The clustering of the first galaxy haloes

Darren S. Reed,† Richard Bower, Carlos S. Frenk, Adrian Jenkins and Tom Theuns

1 Theoretical Astrophysics Group/ISR-1, Los Alamos National Laboratory, PO Box 1663, MS 627, Los Alamos, NM, 87545 USA
2 Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE
3 Department of Physics, University of Antwerp, Campus Groenenborger, Groenenborgerlaan 171, B-2020 Antwerp, Belgium

Accepted 2008 December 3. Received 2008 November 5; in original form 2008 April 1

ABSTRACT
We explore the clustering properties of high-redshift dark matter haloes, focusing on haloes massive enough to host early generations of stars or galaxies at redshift 10 and greater. Haloes are extracted from an array of dark matter simulations able to resolve down to the 'mini-halo' mass scale at redshifts as high as 30, thus encompassing the expected full mass range of haloes capable of hosting luminous objects and sources of reionization. Halo clustering on large-scales agrees with the Sheth, Mo & Tormen halo bias relation within all our simulations, greatly extending the regime where large-scale clustering is confirmed to be 'universal' at the 10–20 per cent level (which means, e.g., that 3σ haloes of cluster mass at z = 0 have the same large-scale bias with respect to the mass distribution as 3σ haloes of galaxy mass at z = 10). However, on small-scales, the clustering of our massive haloes (\(\gtrsim 10^9\ h^{-1}\ M_\odot\)) at these high redshifts is stronger than expected from comparisons with small-scale halo clustering extrapolated from lower redshifts. This implies 'non-universality' in the scale-dependence of halo clustering, at least for the commonly used parametrizations of the scale-dependence of bias that we consider. We provide a fit for the scale-dependence of bias in our results. This study provides a basis for using extraordinarily high-redshift galaxies (redshift \(\sim 10\)) as a probe of cosmology and galaxy formation at its earliest stages. We show also that mass and halo kinematics are strongly affected by finite simulation volumes. This suggests the potential for adverse affects on gas dynamics in hydrodynamic simulations of limited volumes, such as is typical in simulations of the formation of the ‘first stars’, though further study is warranted.

Key words: methods: N-body simulations – galaxies: formation – galaxies: haloes – cosmology: theory – dark matter.

1 INTRODUCTION
Dark matter haloes are formed from fluctuations in the matter density field as characterized by the matter power spectrum. A theoretical understanding of the relationship between the matter power spectrum and the numbers and clustering of dark matter haloes provides an essential link between cosmological parameters and the properties of haloes in which observable (or potentially observable) galaxies live. Haloes form from gravitationally induced non-linear collapse of mass overdensities that grow from primordial density fluctuations. Thus, halo clustering properties in a universe composed of cold dark matter are determined entirely by the matter power spectrum, or equivalently, by the cosmological parameters. The aim of this paper is to use dark matter simulations to quantify the relation between the matter power spectrum and resulting halo clustering properties, focusing on haloes of masses capable of star or galaxy formation during the era of reionization. It is beyond the scope of this paper to incorporate baryon physics to model galaxy formation within haloes. However, the distribution of haloes, which we explore, can be used as a basis for studies that populate haloes with galaxies to model galaxy clustering properties, which can then be used to probe cosmology and galaxy formation at high redshifts.

Halo clustering and its evolution have been studied extensively by a number of authors, mainly focusing on cluster, group, or galaxy mass scales at low redshifts (e.g. White et al. 1987; Mo & White 1996; Jing 1998; Sheth & Tormen 1999; Colberg et al. 2000; Taruya & Suto 2000; Hamana et al. 2001; Yoshikawa et al. 2001; Seljak & Warren 2004; Tinker et al. 2005; Wechsler et al. 2006; Wetzel et al. 2007; Angulo, Baugh & Lacey 2008; Basaklos, Plionis & Ragone-Figueroa 2008). However, a few studies have explored the halo or galaxy distribution in the reionizing era (e.g. Zahn et al. 2007; Cohn & White 2008; Iliev et al. 2008; Ricotti, Gnedin &
Shull 2008), an epoch, just beginning to be targeted by a number of surveys, expected to provide a new probe of cosmology and the physics of galaxy formation. At these high redshifts, haloes large enough to host stars and galaxies are expected to form from much rarer fluctuations (i.e. collapsed from higher sigma overdensities) than haloes that host typical galaxies or even groups and clusters in the low-redshift universe. This could lead to different halo clustering properties, which we will explore using numerical simulations.

The Press & Schechter formalism (Press & Schechter 1974) provides a general analytic framework for understanding halo formation, and has been improved by a number of subsequent authors (e.g. Bower 1991; Lacey & Cole 1993). Utilizing the extended Press & Schechter formalism, Mo & White (1996) developed a theoretical means of predicting halo clustering properties from the linear matter power spectrum. Numerical simulations have been utilized by a number of the fore-mentioned authors to validate and improve upon these works. Recently, we used a suite of high resolution numerical simulations to quantify the numbers of haloes during the reionizing era Reed et al. (2007) (see also e.g., Heitmann et al. 2006; Lučić et al. 2007). This work utilizes the same simulations to analyse the clustering properties of haloes.

