Large-scale bias and stochasticity of haloes and dark matter

Uroš Seljak1 and Michael S. Warren2

1Department of Physics, Jadwin Hall, Princeton University, Princeton, NJ 08544, USA
2Theoretical Astrophysics (T-6), Mail Stop B227, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Accepted 2004 August 10. Received 2004 July 30; in original form 2004 March 30

ABSTRACT

On large scales, galaxies and their haloes are usually assumed to trace the dark matter with a constant bias and dark matter is assumed to trace the linear density field. We test these assumption using several large N-body simulations with 3843–10243 particles and box sizes of 96–1152 h−1 Mpc, which can both resolve the small galactic-size haloes and sample the large-scale fluctuations. We explore the average halo bias relation as a function of halo mass and show that existing fitting formulae overestimate the halo bias by up to 20 per cent in the regime just below the non-linear mass. We propose a new expression that fits our simulations well. We find that the halo bias is nearly constant, \( b \sim 0.65–0.7 \), for masses below one-tenth of the non-linear mass. We next explore the relation between the initial and final dark matter in individual Fourier modes and show that there are significant fluctuations in their ratio, ranging from 10 per cent rms at \( k \sim 0.03 \frac{h}{\text{Mpc}} \) to 50 per cent rms at \( k \sim 0.1 \frac{h}{\text{Mpc}} \). We argue that these large fluctuations are caused by perturbative effects beyond the linear theory, which are dominated by long-wavelength modes with large random fluctuations. Similar or larger fluctuations exist between haloes and dark matter and between haloes of different mass. These fluctuations must be included in attempts to determine the relative bias of two populations from their maps, which would otherwise be immune to sampling variance.

Key words: cosmology: theory – dark matter – large-scale structure of Universe.

1 INTRODUCTION

Determination of the power spectrum of mass fluctuations and its redshift evolution is one of the main goals of modern observational cosmology. Its accurate measurement would allow us to test some of the most fundamental questions in cosmology today, such as the shape of the primordial power spectrum and its relation to fundamental theories of structure formation, the mass of the neutrino and the nature of dark energy.

In general there are two approaches to the measurement of the matter power spectrum. One is to measure galaxies, either in redshift space or in angular position (perhaps supplemented by photometric redshift information) and to assume they trace the dark matter. This assumption is believed to be valid on large scales, where the so-called linear bias model assumes that the galaxy density field is proportional to the matter density field times a free parameter called bias. While power-spectrum measurements of galaxies with modern surveys such as SDSS (Tegmark et al. 2004) or 2dF (Percival et al. 2002) have enormous statistical power, they can only determine the shape of the matter power spectrum and not its amplitude because of the bias uncertainties. This limits their use in the study of the growth-factor evolution, important for investigations of dark energy models. In addition, on small scales information from galaxy clustering is limited by the uncertainties in the relation between the galaxies and the dark matter, which make the bias scale-dependent. For this reason the small-scale information is usually discarded.

The situation with galaxies would not be as dim if we could determine the bias. Here we explore a method to determine galaxy bias based on determination of clustering amplitude of faint galaxies. These are likely to occupy low-mass haloes which, as we show in this paper, have a well-determined large-scale bias that is nearly independent of halo mass. While some fraction of these galaxies are satellites in larger haloes, this can be quantified and corrected for (see Seljak et al. 2004 for a first application of this method to the real data). In addition, there exist populations, such as IRAS galaxies, for which this fraction may be small. In these cases measuring the large-scale power-spectrum amplitude for these galaxies determines the matter power-spectrum amplitude as well.

The main problem with using faint galaxies as tracers of large-scale structure is that in a typical flux-limited survey faint galaxies occupy a small nearby volume, so the sampling variance errors for the power spectrum on large scales are large. However, we can still determine their bias relative to a population of brighter galaxies. If there is no stochasticity between the two populations then a direct comparison of the maps gives an accurate determination.
of the relative bias with no sampling variance. We can then use the power-spectrum determination of the brighter population, with smaller sampling variance errors because of the larger volume covered, to determine the power spectrum of the fainter population and of the matter itself. The limiting source of noise is the stochasticity between these fields, which we explore in this paper.

Another approach to determine matter fluctuations is to use weak-lensing induced correlations between background galaxy ellipticities. These are sensitive to the dark matter fluctuations directly and as such this approach holds the promise to improve upon the limitations of the galaxy clustering methods. Its main limitation is that it traces the dark matter in angular projection and has large sampling variance errors on large scales. This limits the statistical power of the weak-lensing surveys. On small scales the non-linear corrections, noise, intrinsic correlations and other systematic contaminations become significant, all of which may complicate the modelling.

