The imprint of cosmic reionization on galaxy clustering

J. Stuart B. Wyithe1* and Abraham Loeb2⋆

1School of Physics, University of Melbourne, Parkville, Victoria, Australia
2Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

Accepted 2007 September 10. Received 2007 August 23; in original form 2007 June 25

ABSTRACT

We consider the effect of reionization on the clustering properties of galaxy samples at intermediate redshifts (\(z \approx 0.3–5.5\)). Current models for the reionization of intergalactic hydrogen predict that overdense regions will be reionized early, thus delaying the build-up of stellar mass in the progenitors of massive lower redshift galaxies. As a result, the stellar populations observed in intermediate-redshift galaxies are somewhat younger and hence brighter in overdense regions of the Universe. Galaxy surveys would therefore be sensitive to galaxies with a somewhat lower dark matter mass in overdense regions. The corresponding increase in the observed number density of galaxies can be parametrized as a galaxy bias due to reionization. We model this process using merger trees combined with a stellar synthesis code. Our model demonstrates that reionization has a significant effect on the clustering properties of galaxy samples that are selected based on their star formation properties. The bias correction in Lyman-break galaxies (including those in proposed baryonic oscillation surveys at \(z < 1\)) is at the level of 10–20 per cent for a halo mass of \(10^{12} \, M_\odot\), leading to corrections factors of \(1.5–2\) in the halo mass inferred from measurements of clustering length. The reionization of helium could also lead to a sharp increase in the amplitude of the galaxy correlation function at \(z \approx 3\). We find that the reionization bias is approximately independent of scale and halo mass. However, since the traditional galaxy bias is mass dependent, the reionization bias becomes relatively more important for lower mass systems. The correction to the bias due to reionization is very small in surveys of luminous red galaxies at \(z < 1\).

Key words: galaxies: high-redshift – intergalactic medium – cosmology: theory – diffuse radiation – large-scale structure of Universe.

1 INTRODUCTION

The clustering of galaxies is often used to study the power spectrum of the underlying mass distribution (e.g. Tegmark et al. 2006). Since the data do not reflect the clustering of mass but rather the clustering of galaxies, a correction factor termed the galaxy bias must be applied to its analysis. This bias factor is a mass-dependent (and possibly scale-dependent) property of the galaxy population. Alternatively, assuming that the mass power spectrum is known, the masses of the galaxies can be inferred from their bias by comparing the expected clustering of mass with the observed galaxy clustering. In either case, if one is trying to determine galaxy mass using clustering or one is trying to determine the underlying properties of the mass power spectra from observations of galaxy clustering, a theoretical understanding of galaxy bias is required. As a first estimate, the bias can be calculated from linear theory (Mo & White 1996; Sheth, Mo & Tormen 2001). In addition, through comparison with N-body simulations, various corrections to the bias have also been calculated (see Eisenstein et al. 2005, for a summary) to allow more accurate comparisons with improving data. These contributions to the bias are physical in the sense that they are due only to the properties of the dark matter, and so they can be computed (albeit via simulation) from first principles given an input mass power spectrum.

Recently it has been suggested that an inhomogeneous reionization can lead to a modification of the observed clustering of galaxies. Babich & Loeb (2006) calculated the modulation of the number density of the lowest mass galaxies that result from reionization-induced variation in the thermal history among different regions of the intergalactic medium (IGM). Although they have found that the expected effect on the galaxy power spectrum is much larger than the difference between competing models of inflation, their analysis did not extend to high-mass galaxies to which future surveys will be sensitive. Pritchard, Furlanetto & Kamionkowski (2007) considered lower redshifts and more massive galaxies, but used an ad hoc ansatz that the overdensity of galaxies is proportional to the underlying radiation field and concluded that reionization would leave a redshift-dependent imprint on the galaxy power spectrum.
low redshifts that might interfere with measurements of the baryonic acoustic peak. These papers did not attempt to compute the coupling between the mass-to-light ratio of massive galaxies and the large-scale environment. However, galaxy surveys produce clustering statistics for either flux limited surveys, or for volume-limited surveys in a fixed luminosity range. Computation of the effect of reionization on the mass-to-light ratio of massive galaxies is therefore critical for comparison with any real survey.

The aim of this paper is to estimate the astrophysical contribution to the galaxy bias due to the reionization of the IGM. This contribution is model dependent, requiring knowledge of the baryonic physics in addition to gravity. The reionization of the IGM is sensitive to the local large-scale overdensity. In regions that are overdense, galaxies are overabundant for two reasons: first because there is more material per unit volume to make galaxies, and second because small-scale fluctuations need to be of lower amplitude to form a galaxy when embedded in a larger scale overdensity. The first effect will result in a larger density of ionizing sources. However, this larger density will be compensated by the increased density of gas to be ionized. In addition, the recombination rate is increased in overdense regions, but this effect is counteracted by the bias of galaxies in these regions. The process of reionization also contains several layers of feedback. Radiative feedback heats the IGM and results in the suppression of low-mass galaxy formation (Efstathiou 1992; Quinn, Katz & Efstathiou 1996; Thoul & Weinberg 1996; Dijkstra et al. 2004). This delays the completion of reionization by lowering the local star formation rate, but here again the effect is counteracted in overdense regions by the biased formation of massive galaxies. The radiation feedback may therefore be more important in low-density regions where small galaxies contribute more significantly to the ionizing flux. Wyithe & Loeb (2007) have modelled the density-dependent reionization process using a semi-analytic model that incorporates the features described above, and so captures the important physical processes. This model demonstrated that galaxy bias leads to enhanced reionization in overdense regions, so that overdense regions are reionized first.

We show that this early reionization leads to an additional bias in the observed clustering at later epochs in addition to that associated with enhanced structure formation. We find that the correction to the linear bias due to reionization could be significantly larger than other corrections that have been previously considered. Moreover, we show that the bias correction is larger than the uncertainties in current surveys over a wide range of redshifts within $1 \leq z \leq 5$.

The outline of the paper is as follows. In Sections 2 and 3 we describe the effect of reionization on galaxy formation, and summarize galaxy bias. We then outline the reasons why we would expect reionization to yield an additional galaxy bias in Section 4, before presenting a model to allow quantitative predictions of the effect (Section 5). We then apply our model to surveys for Lyman-break galaxies (LBGs) (Section 6) and surveys to measure baryonic acoustic oscillations (Section 7). We then discuss some outstanding issues in Section 8 before summarizing our conclusions in Section 9. Throughout the paper we adopt the latest set of cosmological parameters determined by Wilkinson Microwave Anisotropy Probe (Spergel et al. 2007) for a flat ΛCDM universe.

2 REIONIZATION AND OBSERVED GALAXY FORMATION

The dominant effect of reionization on galaxy formation is believed to involve radiative feedback which heats the IGM following the reionization of a region, and thus results in the suppression of low-mass galaxy formation (Efstathiou 1992; Quinn et al. 1996; Thoul & Weinberg 1996; Dijkstra et al. 2004). Standard models of the reionization process assume a minimum threshold mass for galaxy haloes in which cooling and star formation occur ($M_{\text{cool}}$) within neutral regions of the IGM. In ionized regions the minimum halo mass is limited by the Jeans mass (Barkana & Loeb 2001) in an ionized IGM ($M_{\text{ion}}$). We assume $M_{\text{cool}}$ to correspond to a virial temperature of $10^5$ K, representing the hydrogen cooling threshold, and $M_{\text{ion}}$ to correspond to a virial temperature of $10^4$ K, representing the mass below which infall is suppressed from an IGM in which hydrogen has been ionized (Dijkstra et al. 2004).

Observations suggest that hydrogen was reionized by stars prior to $z \sim 6$ (e.g. White et al. 2003; Fan et al. 2006). However, models of He reionization suggest that it was the rise of quasars (with harder spectra) that resulted in the overlap of He regions at a redshift of $z \sim 3.5$ (e.g. Sokasian, Abel & Hernquist 2003; Wyithe & Loeb 2003). This prediction is consistent with observations that show transmission just blueward of the helium Lyα line at $z \sim 3$ (Jacobsen et al. 1994; Tytler 1995; Davidsen, Kriss & Zheng 1996; Hogan, Anderson & Rugers 1997; Reimers et al. 1997; Heap et al. 2000; Kriss et al. 2001; Smette et al. 2002). In addition, the double reionization of helium results in the temperature of the IGM being approximately doubled (Schaye et al. 2000; Theuns et al. 2002a,b). Thus we assume the IGM temperature to change from $T_{\text{IGM}} \sim 10^3$ to $T_{\text{IGM}} \sim 2 \times 10^4$ K between $z \sim 4$ and $z \sim 3$. Calculation of the accretion of baryons from an adiabatically expanding IGM into a dark matter potential well show that the minimum virial temperature for significant accretion is proportional to the temperature of the IGM (Barkana & Loeb 2001). Thus, when helium is reionized at $z \sim 3.5$, the value of $T_{\text{min}}$ is doubled from $T_{\text{ion}}$ to $2T_{\text{ion}}$. When considering Helium reionization we assume a sudden heating ($\Delta z \lesssim 0.1$). However, we note that the period of heating could be more prolonged.