We argue that halo clustering is at least somewhat ‘universal’ with respect to the linear matter fluctuation field. If halo clustering is exactly ‘universal’, then haloes of different mass and redshift formed from linear fluctuations of the same rarity will have identical clustering properties (see Section 2). For example, 3σ haloes of cluster mass at z = 0 would have the same relative clustering with respect to the mass distribution as 3σ haloes of galaxy mass at z = 10. Universality of halo clustering has been explored at lower redshifts by authors mentioned earlier, but has not been thoroughly addressed during the reionization period. We argue later that approximate universality holds for large-scale halo clustering to redshift 30, though it breaks down at small scales for the more massive haloes at z ≥ 10. The universality of large-scale clustering that we will show is a convenient feature (also held by the halo mass function; see e.g. Jenkins et al. 2001; Reed et al. 2003, 2007; Lučić et al. 2007) that should enable the results of our simulations, which are evolved only to z = 10, to be applicable to rare haloes over a wide redshift range. Rare, massive haloes attainable in current and future generations of high-redshift clustering surveys should have large-scale clustering strength relative to the mass, that is bias, similar to that of equally rare, massive haloes at lower redshifts (e.g. clusters, quasi-stellar object hosts).

The mass range of high-redshift haloes that we are able to resolve includes virtually all haloes large enough to form stars or galaxies (see e.g. Reed et al. 2005) and references therein for a review of high-redshift star formation. The first stars are expected to form within ‘mini-haloes’ of ~10^5−7 M⊙ where primordial gas cools and contracts mainly by H2 cooling (e.g. Bromm, Coppi & Larson 1999, 2002; Abel, Bryan & Norman 2000, 2002; Yoshida et al. 2003; O’Shea & Norman 2007). The ‘first galaxies’ may form within haloes of virial temperatures greater than ~10^6 K (~10^6 M⊙ at z ~ 10), where gas is able to cool via collisionally induced atomic hydrogen cooling, a more effective process. One should keep in mind that metal enrichment or other feedback effects may have a large influence on what types of haloes can cool and transform gas into stars sufficiently to resemble a ‘galaxy’.

Using extremely high-redshift galaxies for cosmology has some particular advantages. Their small masses allow the matter power spectrum to be probed on very small scales. For example, warm dark matter models with filtering lengths ≤100 h^−1 kpc can be tested if the abundance and bias of 10^9 h^−1 M⊙ hosts of the ‘1st galaxies’ can be measured. Additionally, ‘1st galaxies’ can potentially enable observations of the first stages of galaxy formation.

 Halo clustering can, in principle, be used to constrain both cosmology and the physics of galaxy formation with extremely high-redshift galaxies (z ≥ 6) as found, for example, by the Lyman-break (dropout) technique (e.g. Steidel et al. 1999) or by the redshifted Lyman α emission line (e.g. Taniguchi et al. 2005). We note that our redshift 10 haloes include the likely mass range of candidate z ~ 10 Lyman α emitters found behind a lensing cluster by Stark et al. (2007). Because the survey volume of that study is highly uncertain, so is the abundance, and hence the mass, of dark halo hosts of these objects. As in future ever more sensitive observations target similar galaxies, and with the upcoming launch of James Webb Space Telescope (JWST), their clustering properties should be measurable. This will constrain the properties of the haloes in which they live. It is thus essential to develop our theoretical understanding of halo and galaxy clustering in this regime.

In this study, we examine halo bias in the reionizing universe. In Section 2, we review briefly theoretical studies of halo bias. In Section 3, we discuss our suite of simulations and analysis techniques. We present measurements of halo bias and its scale-dependence from our simulations in Section 4. In Section 5, we analyse the kinematics of haloes and mass, and address the sensitivity of the pairwise velocity dispersion to the volume of the simulation. We conclude with a discussion of how our results can be used as a basis for using reionizing era galaxies to explore cosmology and galaxy formation.

2 HALO CLUSTERING THEORY

Clustering can be quantified by counting the numbers of pairs as a function of separation relative to that of a random distribution, as in the two-point correlation function,

\[ \xi(r) = \frac{N_{\text{pairs}}(r)}{N_{\text{pairs random}}(r)} - 1. \]  

(1)

Haloes can be either more or less clustered than the matter fluctuation spectrum, as described by the bias, which we compute as

\[ b = \sqrt{\xi_{\text{inh}}(r)/\xi_{\text{ran}}(r)}, \]  

(2)

where \( \xi_{\text{inh}} \) is the halo two-point correlation function and \( \xi_{\text{ran}} \) is the same for the mass. The two-point correlation function is directly related to the power spectrum by the Fourier transform, expressed as

\[ \xi(r) = \frac{1}{2\pi^2} \int_0^\infty P(k) \frac{\sin(kr)}{kr} k^2 dk. \]  

(3)

On large, linear scales, halo formation modelled by extended Press & Schechter formalism (e.g. Bower 1991; Lacey & Cole 1993) can be used to predict halo clustering. In this model, halo formation occurs when linear fluctuations within a global Gaussian random density field grow gravitationally to reach a critical threshold for collapse, \( \delta_c \), computed for spherical overdensities. Mo & White (1996) utilized this model to predict halo bias and demonstrated that bias should be scale-independent of sufficiently large scales. In this formalism, the bias is closely related to the mass function, which has been shown by Sheth & Tormen (1999), Sheth, Mo & Tormen (2001, hereafter SMT) and Sheth & Tormen (2002) to be better described by an ellipsoidal halo collapse model. The ellipsoidal collapse model of SMT results in large-scale bias consistent with simulations of, for example, Sheth & Tormen (1999) and Colberg