A possible approach to achieve the best of both worlds is to combine the weak-lensing and galaxy clustering surveys: one can use the weak lensing to determine the galaxy bias and then use the 3D galaxy clustering information to improve on the statistical errors. One way to do this is to use galaxy–dark matter cross-correlation analysis from weak lensing and combine it with the galaxy auto-correlation analysis. If galaxies are tracing perfectly the dark matter then it suffices to have a few well-measured modes in both fields to determine the galaxy bias. This has been proposed as a way to get around the sampling variance in weak-lensing surveys (Pen 2004). In the absence of stochasticity it gives the bias (and so the dark matter power spectrum) without the usual sampling variance errors, assuming the analysis is done on the same patch of the sky and with the correct radial weighting of the galaxies to match that of the dark matter.

In both of these cases the underlying assumption is that there is no stochasticity between these fields on large scales. While there have been analytic attempts to address this assumption (Matsubara 1999), it has not been tested well with simulations in the past due to the lack of sufficient dynamic range (but see Casas-Miranda et al. 2002 for a related study). One must resolve the haloes small enough to be suitable as galaxy hosts (with typical masses at or below $10^{12} \, \text{M}_\odot$). At the same time, the simulations must be large enough so that many long-wavelength modes are sampled to determine the statistics of interest. We achieve this by using a set of new simulations with a larger dynamical range. The number of particles in these simulations, $10^9$–$10^7$, and their box size, $100$–$1000 \, \text{h}^{-1} \, \text{Mpc}$, allow a much better exploration of the halo bias and stochasticity on scales larger than those available before.

In addition to exploring the relation between haloes and matter we can also investigate the relation between the initial and final matter distribution. Weak-lensing measures the non-linear matter field, while for the study of the linear growth factor, one would like to know the relation between galaxies and the linear matter field instead. We explore the relation between the final and initial dark matter density field on large scales, where this relation is believed to be perfect. This case is amenable to perturbation theory analysis and as such allows one to interpret and verify the numerical simulation results.

Finally, we revisit the question of halo bias as a function of halo mass with the new simulations. This has been addressed by the previous generation of simulations using 256$^3$ particles and box sizes of order $(100–140) \, \text{h}^{-1} \, \text{Mpc}$ (Jing 1998; Sheth & Tormen 1999; Sheth, Mo & Tormen 2001). This, as noted by the authors themselves, is barely adequate for this purpose because of large shot noise at high halo masses and insufficient number of large-scale modes where linear evolution is valid. The goal of these papers was to provide expressions which fit over a range of power-law simulations and were not specifically optimized for realistic $\Lambda$CDM models. They provide expressions that fit the simulations available at the time to a reasonable accuracy, but which can be systematically wrong by as much as 10–20 per cent. To put things into a current context, the statistical error on the amplitude of galaxy clustering from SDSS using $k < 0.2 \, \text{h} \, \text{Mpc}^{-1}$ modes is 1 per cent (Tegmark et al. 2004), so a perfect bias determination, for example using the faint galaxies as described above, would allow us to reach this accuracy on the matter power spectrum. In this era 20 per cent accuracy no longer suffices and the goal of the present paper is to provide more accurate expressions for halo bias as a function of mass and cosmological model.

2 Simulations

The N-body code we use in this paper is the Hased Oct-Tree code (HOT), a parallel, tree-based code (Warren & Salmon 1993). This code was compared to a variety of other simulation codes in Frenk et al. (1999), and further validation studies will be presented elsewhere. For this paper we performed several simulations with this code. The smallest was a $96 \, \text{h}^{-1} \, \text{Mpc}$ box size, $512^3$-particle run (HOT1). This simulation has a particle mass of $5.5 \times 10^8 \, \text{M}_\odot$ and is useful for probing the halo bias at the low-mass end, below $10^{13} \, \text{h}^{-1} \, \text{M}_\odot$. It suffers from the small box size which makes the investigations on linear scales difficult and makes the shot-noise fluctuations for higher-mass haloes (with lower halo numbers) very large. Next up in size is a simulation with $288 \, \text{h}^{-1} \, \text{Mpc}$ box and 768$^3$ particles (HOT2). This is the main simulation that we use in this paper, as it has an optimal combination of box size and particle mass for our purposes. It samples the Fourier modes down to $k \sim 0.02 \, \text{h} \, \text{Mpc}^{-1}$ and has many modes at $k \sim 0.1 \, \text{h} \, \text{Mpc}^{-1}$, where the power spectrum is still close to linear. This is also the typical scale probed by the current surveys such as SDSS and 2dF. The particle mass for this simulation is $4.4 \times 10^9 \, \text{M}_\odot$ and can resolve haloes down to a few times $10^{12} \, \text{h}^{-1} \, \text{M}_\odot$, which is sufficient for typical galaxies in a flux-limited sample. Finally, for determination of the halo bias at the high-mass end we use a simulation with $1152 \, \text{h}^{-1} \, \text{Mpc}$ box size and 768$^3$ particles (HOT3). This simulation has a large enough box to sample long-wavelength modes well, but its particle mass of $2.8 \times 10^{12} \, \text{h}^{-1} \, \text{M}_\odot$ does not allow us to resolve galactic-size haloes and we limit its use to group- and cluster-size haloes.