3 GALAXY BIAS

Strong clustering of massive galaxies in overdense regions implies that these sources trace the higher density regions of IGM. The clustering of galaxies is driven by two effects. The first effect is the underlying clustering of the density field. This clustering may be expressed via the mass correlation function between regions of mass $M_1$ and $M_2$, separated by a comoving distance $R$ (see Scannapieco & Barkana 2002 and references therein)

$$
\xi_m(M_1, M_2, R) = \frac{1}{2\pi^2} \int dk k^2 P(k) \times \int W(kR_1)W(kR_2),
$$

(1)

where

$$
R_{1,2} = \left( \frac{3M_{1,2}}{4\pi \rho_m} \right)^{1/3},
$$

(2)

$W$ is the window function (top-hat in real space), $P(k)$ the power spectrum and $\rho_m$ is the cosmic mass density. The dark matter halo correlation function for haloes of mass $M$ is obtained from the product of the mass correlation function $\xi_m(M, M, R)$ and the square of the ratio between the variances of the halo and mass distributions. This ratio, $b$, is defined as the halo bias. This bias has been discussed extensively in the literature, (e.g. Mo & White 1996; Sheth et al. 2001). However, we briefly describe a likelihood based interpretation which allows the effects of reionization to be included in a natural way.
To see the origin of bias due to enhanced galaxy formation in overdense regions, consider the likelihood (which is proportional to the local number density of galaxies) of observing a galaxy at a random location. Given a large-scale overdensity $\delta$ of comoving radius $R$, the likelihood of observing a galaxy may be estimated from the Sheth & Tormen (2002) mass function as

$$L_\delta (\delta) = \left(1 + \delta \right)\left[1 + (v - 2)c_2 \right]^{-1/2},$$

where $v = (\delta - \delta_c)/[\sigma(R)]$, $\delta_c = 1.69$ is the critical linear overdensity for collapse to a bound object. Here $\sigma(R)$ is the variance of the density field smoothed with a top-hat window on a scale $R$ at redshift $z$, and $a = 0.707$ and $p = 0.3$ are constants. Note that here as elsewhere in this paper we work with overdensities and variances computed at the redshift of interest (i.e. not extrapolated to $z = 0$). Equation (3) is simply the ratio of the number density of haloes in a region of overdensity $\delta$ to the number density of haloes in the background universe. This ratio has been used to derive the bias for small values of $\delta$ (Mo & White 1996; Sheth et al. 2001). For example, in the Press & Schechter (1974) formalism we write

$$L_\delta (\delta) = (1 + \delta) \left(\frac{dn}{dM} (\bar{\delta}) + \frac{d^2 n}{dM dv} (\nu) \frac{d\nu}{d\bar{\delta}} \right) \left(\frac{dn}{dM} (\bar{\delta}) \right)^{-1},$$

$$\sim 1 + \frac{\nu^2 - 1}{\sigma(M)\nu} \equiv 1 + \delta b_g,$$

where $(dn/dM)(\nu)$ and $(d^2n/dMdv)(\nu)$ are the average and perturbed mass functions, and $b_g$ is defined as the bias factor due to reionization.

The observed overdensity of galaxies is $\delta_{gal} = (4/3)b_g(M, z)\delta$, where $b_g(M, z)$ is the galaxy bias, and the pre-factor of $4/3$ arises from a spherical average over the infall peculiar velocities (Kaiser 1987). The value of bias $b_g$ for a halo mass $M$ may be better approximated using the Press–Schechter formalism (Mo & White 1996), modified to include non-spherical collapse (Sheth et al. 2001):

$$b_g (M, z) = 1 + \frac{1}{\delta_{c}} \left[ \nu^2 + b \nu^{\nu(1-\nu)} - \nu^{2\nu}/\sqrt{\alpha} \right],$$

where $\nu = \delta_{gal}/\sigma(M)$, $\nu^* = \sqrt{\alpha}$. $a = 0.707$, $b = 0.5$ and $c = 0.6$. Here $\sigma(M)$ is the variance of the density field smoothed on a mass scale $M$ at redshift $z$. This expression yields an accurate approximation to the halo bias determined from $N$-body simulations (Sheth et al. 2001). Note that in linear theory the bias (equations 4 and 5) is a function of halo mass, but not of overdensity or scale.

### 4 REIONIZATION-INDUCED GALAXY BIAS IN OBSERVED GALAXY SAMPLES

We next introduce an additional galaxy bias due to reionization. In addition to colour selection criteria, clustering surveys typically consider galaxies that are either selected to be above a minimum flux threshold (e.g. Adelberger et al. 2005) or to lie in a particular absolute magnitude range (e.g. Eisenstein et al. 2005). Suppose that reionization caused an overdensity-dependent change in the flux per unit halo mass by a factor $\mu$. In either of the selection scenarios mentioned above, this effect will result in the host haloes of survey galaxies (i.e. at fixed luminosity) being smaller in that region by an average factor $\mu$. Following the previous formalism, we find the likelihood for observing a galaxy which is subject to a decrease in mass-to-light ratio of $\mu$ in a region of overdensity $\delta$,

$$L_{reion} (\delta) = \left(1 + \delta \right)\left(\frac{dn}{dM} (\nu) \right) \left[1 + \delta \frac{dn}{dM} (\nu) \right]^{-1},$$

$$\equiv 1 + \delta b_{reion} (\delta),$$

where $(dn/dM)(\nu)$ is the perturbed mass function evaluated at $M/\mu$, and $b_{reion}(\delta)$ is defined to be the bias factor due to reionization and which can be a function of $\delta$. Note that since surveys are magnitude limited or measured in logarithmic bins of luminosity, there is no factor of $\mu^{-1}$ as would be required when discussing the number counts per unit luminosity.$^1$

We may then write an expression for the likelihood of observing a galaxy that includes both the bias due to enhanced formation in overdense regions, and a possible effect of reionization

$$L (\delta) = L_\delta (\delta) L_{reion} (\delta)$$

$$\sim 1 + [b_g + b_{reion}(\delta)]\delta.$$

In the second equality we have parametrized the effect of variance in the reionization redshift as an additive contribution to the galaxy bias, and have then noted (in the third equality) that we are working in a regime where $b_\delta \ll 1$. In the next section we will develop a model that will allow us to estimate the magnitude of this effect.

### 5 PATCHY REIONIZATION AND GALAXY BIAS

In this section we describe a model for the effect of reionization on galaxy bias, and show that reionization increases the bias in the observed overdensity of galaxies relative to the underlying density field. In later sections we will use this model to make qualitative predictions for the impact of reionization on observed clustering in a range of galaxy samples. Our intention is not to produce a detailed model in order to make quantitative predictions or comparisons with the data. Such a model would require detailed numerical simulations, and would in any case require a number of uncertain astrophysical assumptions. However, our model is adequate for the purposes of assessing the importance of reionization in clustering measurements, and for making qualitative predictions about its dependence on quantities such as survey redshift and luminosity.

#### 5.1 Reionization redshift and large-scale overdensity

Large-scale inhomogeneity in the cosmic density field leads to structure formation that is enhanced in overdense regions and delayed in underdense regions. Thus, overlap of ionized regions and hence heating of the IGM would have occurred at different times in different regions due to the cosmic scatter in the process of structure formation within finite spatial volumes (Barkana & Loeb 2004). The reionization of hydrogen would have been completed within a region of comoving radius $R$ when the fraction of mass incorporated into collapsed objects in that region attained a certain critical value, corresponding to a threshold number of ionizing photons emitted per baryon. The ionization state of a region is governed by the enclosed ionizing luminosity, by its overdensity, and by dense pockets of neutral gas that are self shielding to ionizing radiation. There is

---

$^1$ There is also no factor of $\mu^{-1}$ to account for depletion as would be appropriate if the enhancement in flux were due to gravitational lensing.
an offset (Barkana & Loeb 2004) $\delta z$ between the redshift when a region of mean overdensity $\delta$ achieves this critical collapsed fraction, and the redshift $\bar{z}$ when the universe achieves the same collapsed fraction on average. This offset may be computed (Barkana & Loeb 2004) from the expression for the collapsed fraction (Bond et al. 1991) $F_{\text{col}}$ within a region of overdensity $\delta$ on a comoving scale $R$,

$$F_{\text{col}}(M_{\text{min}}) = \text{erfc} \left( \frac{\delta - \delta_c}{\sqrt{2(\sigma_{\text{delta}}^2 - \sigma_{\bar{z}}^2)}} \right),$$

yielding

$$\frac{\delta z}{(1 + \bar{z})} = \frac{\delta}{\delta_c} \left( 1 - \frac{\sigma_{\bar{z}}^2}{\sigma_{\text{delta}}^2} \right), \quad (9)$$

where $\sigma_{\text{delta}}$ and $\sigma_{\bar{z}}$ are the variances in the power spectrum at $z$ on comoving scales corresponding to the region of interest and to the minimum galaxy mass $M_{\text{min}}$, respectively. On large scales equation (9) reduces to

$$\delta z \approx (1 + \bar{z}) \frac{\delta}{\delta_c}. \quad (10)$$

The offset in the ionization redshift of a region depends on its linear overdensity, $\delta$. As a result, the distribution of offsets, and therefore the scatter in the ionization redshift may be obtained directly from the power spectrum of primordial inhomogeneities (Wyithe & Loeb 2004). As can be seen from equation (10), larger regions have a smaller scatter due to their smaller cosmic variance. Note that equation (10) is independent of the critical value of the collapsed fraction required for reionization. We also note that since at high redshift the variance of the linear density field increases approximately in proportion to $(1 + z)^{-3}$, the typical delay in redshift is almost independent of cosmic time (in addition to not being a function of collapsed fraction).

Following the reionization of hydrogen, doubly ionized helium remained in the pre-overlap phase. At this time, the mean free path of He II ionizing photons was therefore limited to be smaller than the size of the He II regions. As is the case for hydrogen, the ionization state of these regions was therefore dependent on the local source population. If it is true that quasars were responsible for the reionization of helium, then these are much rarer sources than the galaxies responsible for the reionization of hydrogen. As a result there would be large fluctuations in the He II reionization redshift due to Poisson fluctuations in the number of sources and variations in the opacity of the IGM (Reimers et al. 2006). These fluctuations would not be simply related to the local large-scale overdensity. On the other hand, the arguments regarding the fluctuations in the redshift of hydrogen reionization due to enhanced structure formation in overdense regions must also apply to the reionization of helium, and these will be present in addition to the Poisson noise. As already mentioned, the delay in reionization due to an overdensity $\delta$ is not a function of cosmic time. Thus we see from equation (10) that since the delay is also independent of collapsed fraction (which we expect to be different for hydrogen and helium reionization), the delay $\delta z$ in the redshift of He II reionization for a particular value of the comoving overdensity $\delta$ is equal to the delay for the reionization of hydrogen. As a result, in an overdense region both hydrogen and helium would be reionized early by the same offset in redshift. The large-scale variations in the reionization redshifts of hydrogen and helium lead to a different accretion histories for galaxies, which, in turn lead to different star formation histories, and thus a change in the luminosity of a galaxy given a total stellar mass due to the different age distribution of the stellar population.