© 2009 The Authors. Journal compilation © 2009 RAS, MNRAS 394, 624–632

Downloaded from https://academic.oup.com/mnras/article-abstract/394/2/624/1069171

on 27 July 2018
et al. (2000), and given by

\[
b_{\text{SMT}} = 1 + \frac{1}{\sqrt{\delta_a}} \left[ \sqrt{\alpha(a^2) + \sqrt{b(a^2)^{1-c}}} \right] - \frac{(av)^2}{1-b(1-c)(1-c/2)},
\]

(4)

where \(a = 0.707, b = 0.5, c = 0.6, v = \delta_c/\sigma(m), \delta_c = 1.686, \) and \(\sigma(m)\) is the rms linear overdensity in top-hat spheres of mass \(m\). For an infinite volume,

\[
\sigma^2(m) = \frac{D^2(z)}{2\pi^2} \int_0^\infty k^2 W^2(k; m) dk,
\]

where \(P(k)\) is the linear power spectrum of the density fluctuations at \(z = 0\), and \(W(k; m)\) is the Fourier transform of the real-space top-hat filter and \(D(z)\) is the growth factor of linear perturbations normalized to unity at \(z = 0\) (Peebles 1993). The above bias relation is universal for \(0.5 < z < 1\) and \(m < 0.707\times10^{10}\).

3 TECHNIQUES AND ANALYSIS

3.1 The simulations

This study consists of a subset of the simulations presented in Reed et al. (2007), summarized in Table 1. These runs consist of dark matter particles evolved from linear initial conditions using a modified version of the parallel N-body code \textsc{gadget2} (Springel 2005; Springel et al. 2005). The wide range of simulation volumes allows us to model haloes over a large range in mass, and to determine empirically the effects of finite simulation volume, discussed below. Consequently, our suite of simulations has a large enough effective dynamic range to model haloes covering the full mass range able to form stars in the reionizing era. Haloes are selected by the friends-of-friends algorithm (Davis et al. 1985) where all particles separated by less than 0.2 times the mean interparticle separation are linked into a common halo. We have tested that our results are not sensitive to the choice of halo finder by verifying that haloes defined using a spherical overdensity of 178 times mean, another commonly used halo definition, have clustering properties similar to our friends-of-friends haloes (see Section 4.1 for further discussion).

| \(N_{\text{runs}}\) | \(L_{\text{box}}\) \(h^{-1}\)Mpc | \(m_{\text{part}}\) \(h^{-1}\)M⊙ | \(N_{\text{part}}\) | \(r_{\text{soft}}\) \(h^{-1}\)kpc |
|---|---|---|---|---|
| 11 | 1.0 | 1.1 \times 10^{10} | 400 | 0.125 |
| 1 | 2.5 | 1.1 \times 10^{10} | 1000 | 0.125 |
| 1 | 4.64 | 1.1 \times 10^{10} | 1000 | 0.58 |
| 2 | 11.6 | 1.1 \times 10^{10} | 1000 | 0.58 |
| 1 | 20 | 8.7 \times 10^{9} | 400 | 2.5 |
| 2 | 50 | 8.7 \times 10^{9} | 1000 | 2.4 |

Figure 1. Top panel: correlation function for haloes (\(\xi_{hh}\), solid) and particles (\(\xi_{mm}\), dashed) for all haloes of mass \(\sim 10^{10} h^{-1} M_\odot\) extracted from a range of simulation volumes, at \(z = 10\). Here, and in subsequent figures, box size and particle mass labels have units of \(h^{-1}\)Mpc and \(h^{-1} M_\odot\), respectively. Error bars on \(\xi_{hh}\) (solid) are 1\(\sigma\) Poisson from halo pair counts per bin. Bottom panel: bias, \(\xi_{hh}/\xi_{mm}\), for a range of simulation volumes. Bold portions denote the most reliable range of scales for the computed bias, extending to approximately 0.1\(L_{\text{box}}\); the agreement of different box sizes demonstrates the weakness of finite-volume effects on halo bias.

3.2 Finite volumes and their effects on spatial bias

The effects of the finite simulation volume might be expected to alter the clustering properties of the simulation. Finite simulation volumes lack power with wavelengths larger than the box, and suffer from discreteness of power especially for long wavelength modes (e.g. Barkana & Loeb 2004; Bagla & Ray 2005; Sirko 2005; Bagla & Prasad 2006; Power & Knebe 2006; Lukić et al. 2007; Reed et al. 2007). The resulting effects on clustering can be strong for the small volumes required to model the low-mass haloes relevant for high-redshift star formation. But fortunately, finite-volume effects are similar for dark matter and for haloes, as we show in Fig. 1. As a result, the finite-volume effects on halo bias are much weaker than the effects on absolute clustering properties of mass or haloes. For this reason, the halo bias measured within a simulation is reliable over a substantial range of scales smaller than the length of the box.