The accuracy of the tree code was controlled using the absolute error criterion described in Salmon & Warren (1994), which ranged from $10^{-3} M_{\text{init}}/R_0^2$ per interaction at the start of the each simulation to $10^{-5} M_{\text{init}}/R_0^2$ at the end. Smoother was used, with softening lengths of 7, 20 and 95 comoving kpc for models 1, 2 and 3, respectively. The number of time-steps for model 1 was 1475, 736 for model 2, and 725 for model 3. Model 1 started at a redshift of 50, model 2 at 44, and model 3 at 27. All particle masses were identical, with the initial particle displacements imposed on a cubical lattice.

All of these simulations are normalized to $\sigma_8 = 0.9$ and have $\Omega_m = 0.3$, $\Omega_\Lambda = 0.04$ and Hubble parameter $h = 0.7$. They use realistic transfer functions from CMBFAST (Seljak & Zaldarriaga 1996), $\sigma_8 = 0.9$ corresponds to $\delta_c = 4.624 \times 10^{-2}$ normalization in CMBFAST. In the following all the results are for HOT2 whenever not explicitly specified otherwise.

For the purpose of studying bias as a function of halo mass we also ran a suite of simulations varying one parameter at a time. The box size for these simulations is $192 \, \text{h}^{-1} \, \text{Mpc}$ with 512$^3$ particles. Their force and mass resolution is the same as for HOT2. We ran the
3 STOCHASTICITY OF DARK MATTER

We begin by exploring the relation between the final dark matter density field and the initial density field, rescaled to \( z = 0 \) using the linear growth factor. We Fourier transform both fields and denote individual modes with \( \delta_i(k) \) and \( \delta_f(k) \) (we treat real and imaginary components as separate modes). Fig. 1 shows the ratio \( b(k) = \delta_i(k)/\delta_f(k) \) for \( k < 0.1 \ h^{-1} \) Mpc\(^{-1}\). This is the scale at which one often assumes linear theory to be valid. We see that there are significant fluctuations between the initial and final field, suggesting that there are large corrections to the linear evolution even for \( k < 0.1 \ h^{-1} \) Mpc\(^{-1}\).

We can define the ratio of the power spectra as
\[
\langle b^2(k) \rangle \equiv \frac{P_i(k)}{P_f(k)} = \frac{\langle \delta_i^2(k) \rangle}{\langle \delta_f^2(k) \rangle},
\]
where \( \langle \rangle \) denotes the ensemble average over different realizations of the universe. We can define relative rms fluctuations in \( b \) as
\[
\left( \frac{\sigma_b}{b} \right)^2 = \frac{\langle \left( b - \langle b \rangle \right)^2 \rangle}{\langle \delta_f^2 \rangle}.
\]

This is related to the cross-correlation coefficient \( r \), defined as
\[
r(k) = \frac{\langle \delta_i(k) \delta_f(k) \rangle}{\sqrt{\langle \delta_i^2(k) \rangle \langle \delta_f^2(k) \rangle}}.
\]

The two are related via
\[
\frac{\sigma_b}{b} = \sqrt{2(1-r)}.
\]

Fig. 2 shows \( \sigma_b/b \) as a function of wave mode \( k \), where the average has been done over a large number of wave modes so that \( r \) converges (at very low \( k \) this condition is not satisfied and \( r \) is biased high, which underestimates the rms fluctuations). We see that \( \sigma_b/b \) changes from 10 per cent at \( k \sim 0.02 \) h Mpc\(^{-1}\) to 40 per cent at 0.1 h Mpc\(^{-1}\), above which it rapidly increases and the two fields become incoherent. Fig. 2 also shows the ratio of non-linear to linear power spectrum at \( z = 0 \) for HOT3 (solid). Also shown is the power-spectrum ratio between the two fields (dashed). Results for HOT2 are similar, but show a somewhat larger suppression of non-linear versus linear power spectrum for \( k < 0.1 \) h Mpc\(^{-1}\).

The main result arising from Figs 1 and 2 is that the fluctuations between the linear and non-linear fields are large on large scales, despite the fact that the non-linear power spectrum is very close to the linear one. This is not so evident from the cross-correlation coefficient \( r \) itself, which can be close to 1 and still lead to large rms fluctuations: even for \( r = 0.995 \) the rms fluctuations between the initial and final field are 10 per cent for any given mode.

