Mon. Not. R. Astron. Soc. **366,** 529–536 (2006) doi:10.1111/j.1365-2966.2005.09880.x

Pair correlations and merger bias

Steven R. Furlanetto★ and Marc Kamionkowski★

California Institute of Technology, Mail Code 130-33, Pasadena, CA 91125, USA

Accepted 2005 November 15. Received 2005 November 11; in original form 2005 July 27

ABSTRACT

We analytically study the possibility that mergers of haloes are more highly clustered than the general population of haloes of comparable masses. We begin by investigating predictions for merger bias within the extended Press–Schechter formalism and discuss the limitations and ambiguities of this approach. We then postulate that mergers occur whenever two objects form within a (small) fixed distance of each other. We therefore study the clustering of pairs of points for a highly biased population in the linear regime, for the overall mass distribution in the quasi-linear regime and (using the halo model of clustering) in the non-linear regime. Biasing, quasi-linear evolution and non-linear clustering all lead to non-zero reduced (or connected) three- and four-point correlation functions. These higher order correlation functions can in many cases enhance the clustering of close pairs of points relative to the clustering of individual points. If close pairs are likely to merge, then the clustering of mergers may be enhanced. We discuss implications for the observed clustering of luminous $z = 3$ galaxies and for correlations of active galactic nuclei and galaxy clusters.

Key words: galaxies: formation – galaxies: interactions – large-scale structure of Universe.

1 INTRODUCTION

Galaxy clustering can be a useful tool to study the origin of large-scale structure and to delineate the formation mechanisms of various types of galaxies. For example, it is now well appreciated that objects forming from rare high-density peaks in the primordial density distribution, such as bright galaxies at high redshifts or galaxy clusters today, should be 'biased' (i.e. more highly clustered) relative to the more common lower mass objects that trace the total-mass distribution more closely (Kaiser 1984).

A currently unanswered question is whether the growth history of haloes can affect their clustering properties. Cosmological simulations give confusing results. Kolatt et al. (1999) argued that merger-driven starbursts at $z \sim 3$ occur in small haloes that lie near larger ones: thus they are more highly clustered than typical objects of the same mass (see also Wechsler et al. 2001). The simulations of Gottlöber et al. (2002) showed different clustering at $z = 0$ between objects that had experienced a major merger and those that had not. Kauffmann & Haehnelt (2002) also found a weak enhancement in the cross-correlation between objects undergoing major mergers and the general population, but only at small scales. On the other hand, Percival et al. (2003) found no evidence for excess merger bias at $z = 0$, where recently merged objects were identified as haloes in which at least 50 per cent of constituent particles were not in a progenitor of at least equal mass at a fixed earlier redshift. Scannapieco & Thacker (2003) agreed at $z = 3$, but if they modified the criterion to include all haloes that grew by 20 per cent or more (implicitly including smooth infall), the rapidly growing sample had a substantial excess bias, making their clustering comparable to that of haloes with three times more mass. Most recently, Gao, Springel & White (2005) examined a high-dynamic-range *N*-body simulation at $z = 0$. They found the clustering of low-mass recently merged objects to be suppressed relative to the average. For example, in their lowest mass bin (with a mass ≈2 per cent of the characteristic halo mass), the 20 per cent youngest and oldest haloes are under- and overbiased by ∼40 per cent, respectively. On the other hand, in agreement with Percival et al. (2003), they found that the clustering of more massive objects is nearly independent of their age. The verdict is clearly not yet in: how can we reconcile these disparate results?

The question is not just academic. Clustering is often used to infer information about the host halo mass of particular galaxy populations (e.g. Mo & Fukugita 1996; Adelberger et al. 1998; Giavalisco et al. 1998). The possibility that clustering depends on the merger history – which obviously also strongly affects observables such as the star formation history – would call such inferences into question. One example is the discrepancy between the masses ($\sim 10^{12}\,M_\odot$) of Lyman-break galaxies (LBGs) inferred from their clustering (Coles et al. 1998; Giavalisco & Dickinson 2001; Porciani & Giavalisco 2002; Adelberger et al. 2005) and the dynamical masses ($\sim 10^{11}\,M_\odot$) inferred from the broadening of nebular emission lines and kinematics (Pettini et al. 2001; Erb et al. 2003). This claimed discrepancy may simply be the difference between the mass in the central regions and the total mass (Erb et al. 2003; Cooray 2005), but Wechsler et al. (2001) and Scannapieco & Thacker (2003) have proposed that it may also point to 'merger bias' if LBGs are galaxies that have recently undergone mergers. The problem is even

★E-mail: sfurlane@tapir.caltech.edu (SRF); kamion@tapir.caltech.edu (MK)

more extreme for submillimetre-selected galaxies at $z > 2$: their dynamics imply total masses $M \lesssim 10^{12}\,\mathrm{M_\odot}$ while clustering implies $M \gtrsim 10^{13}\,\mathrm{M_\odot}$ (Blain et al. 2004).

More generally, to what extent does clustering depend on factors other than the halo mass? Will selection techniques that trace recent episodes of star formation (such as Lyman-break or Lyα line selection) yield more highly clustered objects than techniques sensitive to the total stellar mass (such as infrared observations), even if the typical halo masses in the surveys are identical? Quasars and other active galactic nuclei (AGNs) may also be triggered by galaxy mergers. Their clustering has been used to infer the properties of the host galaxy (La Franca, Andreani & Cristiani 1998; Adelberger & Steidel 2005b) and of the quasar (especially its lifetime; Haiman & Hui 2001; Martini & Weinberg 2001; Adelberger & Steidel 2005a). How will the bias of mergers (if it exists) affect such estimates? Will recently merged galaxy clusters trace the underlying mass distribution differently from relaxed clusters? All of these questions have implications for our understanding of both galaxy formation and the large-scale structure of the Universe.

In this paper, we take an analytic approach that complements the numerical studies and may aid in their interpretation. We begin in Section 2 by considering the question of 'merger bias' within the context of the widely used linear-bias model. We show that existing techniques cannot adequately answer this question, so we then go on to consider other approaches. To be more precise, in Sections 3–7, we derive analytic results for the clustering of close pairs of galaxies in several clustering models. We consider the clustering of close pairs when galaxies Poisson sample (i) the overall mass in a Gaussian random field; (ii) the high-density peaks in a primordial Gaussian random field; (iii) the overall mass in the quasi-linear regime and (iv) the overall mass in the non-linear regime described by the halo clustering model. We find that the clustering of close pairs of galaxies can be enhanced, sometimes significantly, relative to the galaxies in many of these cases. We speculate that if close pairs are likely to merge, then a pair bias will imply a merger bias, although we do not make this statement precise. If a pair bias does in fact lead to a merger bias, then our results are consistent with a solution to the LBG puzzle. We also briefly discuss other observable implications of our results.

2 A FIRST LOOK AT MERGER BIAS

We will first attempt to compute the bias of merging objects via their number densities and the 'peak-background split' approach to bias (Efstathiou et al. 1988; Cole & Kaiser 1989; Mo & White 1996). We define the number density $n_m\,\mathrm{d}m_1\,\mathrm{d}m_2$ of mergers between haloes in the mass range $m_1 \to m_1 + \mathrm{d}m_1$ and those in the mass range $m \to m_2 + \mathrm{d}m_2$ via

$$n_m(m_1, m_2, z) = n(m_1, z)\, n(m_2, z)\, Q(m_1, m_2, z)\, \Delta t, \tag{1}$$

where $n(m, z)\,\mathrm{d}m$ is comoving number density, at redshift z, of haloes with masses $m \to m + \mathrm{d}m$ and $Q(m_1, m_2, z)$ is the merger kernel with units of volume per unit time. We take Δt to be some finite time interval within which the mergers of interest take place; note that we assume it to be sufficiently small that the underlying halo populations do not evolve significantly.