Fig. 1 shows an example of the two-point correlation function, \(\xi(r)\), for dark matter and for haloes with virial temperatures greater than \(10^4\) K at \(z = 10\) (approximately \(\sim 10^4 h^{-1} M_\odot\), the relation \(M_{10^4} = 10^{8} \times [10/(1+z)^{1/2}] h^{-1} M_\odot\), and spanning a factor of 10 in halo mass. The bottom panel shows halo bias for several box sizes, demonstrating that any systematic affects on bias due to finite volume are weak provided that the scale sampled is sufficiently smaller than the box size. We find that as a conservative choice, the bias measured from pair separations below \(0.1L_{\text{box}}\) is adequate for our desired level of accuracy (\(\lesssim 10\) per cent) in our simulations. The bold portion of the bias curves in Figs 1 and 2 extend only to this scale. Although the scatter in bias is close to \(\sim 50\) per cent at small scales and large masses where numbers of pairs are small, we demonstrate later that we are able to measure bias at large scales.
to \( \lesssim 10 \) per cent, using \( 0.1L_{\text{box}} \) as a maximum pair separation. Our smallest box size of \( 1 h^{-1} \) Mpc, which might be expected to be most likely to suffer finite-volume effects, does not have enough haloes of this mass for us to compute reliably a correlation function; however, we have confirmed using smaller haloes that the bias found in the smallest boxes agrees with that in larger boxes.

The decrease in bias at small scales is mainly due to the fact that it is not possible to have a halo pair separated by less than twice the halo radius, sometimes referred to as the halo exclusion effect (Benson et al. 2000). On very small scales, the uncertainty increases due to small numbers of closely separated pairs, as reflected by the increased scatter.

We have just shown that halo bias is not affected strongly by box size, an argument that is strengthened by the agreement among overlapping boxes in Fig. 2. However, because individual realizations and different simulation volumes each have a unique form of \( \sigma(m) \), consideration of the universality of halo bias can be strengthened if one uses the relation between \( \sigma \) and \( M \) specific to each realization as the effective universal mass variable. This mitigates the effects of missing large-scale power and reduces run-to-run scatter. For a periodic cosmological simulation, the smoothed rms linear overdensity, \( \sigma \), is given by the discrete analogue of equation (5),

\[
\sigma^2(m) = D^2(z) \sum_k |\delta_k|^2 W^2(k/m),
\]

where \( |\delta_k| \) refers to the linear amplitude of the Fourier modes at \( z = 0 \), and \( D \) and \( W \) are the same as in equation (5). This approach has been successful in the parametrization of the halo mass function (see Reed et al. 2007 for detailed discussion). We discuss further in the Appendix the importance of considering finite-volume effects for estimates of large-scale clustering.

### 4 RESULTS: HALO BIAS

#### 4.1 Scale-dependence

We present halo bias as a function of scale for haloes of a range of masses at redshift 10 in Fig. 2. At large scales, the halo bias approaches the number-weighted large-scale bias predictions of SMT, given by

\[
b_{\text{SMT}}(>m_{\text{min}}) = n(>m_{\text{min}})^{-1} \int_{m_{\text{min}}}^{\infty} b_{\text{SMT}}(m) \frac{dn}{dm} dm,
\]

where \( b_{\text{SMT}}(m) \) is given in equation (4), and we take \( dn/dm \) to be the mass function given by Reed et al. (2007), equation (11). We explore further the large-scale bias in the next section. First, we show the scale-dependence and compare with fits from clusters in simulations of much lower redshifts. The scale-dependence has been parametrized in terms of the halo pair separation, \( r \),

\[
b(m, r, z) = b_{\text{SMT}}(m, z)[1 + b_{\text{SMT}}(m, z)\alpha(r, z)]^\alpha,
\]

with \( \alpha = 0.15 \) by Hamana et al. (2001) and \( \alpha = 0.35 \) by Diaferio et al. (2003), where \( b_{\text{SMT}} \) is the bias relation of Sheth & Tormen (1999), which is similar to \( b_{\text{SMT}} \). The steeper scale-dependence of Diaferio et al. (2003) is consistent with our small haloes, but is not steep enough to match the most massive haloes in our simulations. The Hamana et al. (2001) bias relation (not plotted) is much shallower than that of Diaferio et al. (2003), making for a poor match to the scale-dependence in our data. The bias fit of Tinker et al. (2005), which is parametrized in terms of the non-linear mass correlation function, \( \xi(r) \), rather than the linear variance in spheres, as in the above relation, implies even weaker scale-dependence than that of Hamana et al. (2001) when extrapolated to our mass and redshift range, and so is not consistent with our results.

It requires a large extrapolation to apply low-redshift fits of cluster bias to our higher redshifts, lower masses, and much rarer (larger v) haloes than previously tested, so it is perhaps not surprising that we find a steeper scale-dependence for many of our haloes. Even so, the reasonable level of agreement for our lower mass haloes suggests 'partial universality'. Our haloes at redshift 10 are well fit by the following relation:

\[
b(m, r, z) = b_{\text{SMT}}(m, z)[1 + 0.03b_{\text{SMT}}(m, z)^3\alpha(r, z)^2]\]

(ignoring halo exclusion effects at small separations). We have checked that our fit remains consistent with our simulation outputs from redshifts 10 to 30, though the uncertainties and scatter become greater at high redshifts. This fit, however, is not intended to be valid at redshifts lower than we have considered. In fact, if extrapolated to the lower redshifts probed by Diaferio et al. (2003), our fit has significantly steeper scale-dependence than their simulations, reflecting non-universality. We have confirmed the lack of universality of this parametrization of the scale-dependent bias using separate simulations of larger volumes evolved to low redshifts. Though the halo samples considered here were divided into cumulative samples for illustrative purposes, the above fit agrees with the bias obtained when samples are restricted to a much narrower mass range of a factor of 2. For visual clarity, we have chosen to plot in Fig. 2 the value of \( b_{\text{SMT}} \) for an infinite volume rather than the individual finite-volume corrected value for each simulation. We demonstrate in the Appendix that the finite-volume corrected large-scale bias of each realization is a much better description of the bias measured in simulations, particularly for the smallest boxes.