One can get some understanding of these results by using second-order perturbation theory results (see Bernardeau et al. 2002, for a recent review). To compute the power spectrum to second order one must derive the density field to third order, \( \delta = \delta_1 + \delta_2 + \delta_3 \). Second-order contributions to the power spectrum arise from both the \( \delta_2 \delta_2 \) and \( \delta_1 \delta_3 \) terms. The \( \delta_2 \delta_2 \) term is the mode–mode coupling term, while the \( \delta_1 \delta_3 \) term is the non-linear growth evolution term. These terms have different behaviour in various limits and have differing...
signs in the contribution: while $\delta_2\delta_2$ is strictly positive, $\delta_1\delta_3$ has a negative component. For $k < 0.15 \, h \, \text{Mpc}^{-1}$ the negative contribution wins and the second-order correction to the power spectrum is negative. At its peak around $k \approx 0.1 \, h \, \text{Mpc}^{-1}$ the correction is 5 per cent and slowly decreases towards $k \rightarrow 0$. While the correction is small on average, this is a result of a cancellation between positive and negative contributions. The dominant perturbative corrections come from the mode–mode couplings at wavelengths close to the wavelength of the mode itself: for $k = 0.1 \, h \, \text{Mpc}^{-1}$ the dominant contribution to the positive component is from the modes around $k \approx 0.05 \, h \, \text{Mpc}^{-1}$, which contribute around a third of the total correction or around 5 per cent of the final power spectrum (Jain & Bertschinger 1994). These are long-wavelength modes and in any finite volume there will be large statistical fluctuations in their power relative to the true power. This leads to significant fluctuations in the second-order corrections depending on the actual realization of the mode amplitudes. Thus the final amplitudes of individual modes fluctuate significantly relative to their initial values because of perturbative large-scale effects.

A similar effect is observed in the average non-linear power spectrum, which is suppressed relative to the linear spectrum. The suppression of non-linear power for $k < 0.15 \, h \, \text{Mpc}^{-1}$ relative to the linear one is dominated by long-wavelength mode–mode coupling, so to get to a percentage precision in simulations one needs very large simulation boxes. We find that the difference between HOT2 ($320 \, h^{-1} \, \text{Mpc}$) and HOT3 ($1152 \, h^{-1} \, \text{Mpc}$) is 5 per cent in power at $k = 0.1 \, h^{-1} \, \text{Mpc}$, with larger HOT3 simulation being closer to the linear power spectrum than HOT2. Thus while these mode–mode induced fluctuations are small compared to sampling variance for individual modes, they are not small when one averages over many modes and may dominate the accuracy of amplitude determination on large scales.

4 STOCHASTICITY OF HALOES

Galaxies are believed to form inside dark matter haloes, which are virialized structures of high density. They can be labelled by their virial mass. Observations suggest that about 80 per cent of the galaxies in a typical flux-limited survey form at the centres of haloes with masses ranging between $10^{12} \, h^{-1} \, M_{\odot}$ and $10^{14} \, h^{-1} \, M_{\odot}$, while the remaining 20 per cent of galaxies are non-central and occupy groups and clusters (Guzik & Seljak 2002; Seljak et al. 2004). The exact radial distribution of galaxies inside haloes and the form of the halo mass probability distribution is the subject of a lot of current observational and theoretical effort. Here we will use centres of dark matter haloes as a proxy for galaxy positions. This will not give the correct correlation properties on small scales, where correlations between central and non-central galaxies within the haloes are important, but should be valid on large scales, where haloes can be thought of as point-like. We will show the results for a range of halo masses, which can be roughly thought of as corresponding to galaxies with different luminosities since there is a tight relation between the halo mass and luminosity (McKay et al. 2001; Guzik & Seljak 2002; Seljak et al. 2004). Alternatively, the different samples can be thought of as varying the flux limit of a survey, since going to fainter limits increases the number density of galaxies and thus reduces the shot noise and the same effect is achieved by going to fainter galaxies.

Dark matter haloes are identified from the simulations using the standard friends-of-friends (FOF) algorithm with a linking length of 0.2. The resulting mass functions agree well with the fitting formulae in the literature (Sheth & Tormen 1999; Jenkins et al. 2001). We order them by mass and use subsamples separated roughly by a factor of 2 in mass. As in the previous section we can define the halo fluctuation $\delta_h(k)$ and bias $b(k) = \delta_h(k)/\delta_m(k)$, as well as the cross-correlation coefficient between the two fields (equation 3). Fig. 3 shows the relative rms fluctuations $\sigma/b$ as a function of scale for several halo masses, relative to both the initial and the final density field. We show the case with and without the subtraction of shot-noise contribution to the halo power spectrum (the dark matter power spectrum does not require shot-noise subtraction because of the large number of dark matter particles). The lines without shot-noise subtraction are always above the ones with subtraction and are the relevant ones if one is interested in the stochasticity between the haloes and dark matter. The lower lines for which the shot noise has been subtracted show the remaining stochasticity which is not due to the shot noise. Because of the shot-noise subtraction the cross-correlation coefficient can exceed 1, in which case we do not show the result.