To compute the bias, we simply need to know how each of these terms varies (to linear order) with the mean density δ in some large patch. For example, the Press & Schechter (1974) mass function is

$$n(m, z) = \sqrt{\frac{2}{\pi}}\, \frac{\bar{\rho}}{m^2}\, \frac{\delta_\mathrm{c}(z)}{\sigma} \left| \frac{\mathrm{d}\ln\sigma}{\mathrm{d}\ln m} \right| \exp\left[-\frac{\delta_\mathrm{c}^2(z)}{2\sigma^2} \right], \tag{2}$$

where δ_c is the fractional-overdensity threshold for spherical collapse, $\bar{\rho}$ is the mean background density and σ^2 is the fractional-density variance smoothed on scale m. Note that we follow the convention in which σ is independent of redshift, while $\delta_\mathrm{c}(z)$ is the (linear-extrapolated) density threshold at redshift z. This distribution can be derived in terms of a diffusion problem in (σ^2, δ) space with an absorbing barrier at $\delta = \delta_\mathrm{c}$ (Bond et al. 1991). Such an approach makes it obvious that the abundance of haloes in a region of (linear-extrapolated) overdensity δ and mass M (corresponding to σ_M) will take the same form, but with a shift in the origin (Lacey & Cole 1993):

$$\begin{aligned} n(m, z|\delta, M) = {} & \sqrt{\frac{2}{\pi}}\, \frac{\bar{\rho}}{m^2}\, \frac{\sigma^2[\delta_\mathrm{c}(z) - \delta]}{\left(\sigma^2 - \sigma_M^2\right)^{3/2}} \left| \frac{\mathrm{d}\ln\sigma}{\mathrm{d}\ln m} \right| \\ & \times \exp\left\{ -\frac{[\delta_\mathrm{c}(z) - \delta]^2}{2\left(\sigma^2 - \sigma_M^2\right)} \right\}. \end{aligned} \tag{3}$$

To find the linear bias, Mo & White (1996) first take the large-scale limit $M \to \infty$ (or $\sigma_M \to 0$). The overdensity of haloes in a region of physical volume V is

$$\delta_\mathrm{h} = \frac{n(m, z \mid \delta)\, V\, (1 + \delta_z)}{n(m, z) V} - 1, \tag{4}$$

where δ_z is the true overdensity at redshift z (without linear extrapolation) and the $(1 + \delta_z)$ factor in the numerator accounts for the fact that an overdense region is larger in Lagrangian space than in physical space. Expanding equation (2) to linear order, we find

$$\begin{aligned} \delta_\mathrm{h} & \approx \delta_z \left[1 + \frac{\nu^2 - 1}{\delta_\mathrm{c}(z = 0)} \right] + \mathcal{O}(\delta_z^2) \\ & \equiv b_\mathrm{h}(m, z)\, \delta_z + \mathcal{O}\left(\delta_z^2\right), \end{aligned} \tag{5}$$

where we have let $\nu = \delta_\mathrm{c}(z)/\sigma$. This defines the usual bias $b_\mathrm{h}(m, z)$ for haloes of mass m at redshift z.

2.1 The extended Press–Schechter merger kernel

To compute the merger bias, we need to perform a similar expansion on the kernel Q. The usual model for this quantity comes from the extended Press–Schechter merger rates of Lacey & Cole (1993). Unfortunately, as we will see explicitly below, this formalism is inherently unable to address our problem: the large-scale bias of mergers disappears from the calculation. Letting $S \equiv \sigma^2$, Lacey & Cole (1993) define $f(S_1, \delta_{c1} \mid S_T, \delta_{cT})$ to be the fraction of excursion-set trajectories that first cross $\delta_{c1} > \delta_{cT}$ at $S_1 > S_T$, given that they first cross δ_{cT} at S_T (here the subscript T refers to the total mass). This is exactly equivalent to $n(m, z|\delta, M)$ in equation (3) with the identifications $(S_1 \leftrightarrow m)$, $(S_T \leftrightarrow M)$, $\delta_{c1} = \delta_\mathrm{c}(z)$, and $\delta_{cT} = \delta$; the only difference is that here we assume M to be in a collapsed halo at a later redshift. To obtain the merger rate, we will need $f(S_T, \delta_{cT}|S_1, \delta_{c1})$ instead: given a halo at some early time, this function describes the distribution of objects to which that halo can belong at some later time. By Bayes' theorem, it is simply

$$f(S_T, \delta_{cT} \mid S_1, \delta_{c1})\, \mathrm{d}S_T = f(S_1, \delta_{c1} \mid S_T, \delta_{cT})\, \frac{f(S_T, \delta_{cT})}{f(S_1, \delta_{c1})}\, \mathrm{d}S_T, \tag{6}$$

where $f(S, \delta_\mathrm{c})$ is the unconditional first-crossing distribution (i.e. the normal Press–Schechter halo mass function). The extended Press–Schechter formalism defines $\mathrm{d}^2 p/\mathrm{d}m\mathrm{d}t$, the probability that a halo of mass m_1 will merge with an object of mass $m_2 \equiv m_T - m_1$ within an infinitesimal time interval $\mathrm{d}t$, from the limit of this distribution as $\delta_{cT} \rightarrow \delta_{c1}$. In other words, it is the probability that the object will join a larger halo in the time interval of interest. The total merger rate $n_m(m_1, m)$ is then this limit (transformed to mass and time units) multiplied by $n(m_1)$.

For our problem, we need to know the dependence of each of these quantities on the large-scale density δ_b (defined over some mass with $S_b \ll S_1, S_T$). The unconditional distributions are easy: $f(S, \delta_c) \rightarrow f(S, \delta_c \mid S_b, \delta_b)$, just like the conditional mass function. We are thus left with the progenitor distribution $f(S_1, \delta_{c1} \mid S_T, \delta_{cT})$ within the large-scale region. Recall, however, that this distribution follows from a diffusion problem with origin (S_T, δ_{cT}). It must therefore be independent of the behaviour on scales $S_b < S_T$; we only need to know that it passes δ_{cT} for the first time at S_T to compute the progenitor distribution. This step is obviously where the extended Press–Schechter formalism fails: it is completely unable to incorporate the large scale environment of merger events, so it cannot make predictions about their bias. To see this explicitly, we calculate how merger densities vary with δ_b:

$$n_m(m_1, m \mid \delta_b) \propto n(m_1 \mid \delta_b)\, \frac{\mathrm{d}^2 p(\delta_b)}{\mathrm{d}m\, \mathrm{d}t}\, \Delta t \tag{7}$$

$$\propto f(S_1, \delta_{c1} \mid \delta_b) \times \frac{f(S_T, \delta_{cT} \mid \delta_b)}{f(S_1, \delta_{c1} \mid \delta_b)} \tag{8}$$

$$\propto f(S_T, \delta_{cT} \mid \delta_b). \tag{9}$$

Thus, according to the extended Press–Schechter model, n_m varies with density in precisely the same way as the number density of haloes with the same final mass m_T. Clearly, there is *no* merger bias in this picture, but only because the formalism is unable to address the relevant question.

