Lyman-α Emitters During the Early Stages of Reionization

Andrei Mesinger* & Steven R. Furlanetto
Yale Center for Astronomy and Astrophysics, Yale University, PO Box 208121, New Haven, CT 06520-8121

12 November 2018

ABSTRACT

We investigate the potential of exploiting Lyα Emitters (LAEs) to constrain the volume-weighted mean neutral hydrogen fraction of the IGM, $\bar{x}_{\rm HI}$, at high redshifts (specifically $z \sim 9$). We use "semi-numerical" simulations to efficiently generate density, velocity, and halo fields at $z = 9$ in a 250 Mpc box, resolving halos with masses $M \geq 2.2 \times 10^8 M_{\odot}$. We construct ionization fields corresponding to various values of $\bar{x}_{\rm HI}$. With these, we generate LAE luminosity functions and "counts-in-cell" statistics. As in previous studies, we find that LAEs begin to disappear rapidly when $\bar{x}_{\rm HI} \gtrsim 0.5$. Constraining $\bar{x}_{\rm HI}(z = 9)$ with luminosity functions is difficult due to the many uncertainties inherent in the host halo mass $\leftrightarrow$ Lyα luminosity mapping. However, using a very conservative mapping, we show that the number densities derived using the six $z \sim 9$ LAEs recently discovered by Stark et al. (2007) imply $\bar{x}_{\rm HI} \lesssim 0.7$. On a more fundamental level, these LAE number densities, if genuine, *require* substantial star formation in halos with $M \lesssim 10^9 M_{\odot}$, making them unique among the current sample of observed high-z objects. Furthermore, reionization increases the apparent clustering of the observed LAEs. We show that a "counts-in-cell" statistic is a powerful probe of this effect, especially in the early stages of reionization. Specifically, we show that a field of view (typical of upcoming IR instruments) containing LAEs has $\gtrsim$10% higher probability of containing *more than one* LAE in a $\bar{x}_{\rm HI} \gtrsim 0.5$ universe than a $\bar{x}_{\rm HI} \approx 0$ universe with the same overall number density. With this statistic, an ionized universe can be robustly distinguished from one with $\bar{x}_{\rm HI} \gtrsim 0.5$ using a survey containing only ~ 20–100 galaxies.

Key words: cosmology: theory – early Universe – galaxies: formation – high-redshift – evolution

1 INTRODUCTION

The reionization of hydrogen in the intergalactic medium (IGM) is a landmark event in the early history of structure formation, because it defines the moment at which galaxies (and black holes) affected every baryon in the Universe. As such, it has received a great deal of attention – both observationally and theoretically – in the past several years. Unfortunately, the existing observational evidence is enigmatic (see Fan et al. 2006 for a recent review). Electron scattering of cosmic microwave background photons implies that reionization occurred at $z \sim 10$, albeit with a large uncertainty (Page et al. 2006). On the other hand, quasars at $z \sim 6$ show some evidence for a rapid transition in the globally-averaged neutral fraction, $\bar{x}_{\rm HI}$ (e.g., Fan et al. 2006; Mesinger & Haiman 2004, 2007). However the Lyα absorption is so saturated in the Gunn-Peterson (GP) trough that constraints derived from that spectral region (Fan et al. 2006; Maselli et al. 2007) are difficult to interpret (e.g, Lidz et al. 2006; Becker et al. 2007; Bolton & Haehnelt 2007).

Of particular recent interest have been efforts to constrain

reionization (and star formation at high redshifts) through searches for distant galaxies. Currently, the most efficient way to find distant galaxies is by searching for Lyα emission lines (which result from the re-processing of ionizing photons inside the galaxy; Partridge & Peebles 1967); such surveys now routinely reach $z \gtrsim 6$, where constraints on reionization become interesting (e.g., Hu et al. 2002; Kodaira et al. 2003; Rhoads et al. 2004; Santos et al. 2004; Stanway et al. 2004; Taniguchi et al. 2005; Kashikawa et al. 2006), and are now being stretched to even higher redshifts (Willis & Courbin 2005; Iye et al. 2006; Cuby et al. 2007; Stark et al. 2007; Ota et al. 2007). They offer a number of advantages over more traditional techniques. First, narrowband searches reduce the sky background, especially if placed between the bright sky lines that (nearly) blanket the near-infrared sky (e.g., Barton et al. 2004). Second, they efficiently select galaxies at a known redshift (albeit with some contamination by lower-redshift interlopers). Third, they increase the signal-to-noise by focusing on an emission line. Of course, the disadvantage is that extraordinarily deep spectroscopic follow-up is required to study the detailed properties of the sources (for an illustration of the difficulties, see Stark et al. 2007).