Before proceeding, we draw attention to the approximation of sudden reionization in which the reionization of a volume on a scale $R$ occurs at a redshift $z$. Of course some regions within that volume will have been reionized earlier. However, our point is that, on average, a region of IGM will be reionized earlier by $\Delta z$ within a volume of scale $R$. A critical component of our model is the assumption that the average variation in the redshift at which the gas that ultimately makes the progenitor galaxies was reionized is also equal to $\Delta z$.

Weinmann et al. (2007) have recently employed numerical simulations of reionization to compute whether a galaxy observed at the present time formed in a region of IGM prior to it being reionized, or whether it formed in a region that had already been reionized. In their work the time of formation of a galaxy refers to the identification of the earliest progenitor of the local galaxy above the resolution limit of the simulation. These authors find that more massive galaxies had progenitors that formed in neutral regions while less massive galaxies formed in ionized regions. They also conclude that there is no correlation between the reionization history of field galaxies and their environment or large-scale clustering (however, see discussion below). While very useful in understanding the relation between the early formation histories of galaxies and the reionization process, these findings are not directly applicable to our discussion, which aims to calculate the average effect of reionization on all the progenitors of a low-redshift galaxy rather than on its earliest progenitor.

But interestingly, the numerical results presented by Weinmann et al. (2007) show consistency with the quantitative expectations of our simple model. Their fig. 6 shows the relation between the reionization redshift for a massive galaxy and the local overdensity of massive galaxies within 10 comoving Mpc. The simulations predict a large scatter of $\pm 1$ redshift unit about a mean relation that varies by $\Delta z \sim 0.8$ between present-day galaxy overdensities of $\sim 1$ and 1 on 10 comoving Mpc (cMpc) scales, with overdense environments reionizing at higher redshift. While the scatter is large as expected, the statistical accuracy of the measured mean is significantly below $\Delta z = 0.8$ due to the large sample size of model galaxies. This implies that Weinmann et al. (2007) do indeed infer a relation between large-scale environment and the mean reionization redshift, but that the variation in the mean relation is not significant with respect to the scatter among individual galaxies. We can compare the mean relation of Weinmann et al. (2007) for the reionization redshift of the earliest progenitor from simulations with expectations from our simple model for patchy reionization. On a scale of 10 cMpc, the variance in the density field at $z = 0$ is $\sigma(10 \text{cMpc}) \sim 0.8$. Since the bias is around $b = 1.1$ for the $5 \times 10^{12} M_{\odot}$ galaxies from which the simulated relation was calculated, we expect $1 \sigma$ fluctuations in the overdensity of galaxies at $z = 0$ on 10 cMpc scales to be $\pm 1$. Thus the numerical simulations of Weinmann et al. (2007) predict fluctuations in the reionization redshift around a mean of $z \sim 8.5$ (predicted by their model) of $\delta z \sim 0.4$ for the earliest progenitor of massive local galaxies. Our simple model predicts the fluctuation in the reionization redshift of the IGM with an overdensity of $\delta \sim 0.8$ to be $\delta z \sim (1 + z)(\delta D(\bar{z})) \delta_c \sim 0.6$, where $D$ is the growth factor (from equation 10). This number is similar to the typical fluctuations in the reionization redshift of the earliest progenitor in the simulations of Weinmann et al. (2007). Weinmann et al. (2007) also argue that the way in which a galaxy is reionized (either externally or internally) is not sensitive to the local overdensity of galaxies. Thus numerical simulations predict that the process of reionization is similar in overdense and underdense regions, but that reionization is accelerated in overdense regions. These findings from
numerical simulation support the basic assumptions of our simple model.

5.2 Model of reionization-induced bias

To develop our model we consider a galaxy residing in a halo of mass $M$ at $z \ll z_{\text{esc}}$. This galaxy has accreted its mass via a merger tree, which we generate using the method described in Volonteri, Haardt & Madau (2003). We describe this tree as having a number $N_{q0}(z_i)$ of haloes of mass $M_i(z_i)$ at redshift $z_i$, where the number of redshift steps is $N$, with values of redshift that increase from the redshift of the primary halo in the tree so that $z_0 = z$. These haloes grow in mass due to mergers of progenitor haloes, and due to accretion (which, in the Press–Schechter formalism, is the sum of mergers with haloes below the resolution limit of the merger tree).

First consider haloes above the minimum mass for star formation (which is either $M_{\text{esc}}$ in neutral regions, or $M_{\text{reion}}$ in reionized regions, respectively). At each redshift step, a fraction of the baryonic mass gained by these haloes through accretion is turned into stars, thus

$$\Delta M_{\star,i}(z_j) = f_{\star,i} \frac{\Omega_b}{\Omega_m} [M_i(z_j) - M_i(z_{j+1})] \quad \text{for} \ M > M_{\text{min}}$$

$$\Delta M_{\star,i}(z_j) = 0 \quad \text{otherwise},$$

(11)

where $f_{\star,i}$ is the star formation efficiency. We choose $f_{\star,i} = 0.3$ throughout this paper, though our conclusions are not sensitive to this choice.

In addition, we assume that whenever a progenitor halo $i$ at redshift $z_j$ in the merger tree crosses the minimum mass for star formation through the merger of two subunits $M_{\star,i}(z_{j+1})$ and $M_{\star,i+1}(z_{j+1})$, stellar mass is added in the amount

$$\Delta M_{\star,i}(z_j) = \left[ f_{\star,i} \frac{\Omega_b}{\Omega_m} M_i(z_j) - M_{\star,i+1}(z_{j+1}) \right],$$

(12)

where $M_{\star,i+1}(z_{j+1})$ and $M_{\star,i+1}(z_{j+1})$ are the stellar mass content of the progenitors prior to the merger. Similarly, whenever a progenitor halo $i$ at redshift $z_j$ in the merger tree crosses the minimum mass for star formation through accretion, stellar mass is added in the amount

$$\Delta M_{\star,i}(z_j) = f_{\star,i} \frac{\Omega_b}{\Omega_m} M_i(z_j) - M_{\star,i+1}(z_{j+1}),$$

(13)

where $M_{\star,i+1}(z_{j+1})$ is the stellar mass content of the halo at the previous redshift step. The subtraction of the second term is necessary in each of the latter cases because the minimum mass in a region increases suddenly at the local reionization epoch. The total stellar mass added at each step is the sum of these three contributions. We may then construct a stellar mass accretion history

$$d(M_{\star,i}(z_j)) = \sum_{j=0}^{z_i \rightarrow z_{j+1}} \Delta M_{\star,i}(z_j).$$

(14)

Our scheme neglects any star formation that may occur in recycled gas following a major merger. However, star formation in recycled gas at low redshift is by definition already above the minimum threshold for star formation and so should not be sensitive to the local redshift of reionization. We have used a sudden transition of the minimum virial temperature for star formation, rather than the more gradual transition described by the filtering mass, which takes account of the formation time-scale for a collapsing halo and the full thermal history of the IGM. In the hierarchical build-up of a halo, the merger of collapsed progenitors, combined with accretion of mass on to these progenitors can be identified with the formation of the final halo within the spherical collapse model and Press–Schechter formalism. The sudden transition of virial temperature is therefore the appropriate choice for our model since the star formation is calculated to coincide with the formation of a halo during a merger tree. The fraction of mass in the halo at redshift $z$ that was already collapsed prior to reionization is therefore explicitly accounted for in our model.

Before proceeding we note that the calculation of the merger tree is limited in the large-scale overdensity $\delta$. We demonstrate this explicitly in Appendix A, for a cosmology that ignores the cosmological constant (as is appropriate at high redshift). The calculation presented in Appendix A implies that within the extended Press–Schechter formalism the merger tree is not sensitive to large-scale overdensity. However, it is possible that the extended Press–Schechter formalism does not correctly account for all details of the dependence of the merger history on local overdensity. Such a claim would require comparison with detailed numerical simulations. We therefore briefly describe the ramifications if we suppose that the details of the merger history do indeed depend on the large-scale overdensity in a way not captured by our merger tree. It is clear that galaxies form earlier within overdense regions; this is the galaxy bias. Now suppose for the sake of argument that we have a fixed reionization redshift. The departure of a merger history from the predictions of the merger tree might lead to a particular galaxy at some redshift $z$ accumulating its mass earlier in an overdense region. If this were true, then galaxies in overdense regions would accumulate more mass prior to the reionization redshift, and so would appear to have an earlier mass-to-light ratio at redshift $z$. This earlier formation would counteract the reduced mass-to-light ratio from early reionization in overdense regions. However, the extended Press–Schechter formalism is a linear theory, and we would expect any departure from its predictions to be second order in $\delta$. In contrast the galaxy bias is a first-order correction and we therefore expect galaxy bias to dominate over departures from the merger tree prediction. Thus in the remainder of this paper we will assume that our merger tree formalism correctly captures the dependence of formation time on overdensity. In order to estimate the additional contribution to the bias that is due to changes in the star formation history associated with the reionization variable redshift, we may therefore compute one merger tree within the mean background cosmology, and then change only the reionization redshift to account for the overdensity of the region in which the parent halo formed. The total bias for the observation of this galaxy is then the reionization-induced bias computed from this merger tree, plus the usual galaxy bias. This independence of the merger tree on $\delta$ greatly simplifies calculation of the dependence of the apparent brightness of a galaxy within a halo of mass $M$ on the large-scale overdensity.