Thus, the conclusion of this model is not one that we can trust. In addition to this difficulty, there is the deeper one pointed out by Benson, Kamionkowski & Hassani (2005), who showed that the extended Press–Schechter merger rates are mathematically self-inconsistent (calling into question the association of trajectory jumps with mergers). While it has proven useful in a variety of contexts for galaxy formation, the extended Press–Schechter formalism is manifestly not appropriate for investigating merger bias.

2.2 A density-independent merger kernel

Unfortunately, at this time, there are no fully developed alternatives to the extended Press–Schechter formalism [but see Benson et al. (2005) for first steps in this direction]. We therefore obviously cannot compute the variation of Q with the large-scale density. Instead, we will consider the simplest possible model. We will assume that the merger kernel Q is independent of environment in the Lagrangian space to which the Press–Schechter formalism is native: that is, the merger rate varies with the local density only through the Lagrangian number density of haloes. This would be appropriate if, for example, all Gaussian peaks within a fixed comoving distance merged with each other, and if we neglect extra correlations between neighbouring haloes. In other words, we treat each of the two haloes independently of the other; clearly, this is not completely correct, because the large-scale bias does not describe the small-scale correlations between haloes (e.g. Scannapieco & Barkana 2002). We emphasize, then, that our model is not meant to be quantitatively accurate but only to illuminate the dependence of the merger bias on the halo abundances. In this case, we define the overdensity of mergers via

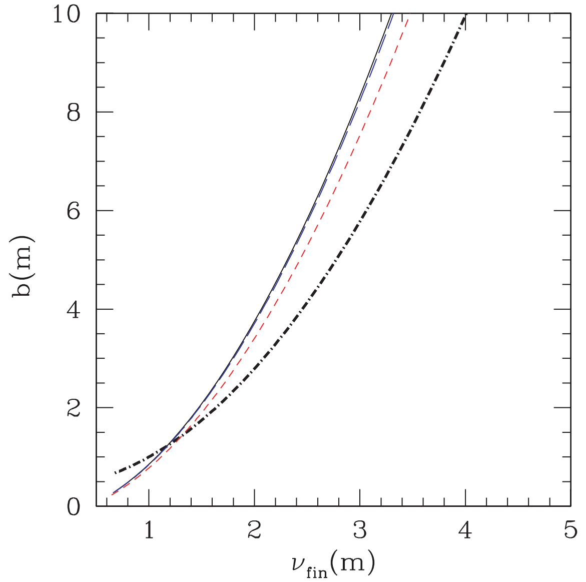

Figure 1. Merger bias at $z = 3$. The dot–dashed line shows the normal halo bias $b_{\rm h}$ for the final merger product. The thin solid, long-dashed and short-dashed curves take $m_2/m_1 = 1, 0.5$ and 0.1, respectively.

$$\delta_{\rm m} \equiv \frac{N_{\rm m}(m_1, m_2, z \mid \delta)}{n_m(m_1, m_2, z)V} - 1, \tag{10}$$

where $N_{\rm m}$ is the number of mergers in this volume. Clearly $N_{\rm m} \propto n(m_1 \mid \delta)\, n(m_2 \mid \delta)\, V\, (1 + \delta_z)$. Expanding to linear order, we find a merger bias

$$b_{\rm m} = 1 + \frac{\nu_1^2 + \nu_2^2 - 2}{\delta_{\rm c}(z = 0)}, \tag{11}$$

where $\nu_1 \equiv \nu(m_1)$, etc.

For a given final mass ν, we can then compute the bias of mergers as a function of the mass ratio. We show some results at $z = 3$ in Fig. 1 as a function of $\nu_{\rm fin} \equiv \nu(m_1 + m_2)$. Interestingly, in this model, $b_{\rm m} > b_{\rm h}$ for $\nu \gg 1$: mergers between massive objects tend to occur in denser regions than an average halo of the final mass (or in other words, younger systems are more biased than older systems). The behaviour reverses at small masses: younger systems are less biased than average. Qualitatively, a dark matter particle in a halo with $\nu \ll 1$ must be in a low-density environment; small-mass objects that have just formed will typically be in lower density environments than an average halo of this type.

Fig. 2 shows the ratio between the merger and halo bias at both $z = 3$ and 0. Note that it appears to asymptote to a constant at large ν. This is simply $b_{\rm m}/b_{\rm h} \rightarrow (\nu_1^2 + \nu_2^2)/\nu_{\rm fin}^2$; the excess bias will thus disappear when one progenitor contains nearly all of the final mass. Also, $b_{\rm m}$ can become negative for sufficiently small mass mergers: such events preferentially occur in *underdense* environments. Note also that in this model the merger bias at fixed $\nu_{\rm fin}$ depends on redshift, even though the halo bias does not; this is because (for a fixed mass ratio) the ratio ν_1/ν_2 does depend on redshift through the scale dependence of the effective slope of the cold dark matter (CDM) power spectrum.

Of course, it is not obvious that taking Q to be constant in Lagrangian space is the most reasonable assumption. We could instead have taken it to be independent of environment in physical (Eulerian) coordinates. Then, the appropriate bias would be

$$b'_{\rm m} = b_{\rm h}(m_1) + b_{\rm h}(m_2) = b_{\rm m} + 1, \tag{12}$$

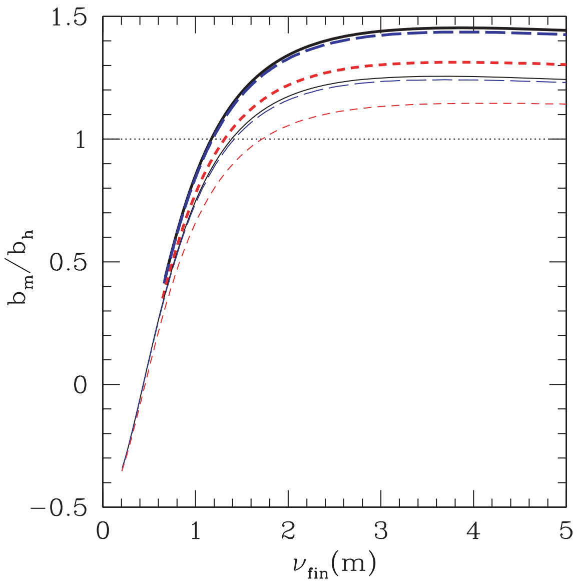

Figure 2. Ratio between the merger bias b_m and the halo bias b_h (of the final product). The solid, long-dashed and short-dashed curves take $m_2/m_1 = 1, 0.5$ and 0.1, respectively. The upper thick and lower thin sets of curves take $z = 3$ and 0, respectively.

because in this case $Q \propto (1 + \delta_z)$ when expressed in the Lagrangian space. This would be appropriate if, for example, mergers occurred only through random collisions in physical space. In this approximation, mergers are even more biased for large ν and less antibiased for small ν. It is not clear which of these assumptions is more physically plausible, but interestingly they both predict positive bias ($b_m/b_h > 1$) for mergers of massive haloes and antibias ($b_m/b_h < 1$) for mergers of sufficiently small haloes.