However, these properties are relatively unimportant for studying the ionization state of the IGM, which can be measured

---

* Email: andrei.mesinger@yale.edu

from the Lyα line photons themselves. In particular, these are absorbed if they pass through neutral gas near the galaxy. This is a consequence of the enormous Lyα optical depth of a neutral IGM: $\tau_{\rm GP} \sim 6.5 \times 10^5 \, \bar{x}_{\rm HI} \, [(1+z)/10]^{3/2}$ (Gunn & Peterson 1965), so even those photons passing through the damping wing of the Lyα resonance will be absorbed (Miralda-Escudé 1998).

Thus, as the IGM becomes more neutral, the Lyα selection technique will detect fewer and fewer objects (even after accounting for cosmological evolution in their intrinsic abundance); the number of such galaxies therefore measures $\bar{x}_{\rm HI}$ (Haiman & Spaans 1999). The optical depth encountered by a galaxy's Lyα photons depends primarily on the extent of the HII region that surrounds it: the photons redshift as they stream through the ionized gas (suffering little absorption), so they are somewhere in the wings of the line by the time they encounter the neutral gas. Thus, the amount of absorption depends sensitively on the size distribution of ionized bubbles during reionization. Early work treated each galaxy or quasar in isolation (Madau & Rees 2000; Haiman 2002; Santos 2004; Haiman & Cen 2005), so the HII regions were rather small even late in reionization. Including clustering dramatically increases the sizes of the ionized bubbles (Furlanetto et al. 2004), which allows Lyα galaxies to be visible farther back in reionization (Furlanetto et al. 2004, 2006; McQuinn et al. 2007).

The current observational picture is ambiguous. Malhotra & Rhoads (2004) compared luminosity functions of Lyman α-emitters (LAEs) at $z = 5.7$ and $z = 6.5$ (bracketing the time at which quasar spectra indicate that reionization ends) and found no evolution in the number density over this range, requiring a substantial ionized fraction at $z \sim 6.5$ (see also Haiman & Cen 2005; Furlanetto et al. 2006; Malhotra & Rhoads 2006). However, Kashikawa et al. (2006) found a decline in the number density of bright LAEs over the same redshift range (with a larger sample of objects), and Ota et al. (2007) suggest a farther decline to $z \sim 7$. On the other hand, Dawson et al. (2007) argued that the LAE luminosity function remains nearly constant from $z = 3–7$ using a compilation of other surveys. Moreover, Stark et al. (2007) found a surprisingly high abundance of faint emitters at $z \sim 9$. These objects are of particular interest to us, because they present the best possibility of being affected by reionization.

The interpretation of these results is complicated by the unknown internal properties of the sources: in particular, how do their Lyα line luminosities depend on the underlying halo mass, and what other factors affect the mapping? A number of recent studies have developed models for the high-redshift LAE population (Le Delliou et al. 2006; Dijkstra et al. 2006; Fernandez & Komatsu 2007; Stark et al. 2007; Kobayashi et al. 2007), but the relevant parameters (such as the star formation efficiency and time-scale and the initial mass function) are still mostly unconstrained, especially if they evolve. Indeed, Dijkstra et al. (2006) have argued that the declining number density of bright LAEs observed by Kashikawa et al. (2006) can be simply a result of the evolving mass function. On the other hand, Kashikawa et al. (2006) argued that the LAE luminosity function evolved more than that of Lyman-break selected galaxies over this range, so that the decline could not be attributed to the galaxies themselves (see also Ota et al. 2007). Most recently, Dawson et al. (2007) argued that the LAEs evolve less rapidly than other galaxy samples. In this paper, we will present new simulations of the luminosity function evolution throughout reionization, focusing on the high-redshift "frontier" at $z = 9$, and try to place some *model-independent* constraints on the sources and the IGM.

