that $z_{12} < 0$ (see equation 19); however, the ratio is too small to draw a firm conclusion. Overall, PDFs of HMACHs may be approximated reasonably well by pure Poisson distribution in the redshift range $6 \leq z \leq 15$ and the range of $\delta \leq 0.5$. Nonlinear Bias of Cosmological Halo Formation
Conditional sub-scale correlation function $\xi_{12}(\delta)$ of minihaloes and the cumulative contribution to $\Delta_{\text{acc}}(\delta)$, inside 20/$h$ Mpc box with $V_{\text{cell}} = (0.45/h)^3$ Mpc. The top- and middle-row subpanels show $\xi_{12}(\delta)$ as a function of two-point distance $r \equiv r_{12}$, and the bottom-row subpanels show $\Delta_{\text{acc}}(\delta; < r) \equiv \left(\frac{|N|}{V_{\text{cell}}}\right)^2 \int_{-r}^{<r} dV_1 dV_2 \xi_{12}(\delta; r_{12})$.

Sames as fig 16 but with $V_{\text{cell}} = (0.15/h)^3$ Mpc.

Same as fig 16 but for LMACHs inside 114/$h$ Mpc box with $V_{\text{cell}} = (3.56/h)^3$ Mpc.

Same as fig 16 but for HMACHs inside 114/$h$ Mpc box with $V_{\text{cell}} = (3.56/h)^3$ Mpc.

Same as fig 20 but with $V_{\text{cell}} = (1.78/h)^3$ Mpc.