Comparison of the simulation results illuminates some of the properties of Q appropriate to halo growth. Gao et al. (2005) found that, for small haloes at $z = 0$, younger objects are less biased than average. This fits, at least qualitatively, with our $Q =$ constant results, which predict $b_m < b_h$ for $\nu \lesssim 1$. However, they also found no evidence for age-dependent clustering in massive objects (see also Percival et al. 2003). This is in conflict with the $Q =$ constant results, which predict a 10–20 per cent enhancement to the merger bias for large ν. Taken at face value, this implies that the merger rate of massive objects must be *suppressed* in dense regions. On the other hand, Scannapieco & Thacker (2003) claimed a positive merger bias for massive haloes in simulations at $z = 3$. The $Q =$ constant model provides an important clue that may explain this apparent redshift evolution: it does indeed predict a larger merger bias at early times. The reason is that the merger bias depends on $\nu(m_1) + \nu(m_2)$ and not simply on $\nu(m_1 + m_2)$. The characteristic scale of the mass function grows with time; because the CDM power spectrum is not a simple power law, the relation between these three quantities changes with time. Thus, although the halo bias at a fixed ν is independent of redshift, the bias of major mergers need not be.

3 CLUSTERING OF PAIRS

The last section showed that, until we have a self-consistent merger kernel Q that correctly incorporates the density dependence of the merger rates, we cannot properly predict the linear merger bias within the Press–Schechter model. It is therefore worth considering other approaches to merger bias to see what light they can shed. In this and the following sections, we will examine a picture in which mergers simply correspond to closely spaced objects. Intuitively, such pairs may merge because of (for example) non-linear gravitational collapse that brings objects closer together. We will consider how close pairs are biased relative to the objects themselves and show that, in general, the pair bias differs from the halo bias.

Consider a population of galaxies with mean spatial density n. Then the differential probability to find a galaxy in an infinitesimal volume element $\mathrm{d}V$ is $\mathrm{d}P = n\,\mathrm{d}V$. The differential probability to find one galaxy in $\mathrm{d}V_1$ centred on a position $\boldsymbol{r}_1$ and another in $\mathrm{d}V_2$ centred on $\boldsymbol{r}_2$ is $\mathrm{d}P = n^2\,\mathrm{d}V_1\,\mathrm{d}V_2\,[1 + \xi(|\boldsymbol{r}_1 - \boldsymbol{r}_2|)]$, where $\xi(r)$ is the galaxy–galaxy autocorrelation function. The correlation function is the excess probability, over random, to find two galaxies in differential volume elements separated by a distance r.

There can never be more than one galaxy in an infinitesimal volume element $\mathrm{d}V$. However, we will soon deal with close pairs of galaxies. We will thus want to know the probability to find *two* galaxies in one small, but finite, volume element δV. To be precise, we take this volume element to be a sphere of radius r_p; then $\delta V = (4\pi/3)r_p^3$. The desired probability is then

$$\delta P = n^2 \int_{\delta V} \mathrm{d}^3 r_1 \int_{\delta V} \mathrm{d}^3 r_2\, [1 + \xi(|\boldsymbol{r}_1 - \boldsymbol{r}_2|)]$$
$$\equiv n^2 (\delta V)^2 \left(1 + \left\langle \delta_p^2 \right\rangle\right), \tag{13}$$

where

$$\left\langle \delta_p^2 \right\rangle \equiv (\delta V)^{-2} \left\langle \left[\int_{\delta V} \mathrm{d}^3 r\, \delta(\boldsymbol{r} - \boldsymbol{x}) \right]^2 \right\rangle$$
$$= (\delta V)^{-2} \int_{\delta V} \mathrm{d}^3 r_1 \int_{\delta V} \mathrm{d}^3 r_2\, \langle \delta(\boldsymbol{r}_1 - \boldsymbol{x}) \delta(\boldsymbol{r}_2 - \boldsymbol{x}) \rangle$$
$$= (\delta V)^{-2} \int_{\delta V} \mathrm{d}^3 r_1 \int_{\delta V} \mathrm{d}^3 r_2 \xi(|\boldsymbol{r}_1 - \boldsymbol{r}_2|) \tag{14}$$

is the variance of the density perturbation smoothed over a spherical top hat of radius r_p. If the correlation function can be approximated by a power law, $\xi(r) \propto r^{-\alpha}$, for $r < r_p$, then

$$\left\langle \delta_p^2 \right\rangle = \frac{9}{2} \xi(r_p) \int_0^1 x_1^2\, \mathrm{d}x_1 \int_0^1 x_2^2\, \mathrm{d}x_2$$
$$\times \int_{-1}^1 \mathrm{d}\mu\, \frac{1}{\left(x_1^2 + x_2^2 - 2x_1 x_2 \mu\right)^{\alpha/2}}. \tag{15}$$

For $\alpha = 0$, the integral evaluates to 2/9. And for $\alpha = 1$, 2 and 3, it evaluates to 0.27, 0.50 and 5.0, respectively. The integral is 0.41 for $\alpha = 1.8$.

To begin, we take the radius of the sphere so that the probability to find three or more galaxies is small compared with that to find two. Roughly speaking (neglecting corrections from higher order clustering that will become apparent below), this requires the probability to find two galaxies in δV to be small compared with that to find one. We thus require the radius r_p to be chosen small enough so that $n\delta V\ (1 + \langle \delta_p^2 \rangle) \lesssim 1$, or usually just $n\delta V \langle \delta_p^2 \rangle \lesssim 1$, since we will often have $\langle \delta_p^2 \rangle \gtrsim 1$.

If two galaxies fall within the same radius-r_p sphere, then we call this a pair. If δP (cf. equation 13) is the probability to find two galaxies in a volume δV, and if $\delta P \ll 1$, then the spatial density n_2 of pairs is $\delta P/\delta V = n^2(\delta V)(1 + \langle \delta_p^2 \rangle)$. The pair–pair autocorrelation function $X(r)$, the excess probability over random to find a pair in each of two volumes δV_1 and δV_3 separated by a distance r_{13}, is defined by

$$\delta P = n_2^2\,\delta V_1\,\delta V_3\,[1 + X(r_{13})]$$
$$= n^4(\delta V_1)^2(\delta V_3)^2\left(1 + \left\langle \delta_{\rm p}^2 \right\rangle\right)^2 [1 + X(r_{13})]\,, \tag{16}$$

where δP is here the joint probability to find one pair in δV_1 and another in δV_3.

A pair of pairs is a quadruplet. To describe the clustering of pairs of galaxies, we will therefore need the four-point correlation function. The joint differential probability to find objects in differential volume elements $\mathrm{d}V_1$, $\mathrm{d}V_2$, $\mathrm{d}V_3$ and $\mathrm{d}V_4$ located, respectively, at positions $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, $\boldsymbol{r}_3$, and $\boldsymbol{r}_4$ is (Peebles 1980)

$$\begin{aligned}\delta P = {} & n^4\,\mathrm{d}V_1\,\mathrm{d}V_2\,\mathrm{d}V_3\,\mathrm{d}V_4 \\ & \times [1 + \xi_{12} + \text{ 5 permutations} \\ & + \zeta(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3) + \text{ 3 permutations} \\ & + \xi_{12}\xi_{34} + \text{ 2 permutations} \\ & + \eta(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3, \boldsymbol{r}_4)].\end{aligned} \tag{17}$$

Here, $\zeta(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3)$ is the reduced (or 'connected') three-point correlation function, η is the reduced (connected) four-point correlation function, and we have introduced the shorthands $r_{ij} \equiv |\boldsymbol{r}_i - \boldsymbol{r}_j|$ and also $\xi_{ij} \equiv \xi(|\boldsymbol{r}_i - \boldsymbol{r}_j|)$. The quantity in brackets is the complete four-point autocorrelation function. For Gaussian perturbations, $\zeta = \eta = 0$.