From Fig. 3 we see that the haloes are even less well correlated to the initial density field than the dark matter is. The stochasticity begins at the level of 20 per cent at low $k$, increasing to 50 per cent at $k \approx 0.1 \, h \, \text{Mpc}^{-1}$. The shot-noise contribution to stochasticity is small for haloes with high spatial density (low-mass haloes), but increases significantly for haloes with low spatial density, as expected. There is no obvious difference in the shot-noise subtracted values, suggesting that the shot noise simply adds an additional component of stochasticity on top of that induced by non-linearities in the relation between the haloes and the initial density field.

The correlation coefficient between the haloes and the final density field is also shown in Fig. 3 (dashed lines). Compared to the halo–initial field correlations the stochasticity is similar on the largest scales, but there is better agreement between the halo and the final dark matter field on smaller scales ($k > 0.1 \, h \, \text{Mpc}^{-1}$).
cross-scale correlation coefficient \( r \) would likely be even larger on small scales if we had modelled the galaxy distributions within the haloes more realistically, since this would lead to an enhancement of correlations on small scales, similar to that seen in the dark matter. Results from \( \text{NFW} \) simulations and analytic results using halo models suggest that the cross-correlation coefficient can remain close to unity up to a fairly high \( k \sim 1 \, h \, \text{Mpc}^{-1} \) (Seljak 2000), but this may not be generic and depends on the details of how galaxies are populated within the haloes, which are quite uncertain. Observational evidence suggests that there is some stochasticity on the 1-Mpc scale, with \( r \sim 0.5 \) (Hoekstra et al. 2002). If \( r < 1 \) it would complicate the interpretation of the results based on the comparison between galaxy–galaxy correlations and galaxy–dark matter correlations, such as those from the galaxy–galaxy lensing analysis (Sheldon et al. 2004). Here we are more concerned with the correlations on large scales, \( k < 0.1 \, h \, \text{Mpc}^{-1} \), where the details of galaxy distribution within haloes are not important and where direct observations are not yet available. The results suggest that the fluctuations between haloes and the initial or final matter field are never below 10–20 per cent.

The dark matter distribution cannot be directly observed, so results shown in Fig. 3 are not directly applicable to any observational test. The closest example to a direct observation of the dark matter is through the weak-lensing effect. Here the light from distant galactic sources is being distorted by the mass distribution along the line of sight. By averaging over the image distortions we can reconstruct the 2D shear and convergence maps. These are given by the line-of-sight projection of the matter density, weighted by a radial function that is very broad. Correlations at a given angular scale receive dominant contributions from a transverse distance at half the distance to the source, but significant contributions are also coming from much smaller transverse separations produced by the mass distribution closer to the observer.

It has been suggested by Pen (2004) that if one cross-correlates the properly radially weighted galaxy field with the weak-lensing maps then one determines the bias of galaxies exactly if the two are perfectly correlated. Under these assumptions one can use the galaxy clustering information to determine the amplitude of dark matter fluctuations with higher accuracy than from the weak lensing itself, because the galaxy clustering can be done in 3D (if redshifts are measured) and so one has more independent modes to reduce the sampling variance compared to the 2D analysis. For this method to work the correlation between the projected matter density and galaxy field must be close to perfect.

To address this assumption one must correlate 2D projections of final dark matter and galaxies. While properly projected weak-lensing 2D maps have been constructed from \( \text{N-body simulations} \) (Jain, Seljak & White 2000; White & Hu 2000), we take a simplifying approach here and cross-correlate the 2D projections of the simulations along each of the three axes. The resulting rms scatter as a function of projected wavevector \( k \) is found to be significantly larger than in the 3D case, a consequence of the projection effects, which cause shorter wavelength modes to contribute to longer wavelength modes in projection. For \( 10^{12} \, h^{-1} \, M_\odot \) haloes, which corresponds roughly to \( L_* \) galaxies, we find that the rms scatter is 20 per cent at \( k \sim 0.03 \, h \, \text{Mpc}^{-1} \) and 40 per cent at \( k \sim 0.1 \, h \, \text{Mpc}^{-1} \). This is reduced by a factor of 2.5 of \( 10^{10} \, h^{-1} \, M_\odot \) galaxies are used instead. This last example is shown in Fig. 4. In reality the stochasticity is likely to be larger for the lensing case, since projections at a fixed angle (rather than at a fixed transverse separation as done here) receive contributions from nearby small-scale structures, for which the stochasticity between the galaxies and the dark matter will be much larger.