A number of factors may affect the universality of the scale-dependence of halo bias. The slope of the linear power spectrum on halo pre-collapse scales is much steeper in our simulations because...
the halo masses are smaller relative to that typical of low-redshift simulations. The steeper power spectrum may affect halo clustering (see e.g. Jing 1998) and cause the apparent non-universality with respect to low-redshift clusters. Another contributing factor may be that the effective definition of a halo changes based on its concentration and environment (Lukić et al. 2008). The low concentrations and dominance of large filamentary structure at high redshift (Gao, Springel & White 2005) may cause the halo finder to combine or split haloes differently relative to their Lagrangian (linear) initial density fluctuations. We have checked the effect of halo definition by verifying that the bias of haloes defined using a spherical overdensity of 178 times mean is consistent, though perhaps slightly higher and steeper, as might be expected from their lower abundance (Reed et al. 2007). Further theoretical work regarding halo formation is needed to understand fully the scale-dependence of halo clustering. We leave the search for a more universal parametrization of the scale-dependence of bias to future studies.

4.2 Large-scale bias

We compute the large-scale bias for our halo sample in Fig. 3. The SMT bias relation is consistent with our data at all redshifts while the Mo & White bias relation is inconsistent with our results. By going to very high redshift, we are able to measure the bias of haloes formed from \( \sim 4\sigma \) fluctuations, representing a major improvement upon previous bias measurements at high redshift. In particular, the Millennium run bias data (from Gao et al. 2005) extend to 3\( \sigma \) haloes at \( z = 5 \), and Cohn & White (2008) compute bias for a narrower mass range at \( z = 10 \). Angulo et al. (2008) measure bias for \( \sim 4.5\sigma \) haloes, but at redshifts of 3 and below. While we have not explicitly plotted uncertainties, we note that the scatter among different simulations can be used to estimate visually an ensemble bootstrap error. This scatter suggests that the data are accurate to about 10 per cent in bias (ignoring any potential systematics). Given the scatter within our data, it is not possible to distinguish between the SMT bias relation and that of more precise fits such as Seljak & Warren (2004), Tinker et al. (2005) or the bias relation that would be derived using the mass function fit to these simulations (Reed et al. 2007).

The agreement with the plotted Millennium data (Springel et al. 2005) for our 2–3\( \sigma \) haloes confirms universality, and suggests that our box-size correction technique is successful. The evidence for universality of halo bias is strengthened by our agreement with Angulo et al. (2008) for 2–4\( \sigma \) haloes, where their results agree with the SMT bias. However, the source of the \( \sim 20 \) per cent disagreement between our data and the lower redshift bias at \( v \approx 1.5 \) from the Millennium run, with which the Angulo et al. (2008) bias appears to also agree, is not clear. Although this could indicate departure from universality at the \( \sim 20 \) per cent level (in bias) because the differences occur at small masses in the smallest boxes where the finite-volume corrections are largest (see Appendix), potential systematics are not easily ruled out. Our work verifies that estimates of halo bias derived from extended Press–Schechter theory and variants are valid for extremely rare haloes. Additionally, our comparison with other studies at lower redshifts significantly extends the confirmed regime of approximate universality (at the 10–20 per cent level) in mass, redshift and \( v \) of large-scale bias.

Some care must be taken when computing bias from haloes in finite volumes. Because the scale-dependence of the bias extends to scales approaching the maximum pair separations where bias measurements remain robust, a simple averaging of the bias over the reliably sampled range in scales would result in a systematic error (i.e. a ‘bias bias’). We thus utilize the scale-dependence of the bias presented earlier to fit more reliably the large-scale bias. The large-scale bias that is plotted in Fig. 3 is then computed by performing a least-squares fit using equation (9) and replacing \( b_{\text{SMT}} \) with the bias to be fit, as follows:

\[
b(m, r, z) = b_{\text{fit}}(m, z)[1 + 0.03b_{\text{fit}}(m, z)^3\sigma(r, z)^2].
\] (10)

To avoid peculiarities of bias for closely separated pairs where numbers are small and bias is steepest, bias is then fit over pairs separated by a range of 0.033–0.1\( L_{\text{box}} \), which also remains within the range shown previously to be most robust against finite box effects.

In Cohn & White (2008), the large-scale bias was taken to be the bias at 1.5\( h^{-1} \) Mpc. They caution that they may not have captured the bias at a scale large enough to reflect the asymptotic large-scale value. Indeed, their large-scale bias estimates are higher than ours over the range of overlap, covering 3–3.7\( \sigma \) haloes, also at \( z = 10 \). Their bias estimates are \( \approx 10–20 \) per cent higher than the SMT bias relation, whereas our results agree with the SMT bias. These differences are consistent with the overestimate in bias expected when estimating the large-scale bias from the value at 1.5\( h^{-1} \) Mpc (see Fig. 2).