We next compute the spectrum of stellar light that results from this star formation history using the stellar population model of Leitherer et al. (1999). We assume a 1/20th solar metallicity population with a Scalo (1998) mass function and begin with the time-dependent spectrum for a burst of star formation. This yields the emitted energy per unit time per unit frequency per solar mass of stars $d^2E/\nu dm(t_0 - t_1)$ at a time $t_0 - t_1$ following the burst, where $t_0$ and $t_1$ are the ages of the universe at the redshift of the primary halo ($z_0$) and $z_1$, respectively. The flux (erg s$^{-1}$ cm$^{-2}$ Hz$^{-1}$) from the galaxy at $z$ then follows from the sum over the starburst associated with each star

\footnote{Model spectra of star-forming galaxies obtained from http://www.stsci.edu/science/starburst99l/}.

© 2007 The Authors. Journal compilation © 2007 RAS, MNRAS 382, 921–936
formation episode. We find

$$f_\nu = \sum_{j=1}^{N_{\text{trees}}} \Delta M_r(z_j) \frac{d^2 f_\nu}{dM} (t_j - t_0) \frac{1}{4\pi D_L(z_0)} (1 + z_0). \tag{15}$$

where $D_L$ is the luminosity distance at $z$. We note that our scheme neglects enrichment of gas prior to star formation, so that all starbursts are assumed to have the same metallicity. Since we are not computing the contribution from star formation in recycled gas, we do not expect this assumption to have a large influence on our results. Moreover, the ultraviolet (UV) spectra of galaxies are very sensitive to the dust content of a galaxy and therefore also sensitive to the metallicity. However, the reionization bias would be sensitive to the ratio of fluxes in which the effect of the dust on a single spectrum will vanish.

We can then use equation (15) to compute the spectra for two star formation histories that correspond to the mean universe and to an overdensity $\delta$, with reionization redshifts separated by $\delta z$. These spectra in turn allow us to determine ratio of fluxes and hence to compute a value of $\mu(\delta)$. As described above, in order to compute the contribution of reionization to galaxy bias we make this comparison directly, using the same merger tree with different values of the reionization redshift to obtain different star formation histories.

Equation (15) implies that the apparent flux of a galaxy will be sensitive to its star formation history, which will in turn be sensitive to the redshift of reionization. We would now like to calculate the typical flux change induced by the effect of a large-scale overdensity on the reionization redshift. In order to achieve this we must determine the change in flux as a function of scale, and at a range of overdensities. As shown above, the delay in the reionization redshift is proportional to the variance of the power spectrum on the scale of interest. Since the variance decreases towards large scales, we find that the fluctuations in reionization redshift should also be smaller on large scales. In order to investigate the scale dependence of the bias, we could therefore compute the difference in star formation histories corresponding to different delays in reionization over a number of spatial scales. However, we find that for individual merger trees, the value of the apparent magnitude change (i.e. the logarithm of the observed flux) is approximately proportional to the delay in reionization. Thus we can estimate the change in magnitude due to reionization using a first-order expansion in $\delta z$

$$2.5 \log_{10} \mu = 2.5 \frac{d \log \left[ f_\nu(z_{ol}) \right]}{dz_{ol}} \delta z \sim 2.5 \log_{10} \left[ \frac{f_\nu(z_{ol})}{f_\nu(z_{ol} + \Delta z)} \right] \frac{\delta z}{\Delta z}. \tag{16}$$

where $z_{ol}$ is the overlap (reionization) redshift, which we assume to be $z = 6$ throughout this paper (e.g. White et al. 2003; Fan et al. 2006), and $\Delta z = 0.25$ is the separation in overlap redshifts of the two star formation histories computed for each merger tree. This is an approximation which greatly simplifies our calculations.

Given a scale $R$ and variance $\sigma(R)$ we can now estimate the contribution of reionization to galaxy bias. For each merger tree $k$, we compute the bias averaged over likelihoods at each $\delta$ in the density field

$$b_{\text{reion},k} = \frac{1}{\sqrt{2\pi} \sigma(R)} \int d\delta \left[ \frac{1 - \Lambda(\delta)}{\delta} \right] \exp \left[ -\frac{\delta^2}{2\sigma^2(R)} \right]. \tag{17}$$

To get the average bias for the galaxy population, we then average the bias evaluated using $N_{\text{trees}}$ different merger trees,

$$b_{\text{reion}} = \frac{1}{N_{\text{trees}}} \sum_{k=1}^{N_{\text{trees}}} b_{\text{reion},k}. \tag{18}$$

In the models presented below we find that the variance in values of $b_{\text{reion},k}$ among the different merger trees is of the order of the mean value of the bias $b_{\text{reion}}$. In the remainder of this paper, we use the above model to estimate the contribution of reionization to galaxy bias for several existing and planned galaxy surveys.

### 6 Lyman-Break Galaxies

As a first application of our model we construct mock spectra corresponding to LBGs (Steidel et al. 2003) at $z = 3$. Examples of the model star formation history and resulting galaxy spectra are shown in Fig. 1 assuming a halo mass of $M = 10^{12} M_\odot$ (Adelberger et al. 2005). Here we include only hydrogen reionization as the mechanism that introduces fluctuations in the star formation history.

In the upper left-hand panel of Fig. 1 we show an example of the star formation rate summed over all haloes that end up as part of the galaxy in a $10^{12} M_\odot$ halo at $z = 3$. Here the solid and dotted lines refer to histories where the overlap of ionized regions occurred at $z = 6.125$ and 5.875, respectively, with a redshift interval of $\delta z$. The effect of the reionization redshift on these star formation histories is more easily seen in the central left-hand panel where we plot the difference in star formation rate between the two histories.

Fig. 1 shows that early reionization initially results in a deficit of star formation. This deficit is then made up continually until $z = 3$. By $z = 3$ the total stellar mass is the same for both histories ($M_\star = f_\nu \Delta_\nu / \Omega_\mu M$) as required for consistency of the model. This behaviour is also demonstrated in the lower left-hand panel of Fig. 1 where we plot the difference in cumulative stellar mass, summing over all haloes that end up as part of the primary galaxy at $z = 3$. This figure shows that the accretion history for stellar mass is effected by the redshift of reionization until times long after the reionization redshift. This is because in the case of an early reionization, a halo which would have formed stars when crossing the cooling threshold during $\delta z$ will now not form stars until it crosses the larger threshold for star formation in an ionized IGM. This may not occur until a substantially later time. The lower left-hand panel of Fig. 1 shows that the amount of mass accreted during the $0.25$ redshift interval assumed for this example is $\sim 10^8 M_\odot$, with a scatter of around a factor of 2. The difference in the accretion rate during this time therefore represents about $10^{-4}$ of the total stellar mass at $z = 3$.

Fig. 1 also shows examples of the resulting galaxy spectrum. In the upper right-hand panel we show the rest-frame luminosity for 10 realizations of $10^{12} M_\odot$ galaxies at $z = 3$. In each case reionization was at $z = 6.125$. The differences between these spectra arise due to the slightly different age distribution of the stellar populations that result from the stochastic build-up of mass in the merger tree. In the central right-hand panel of Fig. 1 we show the corresponding observed flux [including mean Ly$\alpha$ absorption, see e.g. Fan et al. (2006)] for the same 10 LBGs at $z = 3$. The distinctive Lyman-break near 4000 Å is clearly visible in these spectra. Finally, in the lower right-hand panel we show the fractional change ($\delta b_{\nu}$) in the observed flux that is induced by a delay in reionization of $\Delta z = 0.25$ units of redshift. This flux change, which is related to the parameter $\mu$ through equation (16) corresponds to differences in the star formation histories that are comparable to the example shown in the left-hand panels of Fig. 1. The fluctuations are at the
Reionization and galaxy bias

Figure 1. The effect of reionization on the star formation histories of galaxies. These examples correspond to typical LBGs at \( z = 3 \) in the survey of Steidel et al. (2003). Upper left-hand panel: The star formation rate summing over all haloes that end up as part of the galaxy in a \( 10^{12} \) \( M_\odot \) halo at \( z = 3 \). The solid and dotted lines refer to histories where overlap occurred at \( z = 6.125 \) and 5.875, respectively. Central left-hand panel: The difference in star formation rate for the two histories. Lower left-hand panel: The difference in cumulative stellar mass for the two histories. The example shown in the upper and central left-hand panels is shown as the dark line. The light lines show histories from a range of merger trees. Upper right-hand panel: Rest-frame luminosity of 10 realizations of LBGs with a \( 10^{12} \) \( M_\odot \) halo at \( z = 3 \). Central right-hand panel: Observed flux (including Ly\( \alpha \) absorption) for the 10 exemplary galaxies. Lower right-hand panel: The magnitude change induced by a delay in reionization of 0.25 units of redshift.

level of a few tenths to a few per cent. We investigate the bias that arises from the resulting values of \( \mu \) in Section 6.2.

6.1 The colours of simulated Lyman-break galaxies

Our aim in this paper is to evaluate the importance of reionization with respect to galaxy bias in measurements of galaxy clustering. To this end we have constructed model star formation histories that include the effect of reionization, and computed the effect of reionization on the corresponding model galaxy spectra. In order for our results to be applicable to surveys of real galaxies, we must, at a minimum demonstrate that our model produces realistic spectra with colours that would see the model galaxies selected into the survey of interest. Therefore, before describing our results for the reionization-induced bias we demonstrate that our model galaxies have colours and magnitudes that correspond to those of real LBGs.