We can estimate the effect of this scatter on the amplitude determination from the weak-lensing cross-correlation analysis. If the lensing kernel peaks at \( z \sim 0.3–0.4 \) then \( k \sim 0.1 \, h \, \text{Mpc}^{-1} \) corresponds to \( l \sim 100 \). In a 200 deg\(^2\) survey such as the upcoming CFHT Legacy Survey (Van Waerbeke & Mellier 2003) we will have about five independent modes at \( l \sim 30 \) and 50 at \( l \sim 100 \). This means that for galaxies in \( 10^{12} \, h^{-1} \, M_\odot \) haloes the overall linear amplitude will have an error of 20 per cent/\( \sqrt{5} \) \( \sim 9 \) per cent at \( l \approx 30 \) and 6 per cent at \( l \sim 100 \), arising just from this effect (the power-spectrum amplitude error will be twice as large). Additional errors of comparable magnitude will arise from the lensing noise and projection effects. Such a poor determination of the growth factor as a function of redshift is unlikely to improve our current constraints on the dark energy significantly. This source of error was not included in the previous analysis (Pen 2004) and is much larger than the prognosticated errors without it. This complicates the prospects of this method for studies of dark energy through the growth factor evolution. The errors can be reduced with a larger survey area: for a survey covering 25 per cent of the full sky the errors on the power-spectrum amplitude may approach 1 per cent because more modes are being sampled and because the largest modes have the smallest amount of stochasticity. It remains to be seen whether this is ever competitive with a straight weak-lensing analysis on smaller scales.

As discussed in the introduction another method to determine the bias is to combine the clustering analysis of faint galaxies, for which we know the theoretical bias, with the luminous galaxies, for which we can measure the clustering on large scales with a small statistical error. To determine the relative bias between the populations we can simply compare the smoothed maps. In the absence of stochasticity between the two galaxy populations one could determine the amplitude of mass fluctuations directly. Suppose that we want to determine the clustering amplitude of faint galaxies, which are in low-mass haloes (around \( 10^{11} \, h^{-1} \, M_\odot \) for galaxies 2–3 mag below \( L_* \)) and \( L_* \) galaxies, which typically occupy \( 10^{12} \, h^{-1} \, M_\odot \). Fig. 5
shows the relative rms fluctuations in ratios of Fourier mode amplitudes between haloes of mass $10^{11} h^{-1} M_\odot$ and $10^{12} h^{-1} M_\odot$. We see again from Fig. 5 that the scatter is large. Both shot noise and stochasticity due to non-linearities limit this method. The rms fluctuations between $10^{11} h^{-1} M_\odot$ and $10^{12} h^{-1} M_\odot$ haloes are 8 per cent at $k \sim 0.1 h$ Mpc$^{-1}$ and 23 per cent at $k \sim 0.2 h$ Mpc$^{-1}$. This is somewhat smaller than between the haloes and the matter. Moreover, galaxies in redshift surveys provide 3D information, so there are more large-scale modes to reduce the scatter. Nevertheless, any attempt to determine the linear bias using the cross-correlations must include this source of stochasticity in the analysis.

5 HALO BIAS AS A FUNCTION OF MASS

One of the important questions that can be addressed with these simulations is the relation between halo and dark matter power spectrum as a function of halo mass. This relation has been theoretically predicted from the spherical collapse model (Cole & Kaiser 1989; Mo & White 1996) and from the ellipsoidal collapse model (Sheth et al. 2001), which suggest that the halo bias is related to a derivative of the halo mass function. The relation has also been extracted from the numerical simulations, with good quantitative agreement between the simulations in the overlap mass range, but the larger simulation has smaller statistical errors. The smallest simulation (HOT1) has very few modes in the linear regime and the fluctuations in the ratio caused by perturbative effects beyond linear theory are large, so the bias determination from this simulation is somewhat less reliable. On the other hand, all of the low-mass haloes in this simulation have almost the same bias and at the upper end of the mass range there is good agreement in bias with haloes in HOT3 haloes.

The simulations used in this paper are a significant improvement over the previous generation. They contain 8–64 times more particles and cover a wide range of masses. We use HOT1 simulation for haloes in the mass range $(5 \times 10^{10} - 3 \times 10^{11}) M_\odot$, HOT2 for haloes in the mass range $(3 \times 10^{11} - 10^{12}) M_\odot$ and HOT3 for haloes in the mass range $(10^{12} - 10^{13}) M_\odot$ (the latter two are also checked with $768 h^{-1}$ Mpc box 1024$^3$-particle simulation).

Fig. 6 shows the ratio of the (shot-noise corrected) halo power spectrum to the linear mass power spectrum as a function of wavevector $k$. One can see that the assumption of constant bias is reasonable for $k < 0.1 h$ Mpc$^{-1}$ and even beyond, so a linear bias can be defined as an appropriate average over these modes. The exception are the most massive haloes in HOT3 with $b > 1.5$, for which the power spectrum is already suppressed at $k \sim 0.1 h$ Mpc$^{-1}$ due to the fact that the FOF haloes do not overlap and so cannot be closer than two times the virial radius. Here we use all of the modes with $k < 0.1 h$ Mpc$^{-1}$, except for the smallest 96 $h^{-1}$ Mpc simulation where we use $k < 0.15 h$ Mpc$^{-1}$. We note that there is a good agreement between the simulations in the overlap mass range, but the larger simulation has smaller statistical errors. The smallest simulation (HOT1) has very few modes in the linear regime and the fluctuations in the ratio caused by perturbative effects beyond linear theory are large, so the bias determination from this simulation is somewhat less reliable. On the other hand, all of the low-mass haloes in this simulation have almost the same bias and at the upper end of the mass range there is good agreement in bias with haloes of the same mass from HOT2.