To find the pair autocorrelation function, we now consider the case where two of the galaxies (1 and 2) are in one volume (δV_1) centred at $\boldsymbol{r}_1$ and the other two (3 and 4) are in another (δV_3) centred at $\boldsymbol{r}_3$. We also assume that the separation $|\boldsymbol{r}_1 - \boldsymbol{r}_3| \gg r_{\rm p}$. The joint probability to find two galaxies in δV_1 and two in δV_3 is thus

$$\begin{aligned}\delta P = {} & n^4 \int_{\delta V_1} \mathrm{d}^3x_1 \int_{\delta V_1} \mathrm{d}^3x_2 \int_{\delta V_3} \mathrm{d}^3x_3 \int_{\delta V_3} \mathrm{d}^3x_4 \\ & \times\ [1 + \xi_{12} + \xi_{34} + 4\xi_{13} + \xi_{12}\xi_{34} \\ & + 2\xi_{13}^2 + 2\zeta(r_{12}, r_{13}, r_{13}) + 2\zeta(r_{34}, r_{13}, r_{13}) \\ & +\ \eta(r_{12}, r_{13}, r_{14}, r_{23}, r_{24}, r_{34})]\,;\end{aligned} \tag{18}$$

note that in this equation (only), $\xi_{12} = \xi(|\boldsymbol{x}_1 - \boldsymbol{x}_2|)$ and similarly for ξ_{34}. We next note that

$$\int_{\delta V_1} \mathrm{d}^3x_1 \int_{\delta V_1} \mathrm{d}^3x_2 \int_{\delta V_3} \mathrm{d}^3x_3\,\zeta(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$$
$$= (\delta V)^3 \left\langle \delta_{\rm p}^2(\boldsymbol{r}_1)\delta_{\rm p}(\boldsymbol{r}_3)\right\rangle_{\rm c}\,, \tag{19}$$

the (reduced) three-point correlation function (with two of the three points coincident) for the smoothed density field, and

$$\int_{\delta V_1} \mathrm{d}^3x_1 \int_{\delta V_1} \mathrm{d}^3x_2 \int_{\delta V_3} \mathrm{d}^3x_3 \int_{\delta V_3} \mathrm{d}^3x_4\,\eta(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4)$$
$$= (\delta V)^4 \left\langle \delta_{\rm p}^2(\boldsymbol{r}_1)\delta_{\rm p}^2(\boldsymbol{r}_3)\right\rangle_{\rm c}\,, \tag{20}$$

a (reduced) four-point correlation function. Equating equations (16) and (18), we find

$$\begin{aligned}X(r) = {} & \left[4\xi(r) + 2\xi^2(r) + 4\left\langle \delta_{\rm p}^2(\boldsymbol{x})\delta_{\rm p}(\boldsymbol{x}+\boldsymbol{r})\right\rangle_{\rm c}\right. \\ & \left. +\ \left\langle \delta_{\rm p}^2(\boldsymbol{x})\delta_{\rm p}^2(\boldsymbol{x}+\boldsymbol{r})\right\rangle_{\rm c}\right] \Big/ \left(1 + \left\langle \delta_{\rm p}^2\right\rangle\right)^2.\end{aligned} \tag{21}$$

This result becomes exact in the limit that $r \gg r_{\rm p}$ and $n\delta V\ (1 + \langle\delta_{\rm p}^2\rangle) \ll 1$, and it is valid for any galaxy–galaxy two-, three- and four-point autocorrelation functions. We thus find that the calculation of the pair correlation function reduces to the calculation of the correlation of the density $\delta_{\rm p}$ with $\langle\delta_{\rm p}^2\rangle$ and the autocorrelations of $\langle\delta_{\rm p}^2\rangle$, a result that should come as no surprise.

We will define the effective pair bias via $b_{\rm p}^2 \equiv [X(r)/\xi(r)]$; it is the excess bias of pairs relative to individual objects. Note then that, according to Section 2, the net merger bias is $b_{\rm m} = b_{\rm h}\,b_{\rm p}$.

4 PAIR CLUSTERING FOR GAUSSIAN PERTURBATIONS

For Gaussian perturbations, $\zeta = \eta = 0$ and the pair–pair autocorrelation function simplifies to

$$X(r) = \frac{4\xi(r) + 2[\xi(r)]^2}{\left(1 + \left\langle \delta_{\rm p}^2\right\rangle\right)^2}. \tag{22}$$

In the limit of weak correlations, $\langle\delta_{\rm p}^2\rangle$, $\xi \ll 1$, $X(r) \simeq 4\xi(r)$. This is easy to understand: given two galaxies in the first cell, each contributes a factor of $\xi(r)$ to the excess probability to find a galaxy in the second cell (at least to linear order), and for $X(r)$ there are two such galaxies in the second cell. Although of interest academically, this limit is probably not relevant for galaxies or clusters of galaxies, as a value $\langle\delta_{\rm p}^2\rangle \lesssim 1$ requires that we deal with objects that are so rare that their mean separations are $\gtrsim$Mpc.

If $\xi(r) \lesssim 1$ and $\langle\delta_{\rm p}^2\rangle \gtrsim 1$, then the clustering of pairs is suppressed relative to that of individual galaxies, a consequence of the scarcity of pairs relative to individual galaxies. In the limit of strong clustering, $\xi(r)$, $\langle\delta_{\rm p}^2\rangle \gg 1$, the pair correlation function becomes $X(r) \simeq 2\,[\xi(r)]^2/\langle\delta_{\rm p}^2\rangle^2$, which is again suppressed relative to the galaxy correlation function. The applicability of this limit, however, should be questioned, as $\xi \gtrsim 1$ generally implies non-Gaussian perturbations. Interestingly, this simple exercise implies that merger bias can operate in different directions, depending on the regime of interest – as indeed the simulations discussed above find.

5 CLUSTERING OF GAUSSIAN PEAKS

We have just seen that if objects trace the distribution of mass in a system with Gaussian perturbations with some specified correlation function, then the pair correlation function is suppressed relative to the normal correlation function, unless the correlations are weak, in which case it can be enhanced by up to a factor of 4. If, however, objects form only at high-density peaks of a primordial density distribution, then the distribution of these objects will be non-Gaussian. That this is true is easy to see. The one-point probability distribution function for Gaussian perturbations is $P(\delta) \propto \mathrm{e}^{-\delta^2/2\sigma^2}$, where σ^2 is the variance. This distribution has zero mean, no skewness, no kurtosis and no higher order (reduced) cumulants. The one-point probability distribution of high-density peaks is $P(\delta) \propto \mathrm{e}^{-\delta^2/2\sigma^2}$ for $\delta > \nu\sigma$ and $P(\delta) = 0$ for $\delta < \nu\sigma$. This distribution has non-zero mean, non-zero skewness, kurtosis, etc.