In the upper left-hand panel of Fig. 2 we show the position of 100 model LBGs within the primary colour selection\(^3\) (Steidel et al. 2003) for LBGs at \( z \sim 1.7 \) (solid points), \( z \sim 2.3 \) (open points) and \( z \sim 3 \) (crosses). The galaxies at \( z \sim 3 \) are well separated from those at lower redshift due to the Lyman break moving to a wavelength beyond the \( U_n \) band. Our model galaxies at \( z \sim 1.7 \) and 2.3 have similar colours that are close to the selection cut-off. This is consistent with the observed galaxies, which have overlapping redshift distributions when selected via this criteria (Adelberger et al. 2005). For their clustering analysis Adelberger et al. (2005) restricted themselves to objects with \( 23.5 < R < 25.5 \). In the

---

\(^3\)To estimate the colours of LBGs we assume top-hat filters of central wavelength \( \lambda_0 \) and width \( \Delta \lambda \) to approximate the filter set used in Steidel et al. (2003). We use \( AB \) magnitudes throughout this paper. The filters have \((\lambda_0, \Delta \lambda) = (3550, 600)\) for the \( U_n \) band; \((\lambda_0, \Delta \lambda) = (4780, 1100)\) for the \( G \) band; \((\lambda_0, \Delta \lambda) = (6830, 1250)\) for the \( R \) band; \((\lambda_0, \Delta \lambda) = (8100, 1650)\) for the \( I \) band.
Figure 2. Examples of clustering bias in LBGs induced by reionization. Upper left-hand panel: The primary colour selection (Steidel et al. 2003) for LBGs at $z \sim 1.7$ (solid points), $z \sim 2.3$ (open points) and $z \sim 3$ (crosses). Central left-hand panel: The apparent magnitudes and colours of the model galaxies. Lower left-hand panel: The $U_n - G$ colour as a function of the change in $R$-band magnitude induced by reionization. Upper right-hand panel: The bias introduced by reionization in cases where helium reionization at $z \sim 3.5$ is considered in addition to hydrogen reionization (open squares) and where it is not (solid squares). The bias was computed assuming a flux evaluated at a rest-frame wavelength of 1350 Å within a 400-Å window. The galaxy bias is shown by the solid line for comparison. The error bars represent the statistical noise in the simulations due to the finite number of merger trees. Central right-hand panel: The ratio of the component of bias introduced through reionization to the usual galaxy bias. Lower right-hand panel: The factor by which the mass will be overestimated in clustering analyses where reionization is not considered.

central left-hand panel we show the position of our model galaxies in a colour–magnitude diagram. The apparent magnitudes of these model galaxies are consistent with the observed population. Thus our model produces LBGs with both the correct colours, and the correct luminosity. Finally we check that reionization-induced changes in the observed flux are not sensitive to the observed galaxy colour. In the lower left-hand panel of Fig. 2 we show the $U_n - G$ colour as a function of the change in $R$ magnitude induced by reionization. We find no systematic trend of the flux variation with galaxy colour.

6.2 Reionization-induced bias for Lyman-break galaxies

We now present an estimate for the reionization-induced bias in the sample of LBGs. We show results at a scale of $R = 10$ comoving Mpc, which is a factor of $\sim 3$ larger than the clustering length at $z \sim 3$ (Adelberger et al. 2005). We evaluate the bias for fluxes measured at a rest-frame wavelength of 1350 Å, and within a 400 Å wide band (this choice allows us to compare the predicted bias over the redshift range $1.7 < z < 5.5$). In the upper right-hand panel of Fig. 2 we show the bias introduced by reionization in cases where hydrogen reionization alone is considered (solid squares), as well as cases where helium reionization at $z \sim 3.5$ is considered in addition to hydrogen reionization (open squares). Also shown for comparison is the galaxy bias due to enhanced structure formation (solid line). In this figure the error bars represent the statistical noise in the simulations due to the finite number of merger trees. In order to better see the relative contributions of enhanced structure formation and reionization-induced galaxy bias, in the central right-hand panel we show the ratio of the component of bias introduced through reionization to the usual galaxy bias. Reionization represents a 10–20 per cent correction to the galaxy bias in LBG samples at $1.7 < z < 3$. This correction corresponds to a predicted amplitude for the galaxy correlation function that can be 50 per cent larger than the prediction in the absence of consideration of reionization. Thus reionization provides a correction to the clustering amplitude that is in excess of the observational error for the existing LBG samples at $1.7 < z < 3$. 

© 2007 The Authors. Journal compilation © 2007 RAS, MNRAS 382, 921–936
One of the primary uses for measurements of clustering in a galaxy sample is the estimation of host halo mass. This mass estimate is made by measuring the bias, which is then interpreted theoretically in terms of host mass. However, the results summarized in Fig. 2 suggest that existing estimates of the galaxy bias could be systematically in error, at a level significantly larger than the observational error, due to the neglect of the effect of reionization. This in turn implies that estimates of the host masses in galaxy samples are also systematically in error. To evaluate the importance of this systematic error, we estimate the ratio of the inferred host masses with and without the inclusion of reionization, yielding

\[
\ln(M_{\text{reion+gal}}) - \ln(M_{\text{gal}}) \approx \frac{d \ln(M)}{d b} [(b_g + b_{\text{reion}}) - b_g],
\]

or

\[
\frac{M_{\text{reion+gal}}}{M_{\text{gal}}} \approx \exp \left( b_{\text{reion}} \frac{d \ln M}{d b} \right),
\]

where \(d \ln(M)/db\) is evaluated via equation (5). The factor by which the mass will be overestimated in clustering analyses where reionization is not considered is plotted in the lower right-hand panel of Fig. 2. We find that masses in existing LBG surveys (Adelberger et al. 2005) have been overestimated by factors of between 1.5 and 2.

In addition to showing results evaluated at redshifts corresponding to the LBG sample, we also show results for hypothetical galaxy samples \(4 \leq z \leq 5.5\). At these redshifts helium is not doubly reionized, and so all modifications to the star formation history are due to the reionization of hydrogen at \(z \sim 6\). This is in contrast to samples at \(z \lesssim 3\) where helium reionization is complete. Fig. 2 demonstrates that we can see a significant jump in the amplitude of the clustering for galaxy samples of fixed absolute magnitude following the double reionization of helium at \(z \sim 3.5\).

Up to this point we have presented results for \(R = 10\) comoving Mpc. We do not explicitly show results corresponding to other length-scales in this paper for the following reason. On the scales of interest (~3–100 cMpc), the variance is approximately a power law with \(R\), while the mass function is approximately a power law with \(\log(M)\). It turns out that these power laws approximately cancel, leaving the bias induced by reionization almost independent of scale. This independence is a coincidence. A different slope of the primordial power spectrum would have led to a scale-dependent bias. However, the conclusion that the bias is not scale dependent should be treated with caution for two reasons. First, we are unable to rule out scale dependence at the level of a few per cent at the numerical accuracy of our simulations. Secondly, as discussed in Section 8.2, the scale independence might be broken by additional astrophysical effects. Thus, future observations aiming to measure clustering at the per cent level over a large range of spatial scales will need to carefully account for this possibility.

### 6.3 Mass dependence of the reionization bias

Thus far our discussion of LBGs has assumed a halo mass of \(10^{12} \text{M}_\odot\), corresponding to observed Lyman-break galaxies. In this section we describe the dependence of the predicted reionization-induced galaxy bias on the host halo mass, and on the assumed values of \(T_{\text{min}}\). In the upper and central rows of Fig. 3 we show examples of the clustering bias in LBGs induced by reionization for halo masses of \(M = 10^{11}\) and \(10^{13} \text{M}_\odot\). In the left-hand panels we show the bias introduced by reionization in cases where helium reionization at \(z \sim 3.5\) is considered in addition to hydrogen reionization (open squares) and where it is not (solid squares). As before the bias was computed assuming a flux evaluated at a rest-frame wavelength of 1350 Å within a 400-Å window. The usual galaxy bias is shown by the solid line for comparison. In addition, in each case the corresponding results for \(M = 10^{12} \text{M}_\odot\), as presented in Fig. 2, are shown for comparison (light lines).

We find that the contribution to the bias due to reionization is fairly insensitive to the halo mass. To understand this we note that although we would expect the larger haloes to have begun forming earlier, and so to have their star formation histories less effected by the reionization of the IGM, this is offset by the steeper mass function of massive haloes. On the other hand the galaxy bias due to enhanced structure formation in over dense regions is quite sensitive to the halo mass, and so we find that the fractional contribution to the galaxy bias is smaller for more massive systems. As a result, the systematic error introduced into the estimate of halo mass from clustering amplitude is less serious for more massive systems. In the right-hand panels of Fig. 3 we show the factors by which the mass will be overestimated in clustering analyses where reionization is not considered. While haloes with masses near \(M \sim 10^{11}\) would be incorrectly inferred by a factor that could be larger than 3, the systematic error on very massive systems of \(M \sim 10^{13} \text{M}_\odot\) would be at most a level of tens of per cent. This implies that the reionization bias will become more important as future surveys begin to discover populations of less massive galaxies at high redshift.

Our calculations presented thus far have assumed values for the minimum virial temperature of haloes that host star formation to be \(T_{\text{min}} = 10^4\) and \(10^5\) K within neutral and ionized regions of the IGM, respectively. Before concluding this subsection we describe the effect of these assumed values on predictions for the reionization-induced bias. As an alternate example we have repeated the clustering estimates for galaxies of mass \(10^{12} \text{M}_\odot\), but with values of \(T_{\text{min}} = 2 \times 10^4\) and \(0.8 \times 10^5\) K. This jump in virial temperature between neutral and ionized regions of the IGM is therefore 2.5 times smaller than the corresponding jump in our fiducial model. The results are presented in the lower panels of Fig. 3. As expected the smaller jump in virial temperature slightly lowers the size of the reionization-induced bias, and hence reduces the systematic error introduced into the estimate of halo mass from clustering amplitude. However, conclusions regarding the effect of reionization on galaxy bias are not qualitatively sensitive to the model values for \(T_{\text{min}}\).

### 6.4 Reionization and the observed colours of Lyman-break galaxies

The reionization-induced bias should be sensitive to the selection band. In the case of LBGs, we would therefore expect that the clustering amplitude would be sensitive to the band in which the flux selection was performed. Alternatively, overdense regions would therefore be expected to have a slightly bluer population of galaxies.