Linear bias parameter $B_1$ as a function of $\nu$ and $p$ (equation 31). The linear bias parameter of MW corresponds to a case with $p = 1$ ($q = 0$ accordingly). Note that $B_1$ is in general a non-monotonic function of $\nu$: both very abundant ($\nu \ll 1$) and very rare ($\nu \gg 1$) haloes are weakly biased to the first order in $\delta$, while the commonly used $b_{\text{ln}} = B_1(p = 1)$ of MW is a monotonic function of $\nu$. Because $p \neq 1$ in practice, care needs to be taken when using $b_{\text{ln}}$ for very rare haloes.
Figure 1. Unconditional, mean mass functions of MHs found in small-box N-body simulations. Plotted are the mass functions (each top panel) inside 20/h Mpc (black, solid) and 6/h Mpc (purple, dashed) boxes, along with the Sheth-Tormen (blue, dotted) and Press-Schechter (red, dot-dashed) mass functions, all integrated over equal-size logarithmic mass bins, and the ratios of these mass functions (each bottom panel with the same line types) to the mass function inside the 20/h Mpc box. The error bars represent 1σ standard deviation in each mass bin.
Figure 2. Correlation between the number of minihaloes \( N_{5:8} \) and the cell overdensity \( \delta \), when the Eulerian volume of the cell is chosen to be \((0.45/h\ Mpc)^3\). Data points are from N-body simulations in the 6.3/h Mpc box (red, larger dot) and 20/h Mpc box (slate blue, smaller dot), which are sampled by 14\(^3\) and 45\(^3\) cells, respectively. Theoretical predictions for active cells (eq. 25) based on \((dn/dM)_{\text{N-body}, b}\) (solid, black; eq. 15), \((dn/dM)_{\text{PS}, b}\) (dotted, blue; eq. 9), the one by the linear bias approximation (long-dashed, blue; eq. 13) combined with \((dn/dM)_{\text{N-body}, b}\) and the one by the 2nd order approximation with a zero-point offset \( B_0 \) (dot-dashed, blue; equations 29-34) also combined with \((dn/dM)_{\text{N-body}, b}\) are plotted for comparison.
Figure 3. Correlation between the fraction of mass collapsed into minihaloes ($f_{\text{coll}, 5:8}$) and the cell overdensity $\delta$ in the 6.3/h Mpc box, where the box is sampled by $14^3$ grid-cells. Conventions for plotting follow those of Fig. 2.
Figure 4. Same as Fig. 2 except that the volume of the cell is now \((0.15/h \text{ Mpc})^3\). The 6.3/h Mpc box (red circle) and 20/h Mpc box (green dot) are sampled by 44\(^3\) and 135\(^3\) cells, respectively.
Figure 5. Correlation between the fraction of mass collapsed into minihaloes ($f_{\text{coll}, 5.8}$) and the cell overdensity $\delta$ in the 6.3/h Mpc box and 20/h Mpc box, sampled by $44^3$ and $135^3$ cells, respectively. Conventions for plotting follow those of fig 3.
Figure 6. Correlation between the fraction of mass collapsed ($f_{\text{coll}, 8:9}$) into LMACHs ($M = 10^8 - 10^9 M_{\odot}$) and the cell overdensity $\delta$ in the 114/h Mpc box, where the box is sampled by $32^3$ grid-cells. Conventions for plotting follow those of Fig. 2 except for the data points (red point).
Figure 7. Correlation between the fraction of mass collapsed ($f_{\text{coll},8.9}$) into LMACHs ($M = 10^8 - 10^9 M_\odot$) and the cell overdensity $\delta$ in the 114/h Mpc box, where the box is sampled by $64^3$ grid-cells. Conventions for plotting follow those of Fig. 2 except for the data points (red point).
Figure 8. Correlation between the fraction of mass collapsed \( (f_{\text{coll}, \, 9:12.5}) \) into HMACHs \( (M = 10^9 - 10^{12.5} M_\odot) \), where the maximum mass is roughly the mass of a cell) and the cell overdensity \( \delta \) in the 114/h Mpc box, where the box is sampled by \( 32^3 \) grid-cells. Conventions for plotting follow those of Fig. 2 except for the data points (red point).
Figure 9. Correlation between the fraction of mass collapsed ($f_{\text{coll}, 8:9}$) into HMACHs ($M = 10^9 - 10^{11.5} M_\odot$, where the maximum mass is roughly the mass of a cell) and the cell overdensity $\delta$ in the $114/h$ Mpc box, where the box is sampled by $64^3$ grid-cells. Conventions for plotting follow those of Fig. 2, except for the data points (red point).
Figure 10. (A) PDFs of the number of minihaloes $N_{5.8}$ at given overdensity $\delta$ (denoted on top of each subpanel) of cells with Eulerian volume $(0.45/h \text{ Mpc})^3$ observed in $20/h \text{ Mpc}$ box. The horizontal and vertical axes represent $N_{5.8}$ and $\log P_{\text{cell}}(N_{5.8}|\delta)$, respectively. The data from simulation (histogram) is compared to the pure Poisson PDF (black, dotted) and the super-Poissonian PDF (equation 23; red, dashed). (B) Variances. Plotted are the observed total variance $(N - [N])^2$ (black, solid), the purely Poissonian variance $[N]$ (blue, dashed), the observed excess $\Delta_{\text{sec (obs)}} = (N - [N])^2 - [N]$ (red, dotted; eq. 20), and a value calculated from the sub-cell correlation function, $\Delta_{\text{sec (fit)}}$ (cyan, dot-dashed; eq. 19). The range of $\delta$ and $z$ are selected such that the number of cells with given $\delta$ (binned properly as described in Section 4.3) at given $z$ exceeds 200 for the statistical reliability of the calculated variances. The ratio $\Delta_{\text{sec}}/[N]$, plotted in the bottom panels, quantifies the excess of variance over the purely Poissonian one, $[N]$. 
Figure 11. (A) Same as Fig. 10A, but of cells with Eulerian volume $(0.15/h\,\text{Mpc})^3$ in $20/h\,\text{Mpc}$ box. (B) Same as Fig. 10B. Here we also find that $\Delta_{\text{sc}} > 0$ with this cell size at any $\delta$ at $z \leq 28$. Note the hugely nonlinear regime of $\delta$ observed at $z \leq 12$. 