However, because of the difficulties involved, it is advantageous to consider other signatures of reionization. In particular, the clustering of LAEs can be a powerful probe of reionization (Furlanetto et al. 2004, 2006; McQuinn et al. 2007). This is because their visibility is modulated by the pattern of ionized bubbles: large bubbles (which surround overdensities with many galaxies) have a small damping-wing optical depth, so that most sources will be visible, while small bubbles (where galaxies are rare to begin with) will appear to be entirely empty. Thus, during reionization observable LAEs should appear *more* clustered than the underlying population, with the boost decreasing as the ionized bubbles grow (which happens relatively quickly). Clustering of the overall galaxy population should evolve much more slowly than number densities, so a rapid change in the clustering is a good indicator of reionization (McQuinn et al. 2007).

Existing work has examined clustering primarily in the context of the power spectrum (or correlation function) of the galaxies (Furlanetto et al. 2006; McQuinn et al. 2007; see also the appendix in McQuinn et al. 2007 for a brief discussion on void and peak statistics). Most recently, McQuinn et al. (2007) suggest that the lack of clustering evolution in the $z = 5.7$ and $z = 6.6$ LAEs rules out a substantially neutral universe at $z = 6.6$. However, the bubble modulation is actually highly non-gaussian. It is therefore useful to consider other statistics, especially early in reionization (when sources are so rare that measuring the scale-dependent power spectrum robustly will be extremely difficult). Here we take a first step in this direction by considering a "counts-in-cells" measurement of the clustering.

This paper is organized as follows. In §2, we briefly outline the main points of our "semi-numerical" simulation, originally presented in Mesinger & Furlanetto (2007a). In §3 we present the Lyα optical depth distributions of galaxies in our simulations, as functions of halo mass and $\bar{x}_{\rm HI}$. In §4 we show the resulting LAE luminosity functions, comparing with the Stark et al. (2007) sample in §4.1. In §5, we present statistics using "counts in-cell". Finally in §6 we summarize our key findings and offer some conclusions.

Unless stated otherwise, we quote all quantities in comoving units. We adopt the background cosmological parameters (Ω_Λ, $\Omega_{\rm M}$, Ω_b, n, σ_8, H_0) = (0.76, 0.24, 0.0407, 1, 0.76, 72 km s^{-1} Mpc^{-1}), consistent with the three–year results of the *WMAP* satellite (Spergel et al. 2006).

2   SEMI-NUMERICAL SIMULATIONS

We use an excursion-set approach combined with first-order Lagrangian perturbation theory to efficiently generate density, velocity, halo, and ionization fields at $z = 9$. This "semi-numerical" simulation is presented in Mesinger & Furlanetto (2007a), to which we refer the reader for details. A similar halo-finding scheme has also been presented by Bond & Myers (1996) and a similar scheme to generate ionization fields has been presented by Zahn et al. (2007).