In Fig. 4 we show the change in \(U_p - G\) colour for LBGs at \(z \sim 1.7\) (solid points), \(z \sim 2.3\) (open points) and \(z \sim 3\) (crosses), due to a fluctuation in the reionization redshift of \(\delta z = 0.25\). Galaxies in overdense regions, have systematically bluer colours due to their younger stellar populations. The example shown for LBGs has fluctuations in \(\Delta(U_p - G)\) colour at the ~0.01–0.02 level given a fluctuation in the redshift of overlap amounting to \(\delta z = 0.25\) redshift units. On a scale of 10 comoving Mpc, the fluctuation in the overlap redshift around a mean of \(z = 6\) is \((\delta z \delta^2)^{1/2} \sim 0.6\) (from equation 10). Hence the expected colour variation between overdense and underdense regions would be \(\Delta(U_p - G) \sim 0.03–0.05\) mag. This expected correlation between galaxy colour and
overdensity would be evidence for the reionization-induced galaxy bias, and could be used to calibrate its effect empirically.

This systematic variation in colour is much smaller than the range of colours in the observed samples. However, high-redshift samples are selected to be redder than a certain limit. In practice one would therefore have to be careful that the systematically bluer colours did not bias the sample against finding galaxies in overdense regions. At the redshifts of LBGs, the shift of the Lyman break with redshift primarily effects the $U_n - G$ colour. On the other hand reddening effects both the $U_n - G$ and $G - R$ colours. As a result, LBGs are selected to lie above a line with positive gradient in the ($U_n - G$) – ($G - R$) colour–colour space. We note that like reddening, the reionization-induced colour change will be in both bands, and will therefore transform the position of the galaxy in colour–colour space in a direction parallel to the selection criteria for LBGs. As a result, we do not expect the reionization-induced colour change to introduce a bias through the survey selection criteria.

7 SURVEYS FOR BARYONIC OSCILLATIONS

We next apply our model to surveys that aim to measure baryonic acoustic oscillations in the clustering of galaxies at $z < 1$. These surveys require exquisite accuracy of the clustering amplitude, and so the effect of reionization on galaxy bias could be particularly important. We consider two surveys, the existing Sloan Digital Sky Survey (SDSS) Luminous Red Galaxy (LRG) survey, and the planned WiggleZ survey.

7.1 Luminous red galaxies

First, we discuss the effect of reionization on the star formation histories of SDSS LRG at $z = 0.3$ (Eisenstein et al. 2001). By selection, LRGs are old galaxies with passively evolving stellar populations and no recent star formation. Thus, in order to model LRGs at $z \sim 0.3$ we arbitrarily shut off star formation in the galaxies
that while the correction to the galaxy bias due to reionization predicted by our models is at a very low level for the LRG sample, it may nevertheless be comparable to the largest correction to linear theory yet described on the scales relevant to baryonic oscillations experiments. On the other hand, our model predicts no dependence of the reionization-induced bias on scale. As a result it is very unlikely that the details of the reionization will adversely affect attempts to use the measurements of baryonic acoustic oscillations as a cosmic standard ruler (Blake & Glazebrook 2003). We return to this point in Section 8.2.

7.2 Blue star-forming galaxies

For our second example we consider the effect of reionization on the star formation histories of galaxies that will be selected by the WiggleZ survey (Glazebrook et al. 2007) at z = 0.8. Unlike the SDSS LRG sample considered in the previous section, galaxies in the WiggleZ survey will be selected as being star-forming via the Lyman break using observations in the near and far UV in addition to optical colours. In modelling these galaxies we therefore do not impose a cut-off in the star formation prior to the observed redshift.

The upper and lower left-hand panels of Fig. 6 show the primary colour selection4 (Eisenstein et al. 2001) for LRGs at z ~ 0.3. The model produces galaxies with the correct colours and observed flux, as is illustrated by the magnitudes and colours of the 100 modelled galaxies which are also plotted in the left-hand panels of Fig. 5. In the upper right-hand panel of Fig. 5 we show examples of observed flux (including Lyα absorption) for 10 model LRGs at z = 0.3. These spectra show less variation than those of the LBGs discussed in the previous section. This lack of variation is a feature of the LRG sample. Due to the lack of star formation in these galaxies, the spectra do not exhibit a sharp Lyman break. In the lower right-hand panel of Fig. 5 we show the fractional change in flux induced by a delay in reionization of 0.25 units of redshift. We see that reionization has a very small effect on the observed flux of these galaxies. The resulting value of bias is quoted in the upper right-hand panel. LRGs are selected to lie within a range of rest-frame absolute r magnitudes, and we therefore calculate the bias at the rest-frame r magnitude. Reionization will decrease the bias by ~0.1 per cent in the LRG sample, and therefore the clustering amplitude by ~0.2 per cent. Also shown is the fractional systematic error in the derived host mass (~1 per cent).

Eisenstein et al. (2005) summarize the various corrections to the linear bias that have been previously considered when interpreting clustering data, including those due to non-linear gravity, and coupling of gravitational modes. On scales larger than ~40 comoving Mpc, the sum of previously considered corrections drops below the 1 per cent level. For this reason among others, the correlation function of galaxies is considered to be a very clean tracer of the underlying large-scale mass distribution, and in particular a perfect sample with which to investigate the baryonic oscillations in the matter power spectrum (Eisenstein et al. 2005). It is therefore important to note

---

4 To estimate the colours of LRGs in this paper, we assume top-hat filters of central wavelength λ₀ and width ∆λ to approximate the SDSS filter set. The filters have (λ₀, ∆λ) = (3543, 564) for the u band; (λ₀, ∆λ) = (4770, 1388) for the g band; (λ₀, ∆λ) = (6251, 1372) for the r band; (λ₀, ∆λ) = (7625, 3524) for the i band.

5 To estimate the UV colours in this paper, we assume top-hat filters of central wavelength λ₀ and width ∆λ to approximate the Galex filter set. The filters assumed have (λ₀, ∆λ) = (1550, 400) for the FUV band; and (λ₀, ∆λ) = (2500, 1000) for the NUV band.
Figure 5. The effect of reionization on the star formation histories of galaxies. These examples correspond to typical SDSS LRG at $z = 0.3$. Upper and lower left-hand panels: The primary colour selection (Eisenstein et al. 2001) for LRGs, together with the locations of our model galaxies. Upper right-hand panel: Observed flux (including Ly$\alpha$ absorption) for 10 example LRGs at $z = 0.3$. Lower right-hand panel: The magnitude change induced by a delay in reionization of 0.25 units of redshift. The corrections to the bias due to reionization are quoted in the upper right-hand panel. We compute this bias using the rest-frame $r$ magnitude.

Figure 6. The effect of reionization on the star formation histories of galaxies that will be selected by the WiggleZ survey (Glazebrook et al. 2007) at $z = 0.8$. Upper and lower left-hand panels: The primary colour selection (Glazebrook et al. 2007) for WiggleZ star-forming galaxies at $z \sim 0.8$, together with the points for our model galaxies. Upper right-hand panel: Observed flux (including Ly$\alpha$ absorption) for 10 example WiggleZ galaxies at $z = 0.8$. Lower right-hand panel: The magnitude change induced by a delay in reionization of 0.25 units of redshift. The corrections to the bias due to reionization are quoted in the upper right-hand panel. This bias is computed at the observed $r$-band wavelength.

$\sim 5$ per cent to the galaxy bias due to reionization is the largest correction to linear theory yet described on the scales relevant to baryonic oscillation experiments (Eisenstein et al. 2005). On the other hand, as mentioned earlier, our simple model predicts that the bias due to reionization is, like the linear bias due to enhanced formation in overdense regions, independent of scale. Thus in an analysis that ignores reionization, the host mass would be misidentified, but because the correction to the linear bias is not scale dependent, the unknown details of the reionization history should not compromise the measurement the baryonic acoustic peak.
8 DISCUSSION

Before concluding we discuss several issues which arise from our results and which will provide interesting areas for future research.

8.1 Implications for the evolution of clustering in galaxy samples

The observed spatial correlation function of galaxies can be used to estimate the mass of the host dark matter halo population through comparison with theoretical calculations. Having determined this mass, the evolution in the clustering of these galaxies can then also be computed and compared with the clustering properties of different populations at later times, with the aim of piecing together the evolution of the galaxy population. Moreover having estimated the host halo mass, the predicted number density of hosts can be compared with the observed number density of objects in order to obtain the fraction of haloes containing a galaxy of the selected type at any one time. By comparing the inferred mass of LGs from clustering data to the observed number counts, Adelberger et al. (2005) concluded that star formation in LGs has a duty cycle approaching unity. This conclusion is consistent with our star formation model in which nearly all model galaxies satisfy the LG colour selection criteria.

In this subsection we consider the interpretation of LG clustering evolution in light of the additional contribution to the observed galaxy bias from reionization. The spatial correlation function of dark matter haloes as a function of radius $r$ can be written in terms of the correlation function of dark matter and the halo bias $b$ as

$$\xi_h(r) = \xi_c b^2(M).$$  \hfill (21)

In practice this correlation function can be approximated using the parameterization

$$\xi_h \approx \left(\frac{r}{r_0}\right)^{-\gamma},$$  \hfill (22)

where $r_0$ is defined as the clustering length, and $\gamma \sim 1.5$ describes the observed clustering of galaxies. More biased samples have larger clustering lengths.

We have argued that reionization will increase the observed value of the bias, by causing galaxies in overdense regions to have lower mass-to-light ratios due to their younger stellar populations. Thus we also expect reionization to increase the observed clustering length of a sample of galaxies at fixed halo mass. As a result, neglect of reionization leads to overestimation of the true clustering length for host dark matter haloes. For small values of $b_{\text{reion}}/b_{\text{gal}}$, equation (22) may be used to estimate the contribution to the observed clustering length ($\Delta r_{0,\text{reion}}$) that results from reionization-induced bias using the expression

$$\Delta r_{0,\text{reion}} \approx r_0 \frac{2 b_{\text{reion}}}{b_{\text{gal}}} \approx 1.25 r_0 \frac{b_{\text{reion}}}{b_{\text{gal}}} \frac{1}{0.2},$$  \hfill (23)

where the units of length-scales in the latter equality are comoving Mpc.