5 RESULTS: HALO PAIRWISE VELOCITY DISPERSION

The kinematics of high-redshift galaxies represent an additional probe of cosmology and galaxy formation, potentially more sensitive than the spatial correlation function (e.g. Zhao, Jing & Borner 2002). The pairwise velocity dispersion, \( \sigma_{\text{rms}}(r) \), has been utilized extensively in low-redshift cosmological surveys (e.g. Davis & Peebles 1983). \( \sigma_{\text{rms}}(r) \) is defined as the one-dimensional rms velocity of pair members in the direction of the line-of-sight connecting the pair, separated by distance \( r \). From Fig. 4, it can be seen that the pairwise velocity dispersion of both mass and haloes tends to be suppressed by finite simulation volume. Particles and
haloes of identical mass in different size volumes have significantly different velocities; note that the pairwise velocity dispersion has an inherent mass-dependence, so objects of similar mass must be considered. In the case of haloes, some kinematic effects could be due to the difference in average halo mass of the sample due to the finite-volume suppression of high-mass haloes (see Reed et al. 2007). Pairwise dispersions for the $\sim10^8\ h^{-1}\ M_\odot$ halo sample, (covering $\sim10^6\ h^{-1}\ M_\odot$) extracted from a range of simulation volumes, at $z=10$ (the same sample as Fig. 1). The finite-volume effect on pairwise velocities decreases with close pair separation, presumably because the infall velocity of close pairs is dominated by the total mass of the pair. Bottom panel: pairwise velocity bias, $\sigma_{12,h}/\sigma_{12,mm}$ for a range of simulation volumes. Bold portions denote the most reliable range of scales previously estimated for spatial bias (extending to approximately 0.1$L_{\text{box}}$). The relative disagreement in pairwise velocities and bias for simulations of different volume indicate significant finite-volume effects in the kinematics of mass and haloes.

Figure 5. Pairwise velocity bias, $\sigma_{12,h}/\sigma_{12,mm}$, at $z=10$ for the same halo samples as in Fig. 2. The bold portions of the curves denote the range of pair separation where spatial bias was found to be relatively robust. In contrast to spatial bias, however, kinematic bias is affected significantly by finite simulation volume because particle and halo velocities are suppressed at different levels.

As an example, it is worth considering the simplest case where there is a one-to-one relation between halo mass and galaxy luminosity, that is if $L_{\text{halo}}(m,z)$ has no scatter, and exactly one galaxy occupies each halo. Here, a measurement of the luminosity function would allow one to halo derive the halo mass for an assumed cosmology directly from the mass function. The inferred mass then determines the halo bias which, in turn, can be used to constrain cosmological parameters, as illustrated in Fig. 6. Note that determining the galaxy bias from observations requires a simultaneous measurement of both matter clustering and galaxy clustering, a difficulty which we later discuss. Since all potentially observable objects at extremely high redshifts live in haloes much larger than $M_*$, the cosmological dependence of bias remains large for all observationally attainable halo abundances. In contrast, at low redshifts, the cosmology dependence of bias grows weak for small abundances. For the most abundant haloes, a bias measurement of 20 per cent accuracy could distinguish between a low $\sigma_8$ of 0.76 and a high $\sigma_8$ of 0.9 if measured from haloes of abundance $\sim1 \ h^{-1} \ M_\odot$ ($\sim10^9\ h^{-1}\ M_\odot$). Although neither of these normalizations are favoured by current cosmic microwave background data, they are in fact within $\sim(2–3)\sigma$ of the recent Wilkinson Microwave Anisotropy Probe (WMAP) 5 yr normalization (Komatsu et al. 2008). Moreover, the point remains that bias is generally more sensitive to cosmology at high redshifts, and at a wide range of halo abundances and mass.

The bias-abundance relation may remain largely unaltered even if intergalactic medium (IGM) absorption is significant. Iliev et al. (2008), for the case of high-redshift Lyman $\alpha$ sources, showed that dimming due to IGM absorption does not affect significantly the relation between observed number density and source clustering.
Once the large-scale bias and cosmology are determined, the scale-dependence of halo bias, to the degree that it is ‘universal’, is also determined. Comparison of the observed scale-dependence of galaxy clustering to that of haloes with the same large-scale bias thus provides a valuable indicator of local environmental effects on galaxy formation. The strong clustering and steep scale-dependence of the clustering for high-redshift haloes may cause feedback from neighbouring galaxies to affect star formation during these early times.

Many possible complications to this simple picture will need to be considered. For example, the large scales of reionizing bubbles (e.g. Iliev et al. 2006; Zahn et al. 2007) may correlate with galaxy clustering on large scales (e.g. Wyithe & Loeb 2007), which leads to the possibility of reionization affecting galaxy formation and galaxy clustering on large scales (e.g. Gnedin 2000) or multiple galaxies may occupy massive haloes, which creates a degeneracy in galaxy abundance between satellite numbers and halo mass. Similarly, there exists a degeneracy between halo mass and satellite numbers in the scale-dependence of clustering on small scales. These degeneracies are difficult to overcome observationally because of the need for clustering measurements on very small angular scales (see e.g. Bullock, Wechsler & Somerville 2002). However, one can use theoretical arguments, semi-analytic modelling and/or smaller volume simulations utilizing hydrodynamics and other detailed physics to infer basic galaxy properties such as the numbers, luminosities, and environmental influences (e.g. Iliev et al. 2006; McQuinn et al. 2007; Mesinger & Furlanetto 2007; Ricotti et al. 2008). Progress may then be made by combining the results of such detailed models of galaxy formation with determinations of halo clustering from dark matter simulations in order to produce detailed predictions for observable clustering properties. This will help to break the degeneracies between galaxy ‘feedback’ and halo clustering on small scales, as well as determine whether haloes tend to host multiple luminous galaxies. As ever deeper observations are made, such studies will allow us to probe galaxy formation at its earliest stages and to explore cosmology at the high-redshift frontier.