Our simulation box is 250 Mpc on a side, with the final density, velocity and ionization fields having grid cell sizes of 0.5 Mpc. Halos with a total mass $M \geq 2.2 \times 10^8 \, M_\odot$ are filtered out of the linear density field using excursion-set theory, with mass scales spaced as $\Delta M/M = 1.2$. Note that we are able to resolve halos with masses less than a factor of two from the cooling mass likely to be pertinent mid-reionization (or more precisely, during the redshift interval $6 \lesssim z \lesssim 10$), corresponding to gas with a temperature of $T \sim 10^4$ K (e.g. Efstathiou 1992; Thoul & Weinberg 1996; Gnedin 2000; Shapiro et al. 1994). Halo locations are then adjusted using first-order Lagrangian perturbation theory. The resulting halo field
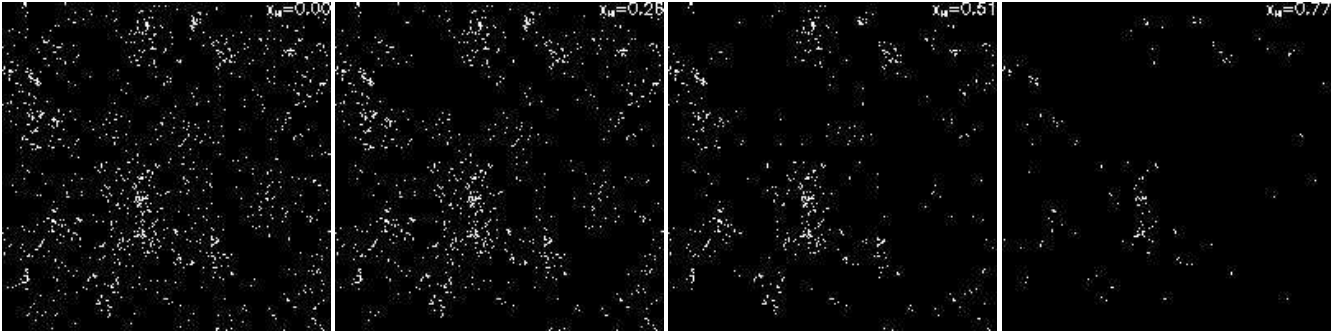
Figure 1. Maps of visible halos at $z = 9$, assuming $M_{\min} = 1.67 \times 10^{10} \, M_\odot$, and $\bar{x}_{\rm HI} \approx 0$, 0.26, 0.51, 0.77, (*left to right*). All slices are 250 Mpc on a side and 20 Mpc deep (corresponding to a narrow band filter with $R = \lambda/\Delta\lambda \sim 125$). The 1500^3 halo field is smoothed onto a 200^3 grid here for viewing purposes.

matches both the mass function and statistical clustering properties of halos in N-body simulations (Mesinger & Furlanetto 2007a).

In constructing the ionization field, the IGM is modeled as a two-phase medium, comprised of fully ionized and fully neutral regions (this is a fairly accurate assumption at high-redshifts preceding the end of reionization, unless the X-ray background is rather strong). Using the same halo field at $z = 9$, we generate ionization fields corresponding to different values of $\bar{x}_{\rm HI}$ by varying a single efficiency parameter,[1] ζ, again using the excursion-set approach (c.f. Mesinger & Furlanetto 2007a; Furlanetto et al. 2004).

This semi-numeric approach is thus ideally suited to the LAE problem, because we are able to "resolve" relatively small halos and simultaneously sample a large, representative volume of ionized bubbles. Note that our "simulations" do *not* make any predictions (and only weak assumptions) about the Lyα luminosities of these sources; we will discuss the mapping from halo mass (the fundamental quantity for our simulations) to observable properties below. This mapping must also be prescribed in state-of-the-art cosmological simulations, which cannot self-consistently include hydrodynamics (and hence star formation) while also subtending a representative volume during reionization (c.f., McQuinn et al. 2007).

3 DAMPING WING OPTICAL DEPTH DISTRIBUTIONS

To study the effects of reionization, we first need to track the absorption of line photons from neutral gas in the IGM. We divide the absorption into two parts: the resonant and damping wing components. This is convenient because they correspond to two spatially distinct sets of absorbers. Resonant absorption occurs whenever a photon that begins blueward of line center redshifts into resonance

(either inside the HII region surrounding the source or in the neutral gas outside). Because the line-center optical depth is so large, this component can lead to nearly complete absorption – but only for photons on the blue side of the line (e.g., Santos 2004). We do not model this component in detail in this work, assuming that a constant fraction of the Lyα line gets resonantly absorbed.

On the other hand, photons that begin redward of line center only redshift farther away. It is therefore only the damping wings of the line that affect them, and the amount of absorption, $\exp[-\tau_D]$, where τ_D is the damping-wing optical depth, will depend sensitively on the size of the host HII region but be *insensitive* to the precise $x_{\rm HI}$ inside the ionized region. It is this component that evolves most rapidly through reionization. Figure 1 shows the visible halos at $z = 9$, with $M \exp[-\tau_D] > 1.67 \times 10^{10} M_\odot$, and $\bar{x}_{\rm HI} \approx 0$, 0.26, 0.51, 0.77, (*left to right*); the obscuration from damping wing absorption is obvious.