The clustering evolution of LGs was discussed by Adelberger et al. (2005). They measure the clustering length of LGs at $z \approx 1.7$, $z \approx 2.3$ and $z \approx 3$, obtaining $r_0 = 5.7$, $r_0 = 6.0$ and $r_0 = 6.4$ comoving Mpc, respectively, corresponding to halo masses of $10^{12} \pm 0.2$, $10^{12} \pm 0.3$ and $10^{11.5} \pm 0.3$ M$_{\odot}$. Using simulations, Adelberger et al. (2005) calculated the clustering lengths that these galaxies should have at lower redshifts of $z \sim 1$ and 0.2, and then compared these evolved clustering lengths to clustering studies of various populations of galaxies from other surveys. In particular, Adelberger et al. (2005) compared the evolved clustering length for LGs to galaxies in the DEEP survey (Coil et al. 2004), and in the SDSS (Budavári et al. 2003). Adelberger et al. (2005) find that the LG clustering length should evolve to a value that is consistent with redder elliptical galaxies ($r_0 \approx 9.4$ comoving Mpc at both $z = 1$ and 0.2), but which is larger than the clustering length for both the whole DEEP galaxy sample at $z \sim 1$ ($r_0 \approx 4.6$ comoving Mpc) and the blue SDSS galaxies at $z = 0.2$ ($r_0 = 6.4$ comoving Mpc).

Based on these results Adelberger et al. (2005) argued that the descendents of LGs will have clustering strengths that are significantly in excess of typical galaxies in optical magnitude-limited surveys at low redshift, and therefore that LGs must have stopped forming stars before $z \sim 1$. However, the results of this paper show that the clustering length at $z \sim 3$ has been overestimated by $\Delta r_{0,\text{reion}} \sim 1.5$ comoving Mpc. Since the reionization-induced bias decreases in influence towards low redshift, and is small below $z = 1$ (see following sections) we conclude that, after accounting for the reionization-induced bias, the clustering of the hosts of LGs may well be comparable to the blue population of galaxies at $z < 1$. Indeed, as shown in fig. 13 of Adelberger et al. (2005), the value of $\Delta r_{0,\text{reion}}$ computed for LGs at $z \sim 3$ is comparable to the difference in the clustering length of normal ellipticals and normal blue galaxies in the SDSS at $z = 0.2$. Thus the effect of reionization on the observed clustering of galaxies should be accounted for in studies that aim to link galaxies at a range of epochs through the evolution of their clustering properties.

8.2 Helium reionization and scale-dependent bias

The previous section (Section 7) ended with the positive suggestion that reionization will not impact measurement of the baryonic acoustic peak in samples of moderate redshift star-forming galaxies, due to the independence of the reionization-induced bias on scale. Before concluding this paper, we describe an additional astrophysical situation which may compromise this favourable conclusion.

Our simple model predicts the bias introduced through reionization to be independent of scale. However, this model ignores several astrophysical effects that could introduce additional fluctuations in temperature within the IGM, and hence also introduce additional dependencies of the star formation history large-scale overdensity. These additional fluctuations might include a scale dependence that is different to that of cosmic variance, and could therefore introduce a scale-dependent component of reionization-induced galaxy bias.

For example, consider the epoch after He II overlap. At that time the mean free path for He II ionizing photons is limited by abundance and cross-section of Lyman-limit systems, since the diffuse He II has previously been ionized. During this epoch, heating of the IGM will be sourced by recombinations of He II ions. Now the recombination time is an order of magnitude longer than the Hubble time at the mean IGM density. However, in the overdense regions containing filaments and sheets the recombination time would be shorter, and could approach a Hubble time in high-density regions. As a result, while regions of low overdensity would cool adiabatically by the cosmic expansion, heating due to photoionization of He II could be substantial in the overdense regions. This effect would introduce temperature fluctuations inside overdense regions of the IGM on scales larger than the mean free path.

Thus, it is possible that the ionizing photon mean free path introduces a length-scale below which reionization-induced bias is
independent of scale, but above which the reionization-induced bias is scale dependent. At $z \sim 3$ the ionizing photon mean free path is $\sim 100$ comoving Mpc (e.g. Bolton et al. 2006). This scale is uncomfortably close to the scale of the baryonic acoustic peak, implying that careful account may need to be taken of reionization-induced bias in galaxy surveys that select star-forming galaxies. A proper analysis of this possibility would require full numerical modelling and is beyond the scope of the present paper.

8.3 Improvements to the model

In the future, our simple model could be improved in several ways. Our model predicts that galaxies in regions that are reionized earlier form a larger fraction of their stellar mass at later times, implying that these galaxies form a greater fraction of their stellar mass in more massive haloes. We have assumed that the star formation efficiency is independent of halo mass. However, if high-redshift galaxies are subject to mass-dependent feedback effects (such as supernova feedback), then the star formation history would be altered. The presence of feedback in low-mass haloes would result in a larger fraction of the final stellar mass being formed after reionization, and hence in an increase in the sensitivity of the final mass-to-light ratio to the local reionization redshift. In addition, one could incorporate metal enrichment of the star formation. Stars forming in galaxies within overdense regions where reionization occurs early have their star formation, and hence their metal enrichment, delayed. As we have discussed, the resulting stellar populations are therefore observed to be younger at a low redshift $z$. Since there is a delay in the enrichment of the IGM following a burst of star formation, the younger stellar populations will have slightly lower metallicity. Thus the metallicity of populations observed at $z$ should be slightly dependent on the reionization history. We expect that the lower metallicity in the younger populations would tend to reduce the magnitude of the reionization-induced bias, since more highly enriched populations are bluer, and have lower mass-to-light ratios (though this would be a second-order effect). On the other hand, since we have not included metal enrichment, our model also underestimates the variation between the UV fluxes of stellar populations with different ages, and hence also underestimates the contribution of reionization to the galaxy bias. In addition to metallicities, one might also attempt to include the effects of dust, which leads to larger extinction in younger galaxies (Shapley et al. 2001). This would redder the spectra of galaxies in regions that were reionized at earlier times.

9 CONCLUSIONS

We have developed a model to estimate the effect of reionization on the clustering properties of galaxy samples at intermediate redshifts. Current models of the reionization of the IGM predict that overdense regions will be reionized early due to the presence of galaxy bias. The IGM in these regions is heated through the absorption of the ionizing radiation. The heating leads to an increased Jeans mass, and so reionization suppresses the formation of low-mass galaxies. The suppression of low-mass galaxy formation in turn delays the build-up of stellar mass in the progenitors of massive low-redshift galaxies. As a result of this delayed build-up, the stellar populations observed in galaxies at later times are on average slightly younger in overdense large-scale regions of the Universe. Stellar populations fade as they age and so the resulting age difference would lead to a lower mass-to-light ratio for galaxies in overdense regions. In volume-limited surveys, such as those now being employed for large-scale clustering studies, a fixed observed flux threshold therefore contains lower mass galaxies (on average) in overdense regions with a corresponding increase in the galaxy number density.

We have parametrized the reionization-induced increase of the observed galaxy density in overdense regions in analogy with the traditional galaxy bias. Our modelling uses merger trees combined with a stellar synthesis code. We have used this model to demonstrate that reionization can have a significant and detectable effect on the clustering properties of galaxy samples that are selected based on their star formation activity.

In existing samples of LBGs, the bias correction for reionization is at the level of 10–20 per cent, leading to correction factors between 1.5 and 2 in the mass inferred from clustering amplitudes. This effect is present in existing samples of LBGs at $1 \lesssim z \lesssim 3$ (Steidel et al. 2003), and provides a systematic correction to existing analyses that is in excess of the statistical errors (Adelberger et al. 2005). For example, the reionization-induced bias qualitatively changes the conclusion of Adelberger et al. (2005) that LBGs stop forming stars at $z \gtrsim 1$, and evolve into red elliptical galaxies by $z \sim 2$. Rather, allowing for reionization-induced bias implies that LBGs could evolve into the blue populations observed at low redshift with clustering lengths that are smaller than the massive red galaxy population.

The reionization of helium, and the associated additional heating of the IGM may lead to a sharp increase in the amplitude of the correlation function of $\sim 50$ per cent for galaxies at fixed luminosity in the redshift range $3 \lesssim z \lesssim 4$. Our model predicts that the reionization introduced bias is approximately independent of scale. However, we are unable to rule out scale dependence at the level of a few per cent due to the limited numerical accuracy of our calculations. Further astrophysical complexities not addressed in our model could alter this conclusion. Future experiments aimed at measuring galaxy clustering at a precision level of a few per cent over a large range of spatial scales (with a goal of constraining the initial conditions from inflation or the nature of dark matter and dark energy), will need to carefully account for this possibility.

We find that the contribution to the bias due to reionization is fairly insensitive to halo mass. This is in contrast to the galaxy bias from enhanced structure formation in overdense regions which is a function of halo mass. Hence the fractional contribution to the galaxy bias by reionization is smaller for more massive systems, and as a result the systematic error introduced into the estimate of halo mass from clustering amplitude is less serious for more massive systems. We find that while the reionization bias is already effecting clustering studies of LBGs, it will become even more important as future surveys begin to discover populations of less massive high-redshift galaxies.

The reionization-induced bias should be sensitive to the type of selected galaxies. In the case of LBGs, we would therefore expect that the clustering amplitude would depend on the band in which the flux selection was performed. Alternatively, overdense regions would therefore be expected to have a slightly bluer population of galaxies. For LBGs our model predicts the systematic offset in $U_g - G$ colour to be $\sim 0.03$–$0.05$ mag. Note that this offset refers to the correlation between colour and large-scale overdensity within a colour-selected sample, and not to the galaxy population on average. The average galaxy population could show a different behaviour. For example, galaxies in overdense regions are observed at low redshift to be redder because they formed earlier, or because their cold gas was heated by mergers or stripped by the hot IGM in clusters. A correlation between galaxy colour and overdensity within a LBG

© 2007 The Authors. Journal compilation © 2007 RAS, MNRAS 382, 921–936
sample would be evidence for the reionization-induced galaxy bias, and could be used to calibrate its effect empirically.