$\log[\text{PDF}(N_{5.8}|\delta)]; V_{\text{cell}} = (0.15/h\,\text{Mpc})^3$
Figure 12. (A) PDFs of LMACHs at given overdensity $\delta$ in cells with Eulerian volume $(3.56/h \text{ Mpc})^3$ in $114/h \text{ Mpc}$ box. (B) Same as Fig. 10(B). Here we also find that $\Delta_{\text{sc}} > 0$ with this cell size at any $\delta$ at $z \leq 20$. 
Figure 13. (A) PDFs of LMACHs at given overdensity $\delta$ in cells with Eulerian volume $(1.78/h\,\text{Mpc})^3$ in $114/h\,\text{Mpc}$ box. (B) Same as Fig. 10(B). Here we also find that $\Delta_{\text{acc}} > 0$ with this cell size at any $\delta$ at $z \leq 20$. 
Figure 14. (A) PDFs of HMACHs at given overdensity $\delta$ in cells with Eulerian volume $(3.56/h \text{ Mpc})^3$ in $114/h \text{ Mpc}$ box. (B) Same as Fig. 10(B). Here we also find that $\Delta_{\text{scc}} > 0$ with this cell size at any $\delta$ at $z \leq 16$ except at $\delta = -0.1$ at $z = 15.6$ where $\Delta_{\text{scc}} = 0$. Overall, at $z = 15.6$, PDFs of HMACHs at any $\delta$ may be approximated well by pure Poisson distribution, as the ratio $\Delta_{\text{scc}}/\langle N \rangle$ is small.
Figure 15. (A) PDFs of HMACHs at given overdensity $\delta$ in cells with Eulerian volume $(1.78/h \text{ Mpc})^3$ in 114/$h$ Mpc box. (B) Same as Fig. 10(B). The outliers compared to pure Poisson distribution is overall small, albeit $\Delta_{\text{sc}}$ is comparable to $[N]$ at $\delta \gtrsim 0.5$. In some cases $\Delta_{\text{sc}} < 0$, which may be an indication of overall anti-correlation in the sub-cell scale such that $\xi_{12} < 0$ (see equation 15); however, the ratio is too small to draw a firm conclusion. Overall, PDFs of HMACHs may be approximated reasonably well by pure Poisson distribution in the redshift range $6 \lesssim z \lesssim 15$ and the range of $\delta \lesssim 0.5$. 

Nonlinear Bias of Cosmological Halo Formation
Figure 16. Conditional sub-scale correlation function $\xi_{12}(\delta)$ of minihaloes and the cumulative contribution to $\Delta_{acc}(\delta)$, inside $20/h$ Mpc box with $V_{\text{cell}} = (0.45/h)^3$ Mpc. The top- and middle-row subpanels show $\xi_{12}(\delta)$ as a function of two-point distance $r \equiv r_{12}$, and the bottom-row subpanels show $\Delta_{acc}(\delta; < r) \equiv \left(\frac{N}{V_{\text{cell}}}\right)^2 \int_{r}^{<r} dV_1 dV_2 \bar{\xi}_{12}(\delta; r_{12})$. 
Figure 17. Sames as fig. 16 but with $V_{\text{cell}} = (0.15/h)^3 \text{Mpc}$. 
Figure 18. Same as fig 16 but for LMACHs inside 114/h Mpc box with $V_{\text{cell}} = (3.56/h)^3$ Mpc.
Figure 19. Same as fig 18 but with $V_{\text{cell}} = (3.56/h)^3 \text{Mpc}$. 
Figure 20. Same as fig 16 but for HMACHs inside $114/h$ Mpc box with $V_{\text{cell}} = (3.56/h)^3$ Mpc.
Figure 21. Same as fig 20 but with $V_{\text{cell}} = (1.78/h)^3\text{ Mpc}$. 
Figure 22. Linear bias parameter $B_1$ as a function of $\nu$ and $p$ (equation 31). The linear bias parameter of MW corresponds to a case with $p = 1$ ($q = 0$ accordingly). Note that $B_1$ is in general a non-monotonic function of $\nu$: both very abundant ($\nu \ll 1$) and very rare ($\nu \gg 1$) haloes are weakly biased to the first order in $\delta$, while the commonly used $b_{\text{lin}} = B_1 (p = 1)$ of MW is a monotonic function of $\nu$. Because $p \neq 1$ in practice, care needs to be taken when using $b_{\text{lin}}$ for very rare haloes.