We close with a brief discussion of the (perhaps surprisingly) large suppression to the halo pairwise velocity dispersion by finite simulation volumes. Pairwise velocity dispersions are reduced by up to and beyond 50 per cent in our simulations, even on scales where halo spatial bias is well behaved. The general suppression of halo and mass velocities may have significant effects on hydrodynamic simulations where small volumes are often required because of computational expense. For example, volumes of \( \lesssim 1 \) Mpc are commonly employed in simulations that model the cooling of gas within high-redshift haloes to form the ‘first stars’ (e.g. Bromm et al. 1999, 2002; Abel et al. 2000, 2002; Yoshida et al. 2003; O’Shea & Norman 2007). This suggests the possibility that gas heating due to shocks of infalling material of merging haloes could be inhibited due to the suppression of pairwise velocities. Such effects have the potential to alter gas cooling and the ionization level of halo gas, which could affect star formation. However, because the velocity suppression diminishes for close pairs, it is not clear whether the kinematic suppression is actually a major problem for gas dynamics or star formation within simulations of the high-redshift universe. Any such problems could be minimized by the ‘zoom’ resimulation technique, consisting of a high resolution subvolume embedded within a larger and lower resolution cosmological volume, as in for example Gao et al. (2007). Further investigation of finite-volume effects on kinematics and any associated effects on gas properties are warranted.
ACKNOWLEDGMENTS

DR is a post-doc LANL, and is supported by the DOE through the IGPP, the LDRD-DR and the LDRD-ER programs at LANL. CSF acknowledges a Wolfson Royal Society research merit award. We thank Salman Habib, Katrin Heitmann, Zarija Lukić, Martin White, Joanne Cohn, Jeremy Tinker and Brian O’Shea for helpful discussions. We thank Volker Springel for use of an enhanced version of gadget-2. We thank Liang Gao for making his Millennium simulation results available to us. We are grateful for the suggestions of the anonymous referee.