This non-Gaussianity introduces non-zero reduced three- and four-point correlation functions (Politzer & Wise 1984; Bardeen et al. 1986; Jensen & Szalay 1986; Melott & Fry 1986), even if the total density perturbation amplitude is linear, $\xi \lesssim 1$. Although the exact expressions can be complicated, they simplify considerably when $\nu \gg 1$. In this limit, the full n-point correlation function can be written in terms of the galaxy two-point correlation function $\xi_{\rm g}(r)$ as (Politzer & Wise 1984)

$$1 + \xi_{\rm g}^{(n)}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n) \simeq \prod_{i>j}[\xi_{\rm g}(r_{ij}) + 1]. \tag{23}$$

The galaxy correlation function $\xi_{\rm g}(r)$ is the correlation function for the objects, rather than for the total mass. Thus, we can have $\xi_{\rm g} \gtrsim 1$ even in the linear regime, $\xi(r) \lesssim 1$, if the objects are highly biased tracers of the mass distribution. In this case, we can simply replace the expression in brackets in equation (18) with $[1 + \xi_{\rm g}^{(4)}]$ from equation (23). Then, the pair autocorrelation function becomes

$$X(r) \simeq [1 + \xi_{\rm g}(r)]^4 - 1. \qquad (24)$$

This equation is the central result of this section. It says that, if objects trace the distribution of peaks in a Gaussian density distribution, then the clustering of pairs can be strongly enhanced relative to the clustering of individual objects. Equation (24) is valid for highly biased objects ($\nu \gg 1$) on scales at which the underlying matter fluctuations are linear (even if fluctuations in the population of the objects is *not* small; Politzer & Wise 1984). It thus applies to haloes well above the characteristic mass scale (such as submillimetre galaxies at $z \sim 3$ or extremely massive clusters at the present day). Physically, higher order clustering – in particular, the four-point correlation function from equation (23), which provides non-zero reduced three- and four-point functions – of high-density peaks is enhanced with this type of non-Gaussianity, and this favours the clustering of pairs over individual objects. Thus, if mergers can be equated with close pairs of galaxies, we do expect a significant merger bias in the limit $\nu \gg 1$.

6 QUASI-LINEAR PERTURBATIONS

Equation (21) shows that the pair correlation depends on the three- and four-point correlation functions. The previous section showed that such terms do appear if galaxies are associated with peaks in the density field. However, another way to produce non-zero higher order correlations is through gravitational processes, and it is interesting to consider how such processes could affect pair correlations (and hence the merger bias). We therefore next consider objects that are distributed like the mass for a non-Gaussian mass distribution produced by gravitational amplification, to the quasi-linear regime, of primordial Gaussian perturbations. At redshift $z = 0$, the quasi-linear regime occurs at $\sim$10 Mpc; at redshift $z = 3$, it occurs at $\sim$1 Mpc. The bispectrum and trispectrum for this case can be calculated from cosmological perturbation theory and from them the three- and four-point correlation functions. The expressions can be quite formidable (Goroff et al. 1986), but fortunately for us, Bernardeau (1996; see also Bernardeau et al. 2002) has calculated the quantities required here. In particular, in the non-linear regime,

$$\left\langle \delta_{\rm p}^2(\boldsymbol{x}_1)\delta_{\rm p}(\boldsymbol{x}_2)\right\rangle_{\rm c} = C_{2,1}\left\langle \delta_{\rm p}^2\right\rangle \xi(|\,\boldsymbol{x}_1 - \boldsymbol{x}_2\,|), \qquad (25)$$

and

$$\left\langle \delta_{\rm p}^2(\boldsymbol{x}_1)\delta_{\rm p}^2(\boldsymbol{x}_2)\right\rangle_{\rm c} = C_{2,1}^2\left\langle \delta_{\rm p}^2\right\rangle^2 \xi(|\,\boldsymbol{x}_1 - \boldsymbol{x}_2\,|), \qquad (26)$$

where

$$C_{2,1} = \frac{68}{21} + \frac{1}{3}\frac{{\rm d}\log\left\langle \delta_{\rm p}^2\right\rangle}{{\rm d}\log r_{\rm p}}. \qquad (27)$$

In the limit that $\langle\delta_{\rm p}^2\rangle \gg 1, \xi$, we find

$$X(r) \simeq C_{2,1}^2\,\xi(r). \qquad (28)$$

We note that ${\rm d}\log(\langle\delta_{\rm p}^2\rangle)/{\rm d}\log\, r_{\rm p} = {\rm d}\log\ \xi/{\rm d}\log\, r$. For the scales probed by LBGs, the linear-theory correlation function is roughly $\xi \propto r^{-2}$, while stable clustering leads to a correlation function $\xi(r) \propto r^{-1.8}$. For these correlation-function scalings, $X(r) \simeq 7\,\xi(r)$; i.e. pairs are biased by roughly a factor of 2.6 relative to galaxies. If, on the other hand, $\xi(r) \propto$ constant at small radii (as expected for $P(k) \propto k^n$ with $n = -3$), then $X(r) \simeq 10\,\xi(r)$. We thus find that in the quasi-linear regime, pairs can be biased, perhaps strongly so, compared with the individual objects, even if they trace the mass. This could further enhance the clustering of mergers, if they are associated with pairs of objects. We emphasize that equation (28) is applicable on scales at which the underlying mass perturbations have $\xi \sim 1$ and assumes that the objects of interest exactly trace the mass distribution. They are thus only directly applicable in the limited regime of relatively unbiased objects on moderately small scales, although the qualitative results likely apply to more biased objects as well (see the Discussion at the end of Section 7).

7 HALO CLUSTERING MODEL

We will now briefly consider pair clustering in the highly non-linear regime. In this case, perturbation theory is no longer appropriate, so we will turn to the halo model of the density field. The halo clustering model postulates a distribution of virialized dark matter haloes, each with a radial (r) density profile $\rho_{\rm h}(m;r)$ that depends on its mass m. On large scales, the clustering is that of biased peaks, possibly in the quasi-linear regime, which we already considered above. On non-linear scales, the clustering is described within individual haloes. Of course, in this 'one-halo' regime, the distribution of objects is ultimately due to the interactions between them (such as dynamical friction acting on satellite galaxies). Our treatment is thus only approximate: it predicts the clustering of pairs given a density profile and implicitly ignores interactions. It could, nevertheless, be useful inside clusters of galaxies in which a population of small 'tracer' haloes orbit in a potential dominated by the massive cluster.

For the purposes of illustration, we suppose that all haloes have the same mass and power-law radial density profile: $\rho \propto r^{-\gamma}$ for $r < R$, and $\rho(r) = 0$ for $r > R$. We will only consider correlations on small scales, within an individual halo (which should be appropriate on small scales in the highly non-linear regime). The autocorrelation function for the mass is then (Scherrer & Bertschinger 1991; Cooray & Sheth 2002)

$$\xi(|\,\boldsymbol{r}_1 - \boldsymbol{r}_2\,|) = \frac{\langle\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)\rangle}{\langle\rho\rangle^2} - 1, \qquad (29)$$

where the angle brackets denote an average over all space. The mean density is $\langle\rho\rangle = n_{\rm halo}M$, where $n_{\rm halo}$ is the spatial number density of haloes and M is the halo mass, and

$$\langle\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)\rangle = n_{\rm halo}\int {\rm d}^3x\ \rho(|\,\boldsymbol{r}_1 - \boldsymbol{x}\,|)\rho(|\,\boldsymbol{r}_2 - \boldsymbol{x}\,|). \qquad (30)$$

The integral in equation (30) is particularly simple at zero lag, where the autocorrelation function for the mass is

$$\xi(r = 0) = \left(4\pi n_{\rm halo}R^3\right)^{-1}\frac{(3-\gamma)^2}{(3-2\gamma)} - 1, \qquad (31)$$

for $\gamma < 3/2$. For $\gamma > 3/2$, the divergence at the $r \rightarrow 0$ limit of the integrand can be tempered by measuring correlations over a finite smoothing volume of radius $r_{\rm s}$ (as would occur in any physical observation). Thus, for $\gamma > 3/2$, we find

$$\xi(r = 0) = \left(4\pi n_{\rm halo}R^3\right)^{-1}\frac{(3-\gamma)^2}{|\,3-2\gamma\,|}\left(\frac{r_{\rm s}}{R}\right)^{3-2\gamma} - 1. \qquad (32)$$

For $r \lesssim R$ and $\gamma > 3/2$ (and for $n_{\rm halo}\ R^3 \ll 1$), the mass correlation function scales with radius r as $\xi(r) \propto r^{3-2\gamma}$; for $\gamma < 3/2$, it decreases less rapidly with radius. For $\gamma = 3/2$, the power laws are replaced by logarithms.