We compute the total line center Lyα optical depth along a randomly chosen line-of-sight (LOS) centered on a halo location at $z_{\rm s} = 9.0$. We do this by summing the damping wing optical depth, τ_D, contribution from each neutral hydrogen patch (extending from $z_{\rm begin}$ to $z_{\rm end}$) encountered along the LOS, using the approximation (Miralda-Escude 1998):

$$\tau_D = 6.43 \times 10^{-9} \left(\frac{\pi e^2 f_\alpha n_H(z_{\rm s})}{m_e c H(z_{\rm s})} \right) \tag{1}$$
$$\times \left[I \left(\frac{1 + z_{\rm begin}}{1 + z_{\rm s}} \right) - I \left(\frac{1 + z_{\rm end}}{1 + z_{\rm s}} \right) \right]$$

where $n_H(z_{\rm s})$ is the mean hydrogen number density of the IGM at redshift $z_{\rm s}$, and

$$I(x) \equiv \frac{x^{9/2}}{1-x} + \frac{9}{7} x^{7/2} + \frac{9}{5} x^{5/2} + 3x^{3/2} + 9x^{1/2}$$
$$- \ln \left| \frac{1 + x^{1/2}}{1 - x^{1/2}} \right| .$$

We use eq. (1) to calculate the optical depth for each neutral hydrogen patch, summing the contributions of patches along the LOS for 200 Mpc,[2] wrapping around the simulation box if needed. We construct distributions of τ_D for each halo mass scale and ionization topology (i.e. $\bar{x}_{\rm HI}$). We make sure to process LOSs from every halo of a particular mass scale, cycling through the halo list until

---

[1] This differs from the method recently used by McQuinn et al. (2007), who used a suite of N-body simulations with radiative transfer. They also argued that a faster but effective method was to generate ionization fields at several different redshifts (using a single radiative transfer simulation) but apply them to a halo field at a single redshift. They thus assumed that the ionization topology is only a weak function of redshift (McQuinn et al. 2007). The speed of our approach, which does not require a radiative transfer algorithm, allows us to generate ionization fields at a *single* redshift self-consistently, using the same halo field, merely by adjusting the source efficiencies. However, we confirm that the ionization maps are very nearly redshift-independent for most purposes (including those studied by McQuinn et al. 2007). The exceptions to this are the rare events occurring in $\lesssim 10^{-3}$ of the typical fields of view discussed in §5.

[2] This number was chosen experimentally in order to ensure convergence of the τ_D distributions at the mass scales and neutral fractions studied in this work.

each mass scale undergoes a minimum of 3×10^4 such Monte-Carlo realizations. We also include the component of the source halo's peculiar velocity along the LOS, v, in our estimates of τ_D by substituting $z_s \to z_s + v/c$. [3]

As delineated above, we only compute τ_D at a single observed wavelength, $\lambda_{\rm obs} = 1215.67(1 + z_s + v/c)$ Å, instead of over the entire wavelength extent of the Lyα emission line. To zeroth order, the damping wing optical depth varies only weakly across the typical scale of the Lyα emission line, although its shape does provide useful information about the IGM properties; we refer the interested reader to Mesinger & Furlanetto (2007b); McQuinn et al. (2007). We also emphasize that, in our two-phase IGM approximation, we do not model the resonance contribution to the total optical depth. However, any *mass-independent* attenuation of the Lyα line can be swept into the assumed halo mass $\leftrightarrow$ Lyα luminosity mapping (c.f. §4.1). Most studies (including our own in §4) assume such a mass-independent resonant attenuation, justifying it with the narrow halo mass range (~ 1 dex) probed by present instruments/surveys (e.g. Dijkstra et al. 2006; Stark et al. 2007; McQuinn et al. 2007). We defer a detailed study of this assumption to a future work.