REFERENCES

Abel T., Bryan G., Norman M. L., 2000, ApJ, 540, 39
Abel T., Bryan G., Norman M. L., 2002, Sci, 295, 93
Angulo R. E., Baugh C. M., Lacey C. G., 2008, MNRAS, 387, 921
Bagla J., Prasad J., 2006, MNRAS, 370, 993
Bagla J., Ray S., 2005, MNRAS, 358, 107
Barkana R., Loeb A., 2004, ApJ, 609, 474
Barkana R., Loeb A., 2005, ApJ, 624, 65
Basilakos S., Plionis M., Ragone-Figueroa C., 2008, ApJ, 678, 627
Benson A. J., Cole S. M., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, 311, 793
Bower R., 1991, MNRAS, 248, 332
Bromm V., Coppi P. S., Larson R. R., 1999, ApJ, 527, L5
Bromm V., Coppi P. S., Larson R. R., 2002, ApJ, 564, 23
Bullock J., Wechsler R., Somerville R., 2002, MNRAS, 329, 246
Cohn J., White M., 2008, MNRAS, 385, 2025
Colberg J. M., et al., 2000, MNRAS, 319, 209
Conroy C., Wechsler R., Kravtsov A. V., 2006, ApJ, 647, 201
Davis M., Peebles J., 1983, ApJ, 267, 465
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 381
Diaferio A., Nusser A., Yoshida N., Sunyaev R., 2003, MNRAS, 338, 433
Furlanetto S., Oh S., Briggs F., 2006, Phys. Rep., 433, 181
Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, 66
Gao L., Yoshida N., Abel T., Frenk C., Jenkins A., Springel V., 2007, MNRAS, 378, 449
Gnedin N., 2000, ApJ, 542, 535
Hamana T., Yoshida N., Suto Y., 2001, ApJ, 561, L143
Hamana T., Ouchi M., Shimasaku K., Kayo I., Suto Y., 2004, MNRAS, 347, 813
Hamana T., Yamada T., Ouchi M., Iwata I., Kodama T., 2006, MNRAS, 369, 1929
Heitmann K., Lukić Z., Habib S., Ricker P. M., 2006, ApJ, 642, L85
Iliev I., Mellema G., Pen U.-L., Merz H., Shapiro P., Alvarez M., 2006, MNRAS, 369, 1625
Iliev I., Shapiro P., McDonald P., Mellema G., Pen U., 2008, MNRAS, 391, 631
Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372
Jing Y. P., 1998, ApJ, L9
Kashikawa N. et al., 2006, ApJ, 637, 631
Kovac K., Somerville R., Rhoards J., Malhotra S., Wang J., 2007, ApJ, 668, 15
Komatsu E. et al., 2008, ApJS, in press (arXiv:0803.0457)
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lee K., Giavalisco M., Gnedin O., Somerville R., Ferguson H., Dickinson M., Ouchi M., 2006, ApJ, 642, 63
Lukić Z., Heitmann K., Habib S., Bashinsky S., Ricker P. M., 2007, ApJ, 671, 1160
Lukić Z., Reed D., Habib S., Heitmann K., 2008, ApJ, in press (arXiv:0803.3624)
McQuinn M., Hernquist L., Zaldarriaga M., Dutta S., 2007, MNRAS, 281, 75
Mesinger A., Furlanetto S., 2007, ApJ, 669, 663
Mo H., White S. D. M., 1996, MNRAS, 282, 347
O’Shea B., Norman M., 2007, ApJ, 654, 66
Peebles P., 1993, Principles of Physical Cosmology, Princeton Univ. Press, Princeton, NJ
Power C., Knebe A., 2006, MNRAS, 370, 691
Press W. H., Schechter P., 1974, ApJ, 187, 425
Reed D., Gardner J., Quinn T., Stadel J., Fardal M., Lake G., Governato F., 2003, MNRAS, 346, 565
Reed D. S., Bower R., Frenk C. S., Gao L., Jenkins A., Theuns T., White S. D. M., 2005, MNRAS, 363, 593
Reed D. S., Bower R., Frenk C. S., Jenkins A., Theuns T., 2007, MNRAS, 374, 2
Ricotti M., Gnedin N., Shull M., 2008, ApJ, 685, 21
Seljak U., Warren M., 2004, MNRAS, 355, 129
Shaw J. R., Lewis A., 2008, Phys. Rev. D, 78, 103512
Sheth R., Tormen G., 1999, MNRAS, 308, 119
Sheth R., Tormen G., 2002, MNRAS, 329, 61
Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1 (SM)
Sirko E., 2005, ApJ, 634, 728
Springel V., 2005, MNRAS, 364, 1105
Springel V. et al., 2005, Nat, 435, 629
Stark D., Ellis R., Nicholl J., Smith G., Santos M., 2007, ApJ, 663, 10
Steidel C., Adelberger K., Giavalisco M., Dickinson M., Pettini M., 1999, ApJ, 519, 1
Taniguchi Y. et al., 2005, PASJ, 57, 165
Takahashi R. et al., 2008, MNRAS, 389, 1675
Taruya A., Suto Y., 2000, ApJ, 542, 559
Tinker J., Weinberg D., Zheng Z., Zehavi I., 2005, ApJ, 631, 41
Wechsler R., Zentner A., Bullock J., Kravtsov A., Allgood B., 2006, ApJ, 652, 71
Wetzel A., Cohn J. D., White M., Holz D., Warren M., 2007, ApJ, 656, 139
White S. D. M., Frenk C. S., Davis M., Efstathiou G., 1987, ApJ, 313, 505
White M., Martini P., Cohn J., 2008, MNRAS, 390, 1179
Wyithe J., Loeb A., 2007, MNRAS, 375, 1034
Yoshida N., Abel T., Hernquist L., Sugiyama N., 2003, ApJ, 592, 645
Yoshikawa K., Taruya A., Jing Y. P., Suto Y., 2001, ApJ, 558, 520
Zahn O., Lidz A., McQuinn M., Dutta S., Hernquist L., Zaldarriaga M., Furlanetto S., 2007, ApJ, 654, 12
Zhao D., Jing Y., Borner G., 2002, ApJ, 581, 876

APPENDIX A:

Here, we demonstrate the importance of taking into account the finite simulation volume when considering large-scale halo bias. By computing $\sigma^2(m)$ from equation (6) instead of equation (5), the mean power at each discrete wavenumber present in the simulation is taken into account. This has the effect of shifting $\sigma(m_{\text{halo}})$ usually towards smaller values (larger values of $\sigma$). Fig. A1, which uses equation (5) to compute $\sigma(m_{\text{halo}})$, thereby ignoring the finite volumes, shows that two large effects are present if finite simulation volume is not treated appropriately. First, there is the significant run-to-run scatter in bias, approaching ~50 per cent for the smallest volume, to be compared with the ~10 per cent scatter in Fig. 3 where equation (6) is used to compute $\sigma(m_{\text{halo}})$. Secondly, the inferred simulation bias is significantly higher when finite volume is ignored, especially for the smallest boxes, and the larger haloes within them. This approach of finite-volume compensation does not correct for discreteness of phases, nor does it provide a correction for the deviations as a result of non-linear coupling to only a small number of modes at large scales (Takahashi et al. 2008).
Figure A1. This figure uses the sigma(m) computed for an infinite volume (equation 5) instead of that computed taking into account the power present in each realization (equation 6). Otherwise this figure is identical to Fig. 3 (including the Millennium run results from Gao et al. (2005). The larger run-to-run scatter in bias and the poorer fit to the SMT relation demonstrate the usefulness of correcting for reduced power and discrete modes in finite volumes.

However, the improvement in the run-to-run scatter, the better fit to the SMT bias relation, and the better agreement with the Millennium run results, all provide some level of assurance that this approach of correcting for finite volume is justified. Similar reductions in the run-to-run scatter and improved universality of the halo mass function were demonstrated in Reed et al. (2007), strengthening the importance and validity of correcting for finite simulation volume in cosmological simulations. These corrections allow one to simulate smaller haloes, and hence achieve a larger effective dynamic range, while maintaining their value as cosmological tools.