The pair correlation function follows simply by noting that pairs are distributed in the halo as ρ^2. We can therefore simply replace $\gamma \to 2\gamma$ in the results for the mass correlation functions. Thus, for $\gamma < 3/4$, the zero-lag pair correlation function is

$$X(r=0) = \left(4\pi n_{\rm halo} R^3\right)^{-1} \frac{(3-2\gamma)^2}{(3-4\gamma)} - 1, \tag{33}$$

and for $\gamma > 3/4$,

$$X(r=0) = \left(4\pi n_{\rm halo} R^3\right)^{-1} \frac{(3-2\gamma)^2}{|\,3-4\gamma\,|} \left(\frac{r_{\rm s}}{R}\right)^{3-4\gamma} - 1. \tag{34}$$

The pair correlation function scales, for $\gamma > 3/4$, with radius r as $X(r) \propto r^{3-4\gamma}$, and it decreases less rapidly with r for $\gamma < 3/4$. For $3/4 < \gamma < 3/2$, the pair correlation diverges (modulo the smoothing) at small radii, while the mass correlation function approaches a constant as $r \to 0$.

We thus see that the distribution of pairs and mass differs, and thus that there should be a (scale-dependent) bias between them. Our calculation is applicable in the non-linear regime, when the correlation function is measured at distances $r \ll R$. The pair bias can then be approximated by the square root of the ratio of zero-lag biases. For example, if $\gamma = 1/2$, then the pair bias evaluates to $b_{\rm p} = 2^{5/2}/5 \simeq 1.1$. For $\gamma \to 0$, the pair bias approaches 1, which is what we expect for objects distributed uniformly in a halo. The zero-lag bias may be considerably larger for $3/4 < \gamma < 3/2$, when the pair correlation function diverges as $r \to 0$, while the mass correlation does not.

So far, we have considered pair correlations for a highly biased population in the linear regime as well as for a population that traces the mass in the quasi-linear and non-linear regimes. What about pair correlations for a highly biased population in the quasi-linear or non-linear regimes? It has been argued that in the quasi-linear regime, highly biased tracers are more likely to be found in denser regions (Cole & Kaiser 1989; Sheth & Tormen 1999); calculation of the pair correlation for a population biased in Lagrangian space evolved into the quasi-linear regime could be done following the techniques of Fry (1996), Catelan et al. (1998) and Catelan, Porciani & Kamionkowski (2000), but we leave that for future work. And what about the non-linear regime? Numerical simulations have suggested that the distribution of primordial density peaks in larger virialized haloes (i.e. the non-linear regime) is more highly peaked toward the centres than the mass as a whole (Moore et al. 1998; White & Springel 2000; Santos 2003; Diemand, Madau & Moore 2005). If so, and if, as we have seen, the bias of pair correlations is enhanced with steeper density profiles, then the bias of pair correlations for rare objects in the quasi-linear and non-linear regimes may be even further enhanced.

8 DISCUSSION

In this paper, we have investigated the implications of the extended Press–Schechter and Mo & White (1996) biasing scheme for merger bias and pointed out some shortcomings and ambiguities in this approach. In particular, we showed that this approach yields *no* merger bias, but only because it explicitly ignores the variation of merger rates with the large-scale density field. We then showed that a simple model in which the merger rate scales only with the halo abundances predicts that mergers of massive galaxies will be more biased than the halo population but that mergers of small galaxies will be less biased. Furthermore, the merger bias will evolve significantly with redshift. These results may provide useful clues to reconciling the various simulations (Percival et al. 2003; Scannapieco & Thacker 2003; Gao et al. 2005). However, these techniques are clearly inadequate for understanding merger bias on any quantitative level (at least until a self-consistent merger kernel is available).

We therefore moved on to hypothesize that close pairs in a clustering model are likely to yield mergers. We thus studied the clustering of close pairs in a variety of models in which objects Poisson sample: (i) the mass in a Gaussian random field; (ii) the high-density peaks in a Gaussian random field; (iii) the mass in the quasi-linear regime and (iv) the mass in virialized haloes with power-law density profiles. We find that in many (though not all) cases, close pairs can be more highly clustered than individual objects. If so, and if close pairs are likely to lead to mergers, then the clustering of objects that have undergone recent mergers can be enhanced relative to the clustering of individual haloes of comparable masses. We have thus shown that, in the simplest picture of mergers, an extra bias (of some magnitude) is generic to most clustering models. The actual magnitude of the bias [or the lack of it, as in the simulations of Percival et al. (2003) and Gao et al. (2005)] is therefore revealing something fundamental about the halo-merging process – an area in need of substantial theoretical insight (Benson et al. 2005).

Even if we do identify close pairs with mergers, there are still a multitude of theoretical steps – each fraught with considerable uncertainties – that must be taken to connect close pairs of galactic haloes with, for example, the observational constraints on LBGs. We have considered the behaviour under a variety of limits, but the more general case must be treated numerically. Still, it is interesting to investigate whether pair biasing might be in the right ballpark to account for the discrepancy between the LBG dynamical and clustering masses. According to Adelberger et al. (1998), the bias of LBGs is $b_{\rm LBG} \sim 4.0$, roughly consistent with that expected for $\sim 10^{12}$-$\mathrm{M}_\odot$ objects [see also Adelberger et al. (2005), who estimate a similar median mass for a larger sample of objects at $z = 3$]. Although the abundance of haloes with such masses is consistent with the abundance of LBGs, it requires that *every* such halo houses a galaxy that produces stars at a prodigious rate (Adelberger & Steidel 2000). On the other hand, the linewidths and kinematics of LBGs suggest masses closer to $\sim 10^{11}\,\mathrm{M}_\odot$ (Pettini et al. 2001; Erb et al. 2003). Haloes of these masses have a much higher abundance, allowing consistency with the LBG abundance if the efficiency for $\sim 10^{11}$-$\mathrm{M}_\odot$ haloes to produce extremely luminous objects is relatively low, $\sim$10 per cent – understandable, perhaps, if only recent mergers of $\sim 10^{11}$-$\mathrm{M}_\odot$ haloes produce LBGs. (An alternate possibility is that dynamical mass measurements are only sensitive to a small fraction of the halo and that LBGs are ubiquitous in large dark matter haloes: Cooray 2005.)

The only remaining problem with the small-mass LBG scenario is why the bias $b_{\rm LBG} \sim 4$ is so much larger than the bias $b_{11} \approx 2.4$ expected for a sample with $\sim 10^{11}$-$\mathrm{M}_\odot$ haloes. Adelberger et al. (1998) measure the clustering through a counts-in-cells analysis within boxes of size $11.4\,h_{100}^{-1}$ Mpc. This is within the linear regime at redshifts $z \sim 3$, and with an expected bias $b_{11} \approx 2.4$, the variance in the $\sim 10^{11}$-$\mathrm{M}_\odot$ halo distribution is $\sigma_{\rm gal} \simeq 0.8$. It is also reasonable to assume a pair spacing with $\langle \delta_{\rm p}^2 \rangle \gg 1$. Although the pair bias implied by equation (24) is not linear, most of the weight for the counts-in-cells analysis occurs at the largest radii. We thus estimate from equation (24) a pair bias (i.e. the extra biasing of mergers relative to the objects themselves) of $b_{\rm p} = \sqrt{X(r)/\xi_{\rm g}(r)} \simeq 3.4$. This is more than enough to make the net merger bias ($b_{\rm m} = b_{\rm p} b_{\rm h}$) comparable to $b_{\rm LBG}$. However, note that $\nu \approx 1.6$ for 10^{11}-$\mathrm{M}_\odot$ haloes, so the true amplification should be smaller than the $\nu \gg 1$ limit we have taken. This may be further augmented by quasi-linear effects, which could contribute a comparable pair bias over some fraction of the cell.