For simplification with theoretical comparisons we will scale all the masses relative to the non-linear mass $M_{nl}$, defined as the mass within a sphere for which the rms fluctuation amplitude of the linear field is 1.68. While the theoretical predictions for the bias depend on the cosmological model, most of that dependence is accounted for if the mass is expressed in terms of the non-linear mass. For HOT1–3 simulations with $\sigma_8 = 0.9$ and $\Omega_m = 0.3$ at $z = 0$ the

Bias and stochasticity of haloes and dark matter

Figure 7. Bias as a function of mass in units of the non-linear mass. Points are from 96 $h^{-1}$ Mpc $512^3$ (HOT1, green), 144 $h^{-1}$ Mpc $512^3$ (cyan), 192 $h^{-1}$ Mpc $512^3$ (red), 288 $h^{-1}$ Mpc $768^3$ (HOT2, blue), 1152 $h^{-1}$ Mpc $768^3$ (HOT3, magenta) and 768 $h^{-1}$ Mpc $1024^3$ (black) simulations. Note that in several cases the points from two simulations overlap exactly. Upper (dashed blue) line is the theoretical prediction from Sheth and Tormen (1999). Lower (solid black) line is the expression from equation (5).

Figure 8. Bias as a function of mass in units of the non-linear mass for several cosmological models. We varied one parameter at a time relative to the fiducial concordance model, roughly covering the range of interest. This figure shows that the bias predictions depend predominantly on the non-linear mass, while other cosmological parameters play only a minor, but not entirely negligible, role.

non-linear mass defined with 1.68 overdensity is $8.73 \times 10^{12} h^{-1} M_\odot$. Fig. 7 shows the bias determinations as a function of halo mass from the simulations used in this paper. The dashed line is the theoretical prediction from Sheth & Tormen (1999) (the fitting formula given in Jing (1998) is very similar; while these fitting formulae are not very accurate we find good agreement between the simulation results in these papers and our simulations). We see that these theoretical predictions overestimate the bias below $M_{nl}$ and are a good fit above $M_{nl}$. The largest discrepancy is below $M_{nl}$, where the relative error can be up to 20 per cent. The various simulations are in reasonable agreement among themselves and the scatter between the points at the same mass is mostly due to the shot noise and small volume over which one is averaging. There may be some systematic error due to the fact that the non-linear mass computed from the theoretical power spectrum can differ from the value obtained if one uses the actual realization. We find this can lead up to a 10 per cent effect on non-linear mass and would cause a horizontal shift by this amount. This is of almost no consequence for masses below $M_{nl}$, where bias is only weakly dependent on the mass, but may lead to a larger error at the high-mass end.

We find that the unbiased galaxies with $b = 1$ are at $M = 1.5M_{nl}$ and the bias is rapidly changing above $0.1M_{nl}$, while below this it is essentially constant with the value around 0.68. In all simulations we see bias increasing at the lowest masses (Fig. 8), which is a numerical artefact. For example, such an increase is seen in HOT2 at the low-mass end and is not confirmed in HOT1, where the mass resolution improves by a factor of 8 (Fig. 8). Moreover, this increase at the low-mass end changes into a decrease if we remove unbound particles from the haloes. To be safe we only present results where the difference between the two cases is less than 0.01. Note that in HOT1 we find $b \sim 0.65$ at the low-mass end. Even in the region of overlap with HOT2 the bias in HOT1 is systematically lower by 0.03. This is likely to be due to the sampling variance in HOT1, as can be seen from Fig. 6, which shows considerable fluctuations as a function of wavevector for this simulation. With a 144 $h^{-1}$ Mpc box $512^3$-particle simulation we again find that $b \sim 0.65$–0.68 at the low-mass end and that there is indeed significant scatter due to the small box size (Fig. 7). For this reason the empirical fit given below goes above HOT1 at the low-mass end. It is not entirely clear that this is the correct procedure, as HOT2 could have been already affected by the resolution, but the fact that both unbound and bound haloes give the same result argues against this.

While there is some uncertainty in the bias value at the low-mass end, all simulations agree very well around the non-linear mass. In addition to HOT2 simulation we also have another 512$^3$ simulation with 192 $h^{-1}$ Mpc box simulation and a 768 $h^{-1}$ Mpc box with 1024$^3$-particle simulation that both sample well this regime. At the high-mass end uncertainty increases again because of the small number of high-mass haloes. In addition to HOT3 and the 768 $h^{-1}$ Mpc box with 1024$^3$-particle simulation we use another 768 $h^{-1}$ Mpc simulation with 384$^3$ particles.