Finally, we considered the importance of reionization-induced bias for current and upcoming surveys attempting to detect baryonic acoustic oscillations. We find that the contribution to the bias from reionization is very small in surveys of old stellar population from reionization-induced bias are scale dependent, but above of which the reionization-induced bias is scale dependent. The scale of the mean free path is uncomfortably close to the scale of the baryonic acoustic peak, implying that careful account may need to be taken of reionization-induced bias in galaxy surveys that select star-forming galaxies.

ACKNOWLEDGMENTS

We thank Dan Stark for helpful comments on an early draft of this paper. This work was supported by the Australian Research Council (JSBW) and Harvard University grants (AL). JSBW acknowledges the hospitality of the Institute of Astronomy at Cambridge University where part of this work was undertaken.

REFERENCES

Adelberger K. L., Steidel C. C., Pettini M., Shapley A. E., Reddy N. A., Erb D. K., 2005, ApJ, 619, 697
Babich D., Loeb A., 2006, ApJ, 640, 1
Barkana R., Loeb A., 2001, Phys. Rep., 349, 125
Barkana R., Loeb A., 2004, ApJ, 609, 474
Blake C., Glazebrook K., 2003, ApJ, 594, 665
Bolton J. S., Hahnelt M. G., Viel M., Carswell R. F., 2006, MNRAS, 366, 1378
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Budavári T. et al., 2003, ApJ, 595, 59
Coil A. L., Newman J. A., Kaiser N., Davis M., Ma C.-P., Kocevski D. D., Budavári T. et al., 2003, ApJ, 595, 675
Davidsen A. F., Kriss G. A., Zheng W., 1996, Nat, 380, 47
Dijkstra M., Haiman Z., Rees M. J., Weinberg D. H., 2004, ApJ, 601, 666
Efstathiou G., 1999, MNRAS, 256, 43
Eisenstein D. J. et al., 2001, AJ, 122, 2267
Eisenstein D. J. et al., 2005, ApJ, 633, 560
Fan X. et al., 2006, AJ, 137, 112
Glazebrook K. et al., 2007, preprint (astro-ph/0701876)
Heikal M., Williger G. M., Smette A., Hubeny I., Sahu M. S., Jenkins E. B., Tripp T. M., Winkler J. N., 2000, ApJ, 534, 69
Hogan C. J., Anderson S. F., Rugers M. H., 1997, AJ, 113, 1495
Jacobson P. et al., 1994, Nat, 370, 35
Kaiser N., 1987, MNRAS, 227, 1
Kessler M. A. et al., 1996, ApJS, 123, 3
Leitherer C. et al., 1999, ApJS, 123, 3
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
Press W., Schechter P., 1974, ApJ, 187, 425
Pritchard R. J., Furlanetto S. R., Kamionkowski M., 2007, MNRAS, 374, 159
Quinn T., Katz N., Efstathiou G., 1996, 278, L49
Reimers D. et al., 1997, A&A, 327, 890
Reimers D., Fechner C., Kriss G., Shull M., Baade R., Moos W., Songaila A., Simcoe R., 2006, in Sonneborn G., Moos H., Anderson B.-G., eds, ASP Conf. Ser. Vol 348. Astrophysics in the Far Ultraviolet: Five Years of Discovery with FUSE. Astron. Soc. Pac., San Francisco, p. 41
Scalo J., 1998, The Stellar Initial Mass Function (38th Herstmonceux Conference), 142, 201
Scannapieco E., Barkana R., 2002, ApJ, 571, 585
Schaye J., Theuns T., Rauch M., Efstathiou G., Sargent W. L. W., 2000, MNRAS, 318, 817
Shapley A. E., Steidel C. C., Adelberger K. L., Dickinson M., Giavalisco M., Pettini M., 2001, ApJ, 562, 95
Sheth R., Tormen G., 2002, MNRAS, 321, 61
Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1
Sneth A., Heap S. R., Williger G. M., Tripp T. M., Jenkins E. B., Songaila A., 2002, ApJ, 564, 542
Sokasian A., Abel T., Hernquist L., 2003, MNRAS, 340, 473
Spergel D. N., 2007, ApJS, 170, 377
Steidel C. L., Adelberger K. L., Shapley A. E., Pettini M., Dickinson M., Giavalisco M., 2003, ApJ, 592, 728
Tegmark M. et al., 2006, Phys. Rev. D, 74, 123507
Theuns T., Bernardi M., Frieman J., Hewett P., Schaye J., Sheth R. K., Subbarao M., 2002a, ApJ, 574, L111
Theuns T., Zaroubi S., Kim T.-S., Tzanavaris P., Carswell R. F., 2002b, MNRAS, 332, 367
Thoul A. A., Weinberg D. H., 1996, ApJ, 465, 608
Tytler D., 1995, in Meylan G., ed., Proc. ESO Workshop, QSO Absorption Lines. Springer-Verlag, Berlin, p. 289
Volonteri M., Haardt F., Madau P., 2003, ApJ, 582, 559
White R., Becker R., Fan X., Strauss M., 2003, AJ, 126, 1
Weinmann S. M., Maccio’o, A. V., Ileev I. T., Mallela G., Moore B., 2007, MNRAS, 381, 367
Wyithe J. S. B., Loeb A., 2003, ApJ, 586, 693
Wyithe J. S. B., Loeb A., 2004, Nat, 432, 194
Wyithe J. S. B., Loeb A., 2007, MNRAS, 375, 1034

APPENDIX A: THE DEPENDENCE OF THE GROWTH FACTOR ON LARGE-SCALE OVERDENSITY

In Section 5 we used the fact that the details of the merger tree history are not sensitive to the local large-scale overdensity. In this appendix we demonstrate this independence.

Consider the conditional probability function for the number of progenitors of mass between and that a halo of mass breaks into when one takes a small-redshift step \(dz\): \n
\[
\frac{dN}{dM'} = -\sqrt{\frac{2}{\pi}} \frac{M'}{M} \left[ \sigma_{\text{cm}}^2(M') - \sigma_{\text{cm}}^2(M) \right]^{-1.5} \times \frac{d\sigma_{\text{crit}}}{dz} \left[ \sigma_{\text{cm}}^2(M') \frac{d \log \sigma_{\text{cm}}(M')}{dM} \right] dz. \tag{A1}\n
\]

Equation (A1) is written in terms of comoving densities and variances (labelled with subscript 'cm'). This equation could be modified in the presence of a large-scale overdensity, which we denote \(\delta_{\text{cm}}\) when extrapolated to \(z = 0\) (making it 'comoving'), by adjusting the overdensity for collapse as is done to derive galaxy bias in the Press–Schechter formalism. This adjustment takes place in the term

\[
\frac{d\sigma_{\text{crit}}}{dz} = \frac{d\sigma_{\text{crit}}}{d\delta_{\text{cm}}(z)} \frac{d\delta_{\text{cm}}(z)}{dz} = \frac{d\sigma_{\text{crit}}}{d\delta_{\text{cm}}(z)} \frac{d\delta_{\text{cm}}(z)}{dz}, \tag{A2}\n
\]

where we have explicitly included both the dependence of the growth factor \(D\) on the comoving overdensity, and its normalization at \(z = 0\).

There remains a possible dependence on large-scale overdensity in \(d\sigma_{\text{crit}}/dz\) through the dependence on the growth factor. However, the variation of the evolution in the growth factor due to \(\delta_{\text{cm}}\) is present both in the growth factor and in its normalization, so that

\[
\frac{d\sigma_{\text{crit}}}{dz} = \frac{d\sigma_{\text{crit}}}{d\delta_{\text{cm}}(z)} \frac{d\delta_{\text{cm}}(z)}{dz} = \frac{d\sigma_{\text{crit}}}{d\delta_{\text{cm}}(z)} \frac{d\delta_{\text{cm}}(z)}{dz}. \tag{A2}\n\]
the ratio $D(\delta_{\text{cm}}, 0)/D(\delta_{\text{cm}}, z)$ should be independent of $\delta_{\text{cm}}$ to leading order.

To demonstrate this we compute the case of a zero cosmological constant (as is appropriate at high redshift). Locally, we can model a region of large-scale overdensity $\delta_{\text{cm}}$ as being carved out of a universe with a density parameter $\Omega_m(1 + \delta_{\text{cm}})$, where $\Omega_m$ is the mean density parameter of our Universe. We can then compute the evolution of structure and the local relation between scalefactor and time, within this modified universe. However, when we observe galaxies in the overdense region we see them at a redshift that is determined by the expansion history of the actual universe, not of the overdense region. To compare the evolution of structure in the mean universe and overdense regions, we therefore need to consider the growth factor within the overdense region at a fixed time in the past (corresponding to the expansion of the actual universe), rather than at a fixed scalefactor. Furthermore, the growth factor computed must then also be normalized by the growth factor in the overdense region at the current time in the usual way. We have

$$D \propto [\Omega_m(1 + \delta_{\text{cm}})]^{-1} a(\delta_{\text{cm}}). \quad (A3)$$

Here $a(\delta_{\text{cm}})$ is the overdensity-dependent scalefactor, which is related to the age of the universe at $a(\delta_{\text{cm}})$ through

$$t = \left(\frac{2}{3}\right) H_0^{-1} \Omega_m(1 + \delta_{\text{cm}})^{-1/2}[a(\delta_{\text{cm}})]^{1/2}, \quad (A4)$$

where $H_0$ is the local value of Hubble's constant. Substituting, we find

$$D \propto H_0^{2/3}(1 + \delta_{\text{cm}})^{-2/3} t^{2/3}, \quad (A5)$$

and hence

$$\frac{D(t)}{D(t_0)} = \left(\frac{t}{t_0}\right)^{2/3}. \quad (A6)$$

Thus the growth factor, normalized to the growth factor at the present day is independent of $\delta_{\text{cm}}$ (when the cosmological constant is ignored).

The above arguments imply that since $\delta_{\text{cm}}$ cancels in equation (A2), equation (A1), and therefore a merger tree is insensitive to the large-scale overdensity. While structure forms earlier in overdense regions leading to an enhanced merger rate, there are more haloes with which to merge so that the merger rate per halo remains constant.