A similar, though perhaps even more desperate, problem occurs for submillimetre-selected galaxies. Blain et al. (2004) claim that the clustering of these galaxies indicates halo masses of $\sim 10^{13}\,M_{\odot}$ while kinematic measurements yield values an order of magnitude smaller, even allowing for the mass in the outer regions of the halo. Our results may help resolve this discrepancy as well, if submillimetre galaxies are the products of recent mergers. Moreover, Blain et al. (2004) measured clustering through the rate of incidence of close pairs in their survey fields. They assumed a correlation function of fixed shape $\xi_g(r) \propto r^{-1.8}$ and varied its amplitude until they recovered the observed number of pairs; the inferred correlation length could then be matched to a halo mass. We have shown that the clustering of pairs is *not* the same as the clustering of the underlying objects and depends on the underlying halo population, the scales of interest and even the relation of haloes to the underlying density field. The effective pair bias can be significantly larger than the bias of the haloes themselves, so pair-counting techniques must be approached with care. The precise effects are difficult to predict given the 'pencil-beam' geometries of their surveys, but they certainly merit further study.

Before closing, we note that our results may be applicable elsewhere as well. For example, galaxy clusters are highly biased tracers of the mass distribution today (Bahcall et al. 2003). Their correlation length may be as large as $\sim 25\,h_{100}^{-1}$ Mpc, as opposed to a correlation length ~ 5–$7\,h_{100}^{-1}$ Mpc for the mass. If this bias occurs because clusters form at peaks of the primordial density distribution, then they should experience higher order clustering as described in Section 5. Moreover, at distances $\gtrsim 10\,h_{100}^{-1}$ Mpc, quasi-linear effects should be small. There will thus be testable predictions for the clustering of close pairs of clusters, or – if pairs are associated with mergers – for the clustering of recently merged clusters. As another example, non-trivial merger bias would modify the interpretation of AGN clustering (provided that they are fuelled by merger activity). This would be particularly important for understanding their host properties and their lifetimes (La Franca et al. 1998; Haiman & Hui 2001; Martini & Weinberg 2001; Adelberger & Steidel 2005a,b). We leave further discussion of these possibilities to future work.

ACKNOWLEDGMENTS

We thank the referee, L. Miller, for helpful comments. This work was supported in part by DoE DE-FG03-92-ER40701 and NASA NNG05GF69G.

REFERENCES

Adelberger K. L., Steidel C. C., 2000, ApJ, 544, 218
Adelberger K. L., Steidel C. C., 2005a, ApJ, 627, L1
Adelberger K. L., Steidel C. C., 2005b, ApJ, 630, 50
Adelberger K. L., Steidel C. C., Giavalisco M., Dickinson M., Pettini M., Kellogg M., 1998, ApJ, 505, 18
Adelberger K. L., Steidel C. C., Pettini M., Shapley A. E., Reddy N. A., Erb D. K., 2005, ApJ, 619, 697
Bahcall N. A., Dong F., Hao L., Bode P., Annis J., Gunn J. E., Schneider D. P., 2003, ApJ, 599, 814
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Benson A. J., Kamionkowski M., Hassani S. H., 2005, MNRAS, 357, 847
Bernardeau F., 1996, A&A, 312, 11
Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rep., 367, 1
Blain A. W., Chapman S. C., Smail I., Ivison R., 2004, ApJ, 611, 725
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Catelan P., Lucchin F., Matarrese S., Porciani C., 1998, MNRAS, 297, 692
Catelan P., Porciani C., Kamionkowski M., 2000, MNRAS, 318, L39
Cole S., Kaiser N., 1989, MNRAS, 237, 1127
Coles P., Lucchin F., Matarrese S., Moscardini L., 1998, MNRAS, 300, 183
Cooray A., 2005, MNRAS, 364, 303
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Diemand J., Madau P., Moore B., 2005, MNRAS, 364, 367
Efstathiou G., Frenk C. S., White S. D. M., Davis M., 1988, MNRAS, 235, 715
Erb D. K., Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., Hunt M. P., Moorwood A. F. M., Cuby J., 2003, ApJ, 591, 101
Fry J. N., 1996, ApJ, 461, L65
Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66
Giavalisco M., Dickinson M., 2001, ApJ, 550, 177
Giavalisco M., Steidel C. C., Adelberger K. L., Dickinson M. E., Pettini M., Kellogg M., 1998, ApJ, 503, 543
Goroff M. H., Grinstein B., Rey S.-J., Wise M. B., 1986, ApJ, 311, 6
Gottlöber S., Kerscher M., Kravtsov A. V., Faltenbacher A., Klypin A., Müller V., 2002, A&A, 387, 778
Haiman Z., Hui L., 2001, ApJ, 547, 27
Jensen L. G., Szalay A. S., 1986, ApJ, 305, L5
Kaiser N., 1984, ApJ, 284, L9
Kauffmann G., Haehnelt M. G., 2002, MNRAS, 332, 529
Kolatt T. S. et al., 1999, ApJ, 523, L109
Lacey C., Cole S., 1993, MNRAS, 262, 627
La Franca F., Andreani P., Cristiani S., 1998, ApJ, 497, 529
Martini P., Weinberg D. H., 2001, ApJ, 547, 12
Melott A. L., Fry J. N., 1986, ApJ, 305, 1
Mo H. J., Fukugita M., 1996, ApJ, 467, L9
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
Moore B., Governato F., Quinn T., Stadel J., Lake G., 1998, ApJ, 499, L5
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Percival W. J., Scott D., Peacock J. A., Dunlop J. S., 2003, MNRAS, 338, L31
Pettini M., Shapley A. E., Steidel C. C., Cuby J., Dickinson M., Moorwood A. F. M., Adelberger K. L., Giavalisco M., 2001, ApJ, 554, 981
Politzer H. D., Wise M. B., 1984, ApJ, 285, L1
Porciani C., Giavalisco M., 2002, ApJ, 565, 24
Press W. H., Schechter P., 1974, ApJ, 187, 425
Santos M. R., 2003, PhD thesis, California Institute of Technology
Scannapieco E., Barkana R., 2002, ApJ, 571, 585
Scannapieco E., Thacker R. J., 2003, ApJ, 590, L69
Scherrer R. J., Bertschinger E., 1991, ApJ, 381, 349
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Wechsler R. H., Somerville R. S., Bullock J. S., Kolatt T. S., Primack J. R., Blumenthal G. R., Dekel A., 2001, ApJ, 554, 85
White S. D. M., Springel V., 2000, in Weiss A., Abel T. G., Hill V., eds, The First Stars: Proceedings of the MPA/ESO Workshop held at Garching. Germany, 4–6 August 1999. Springer, New York, p. 327

This paper has been typeset from a TEX/LATEX file prepared by the author.