The solid curve in Fig. 7 is an empirical expression that fits all simulations. Over the range between $10^{-3} < M/M_{nl} < 10^2$ it is given by

$$b_0(x = M/M_{nl}) = 0.53 + 0.39x^{0.45} + \frac{0.13}{40x + 1} + 5 \times 10^{-4}x^{1.5}.$$  \hspace{1cm} (5)

This expression should be accurate to about 3 per cent or better for this model, as suggested from the scatter in Fig. 7.

In Fig. 8 we show the bias as a function of mass for several simulations for which we varied one parameter at a time, roughly spanning the range of interest from cosmological constraints today. We see that there is very little difference in the theoretical
predictions for halo bias as a function of $M/M_{\text{crit}}$, suggesting that instead of deriving full expressions, which depend on all cosmological parameters, one can simply use a single relation with mass in units of non-linear mass, as in equation (5). The deviations from this relation are qualitatively consistent with the predictions given by Sheth & Tormen (1999). We can generalize the results from equation (5) by linearizing the bias relation in terms of cosmological parameters,

$$b(x) = b_0(x) + \log_{10}(x)(0.4(\Omega_m - 0.3 + n_* - 1)
+ 0.3(\sigma_8 - 0.9 + h - 0.7) + 0.8\alpha_8).$$

(6)

This correction should be reasonable for $1 > x > 0.1$, while below that the correction appears to saturate at $-0.4(\Omega_m - 0.3 + n_* - 1) - 0.3(\sigma_8 - 0.9 + h - 0.7) - 0.8\alpha_8$. For massive haloes with $M > M_{\text{crit}}(x > 1)$ the differences among models in the bias predictions from Sheth & Tormen (1999) become larger, but this is difficult to observe in these simulations, where the number of such haloes is small and the bias measurements have large shot noise. In this regime the analytic predictions may be more accurate than equation (5): we do not see much evidence against the analytic expressions from our comparisons (Fig. 7) and analytic expressions can be more easily generalized to more general cosmological models. Note, however, that the simulations used in this paper improve upon the previous generation of simulations over this regime as well.

6 CONCLUSIONS

In this paper we have addressed the relations between the matter density field, haloes and initial density field, focusing on large scales where these are often assumed to be proportional to each other. We focus on two issues. First, what is the scatter between these fields around the average relation? This is expressed here in terms of relative scatter between the mode amplitudes, which is related to the stochasticity parameter $r$, defined as the cross-correlation coefficient between the two fields. While the two are related we emphasize that even small deviations of $r$ from unity may lead to large relative fluctuations between the two fields. These are of interest whenever one is trying to relate the fields to one another to determine their relative amplitudes. One example is the bias determination using the cross-correlation between the weak-lensing signal (tracing the matter density) and the galaxies. Another example is the relative bias determination between two different galaxy populations, which we propose here as an alternative method to determine the galaxy bias, because galaxies in low-mass haloes have a bias of $b \sim 0.7$ independent of their mass. In all cases we find the scatter between the fields in individual modes is significant and one cannot assume the fields are simply proportional to one another. This scatter, coupled with a small number of modes on large scales, makes it difficult to determine the bias (or relative bias) accurately and needs to be included in the predictions of how accurately one can determine the matter power spectrum with these methods.

The second goal of this paper was to revisit the halo bias as a function of halo mass. This relation is a fundamental ingredient of any halo model (see Cooray & Sheth 2002, for a recent review) and plays an important role if one is trying to model galaxy clustering by connecting it to the underlying haloes. The previous generation of simulations (Jing 1998; Sheth & Tormen 1999) had a limited dynamical range and the predictions were not tuned specifically for \(\Lambda\)CDM models. As a result the existing expressions overestimate the bias by as much as 20 per cent in the range below the non-linear mass, which is likely to be the mass range for haloes that host most of the galaxies. We propose a new expression that fits the simulations better. We argue that this expression should be fairly accurate for other cosmological models of interest as well, as long as the mass is expressed in units of non-linear mass. We give corrections for small deviations from this model. The overall accuracy on the bias–halo mass relation is at the level of 0.03 or better (for $b < 1$), which should help with the bias determination from the current generation of observations.

ACKNOWLEDGMENTS

We thank P. McDonald, N. Padmanabhan and L. Teodoro for help and useful comments, S. Habib for encouraging this collaboration and Zheng Zheng and David Weinberg for comments on the manuscript. US is supported by the Packard Foundation, Sloan Foundation, NASA NAG5-1993 and NSF CAREER-0132953. Simulations by MSW were performed on the Space Simulator Beowulf Cluster and ASCI Q at Los Alamos. Work by MSW was performed under the auspices of the US Department of Energy, and supported by its contract #W-7405-ENG-36 to Los Alamos National Laboratory.

REFERENCES

Sheldon E. S. et al., 2004, AJ, 127, 2544

This paper has been typeset from a TeX file prepared by the author.