In Figure 2, we show some of the optical depth distributions generated by the above procedure, for $\bar{x}_{\rm HI}(z = 9) = 0.93, 0.72, 0.51, 0.34$ (*left to right*). The panels show $\tau_D \partial p (> \tau_D, M, \bar{x}_{\rm HI})/\partial \tau_D$, i.e. the probability per $\ln \tau_D$ that a LOS originating from a halo of mass M embedded in an IGM with a neutral fraction of $\bar{x}_{\rm HI}$ has an optical depth of τ_D. Curves correspond to LOSs originating from halos with masses $M = 2.6 \times 10^{11}$, 1.2×10^{11}, 5.4×10^{10}, 2.5×10^{10}, 1.1×10^{10}, 5.1×10^9, and $2.3 \times 10^9 \, M_\odot$ (*left to right*). Solid curves include the peculiar velocity offset of the host halo; dotted curves do not.

In general, the optical depth distributions become narrower and their mean values decrease as the halo mass increases. This is to be expected because the halo bias is a function of mass, with the more massive halos more likely to sit in larger overdense regions – which in turn contain more ionizing sources and larger HII bubbles than less massive halos (e.g. Furlanetto et al. 2006; McQuinn et al. 2007). Furthermore, since the overlap of several HII bubbles makes a smaller *fractional* change to the size of a large bubble hosting the most massive halos, the distributions of τ_D are narrower for massive halos. Both of these effects diminish as reionization progresses, and the τ_D distributions start merging together.

It is also evident from Fig. 2 that galaxy peculiar velocities broaden the optical depth distributions, shifting their means to smaller values. This effect is the strongest for the smallest mass halos, because they are typically surrounded by the smallest HII bubbles (whose edges are at the smallest velocity offsets from the Lyα line center). The effects of source halo velocities on τ_D also diminishes with decreasing $\bar{x}_{\rm HI}$; for $\bar{x}_{\rm HI} \lesssim 0.7$ peculiar velocities play a negligible role in determining τ_D for halos with $M \gtrsim 10^9 M_\odot$. More importantly, Fig. 2 shows that peculiar velocities have virtually no impact on the optical depth distributions in the pertinent,

---

[3] Note that for simplicity we do *not* include the peculiar velocity of the neutral IGM patches, which could be correlated on large scales with the peculiar velocities of the sources, thus diminishing the impact of velocities on τ_D. However, note from Fig. 2 that peculiar velocities do not play a major role in the optical depth distributions except when $\tau_D \gg 1$ and $\bar{x}_{\rm HI} \sim 1$ (where the absorption is so strong that its precise value does not matter for our purposes). The treatment of velocities is uncertain in any case because we ignore the possibility of galactic winds, which can move the Lyα lines redward and decrease the absorption (Santos 2004).
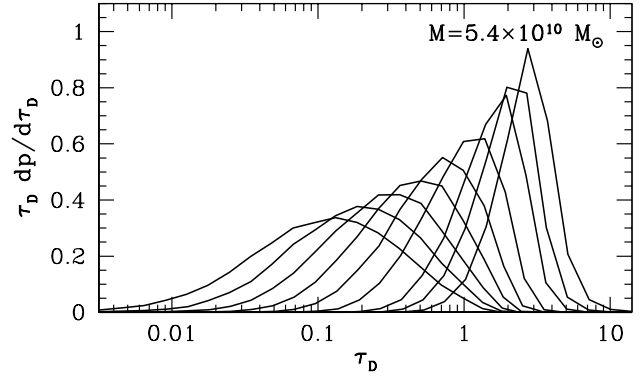
Figure 3. Damping wing optical depth distributions generated using LOSs originating from halos with masses $M = 5.4 \times 10^{10} \, M_\odot$. Curves correspond to $\bar{x}_{\rm HI} = 0.26, 0.34, 0.42, 0.51, 0.61, 0.72, 0.83, 0.88, 0.93$, left to right.

$\tau_D \sim 1$ regime. Hence, they should not have a noticeable effect on any of our conclusions.

Figure 3 quantifies the evolution of the optical depth distributions with $\bar{x}_{\rm HI}$. The distributions were generated using source halos at a fixed mass scale of $M = 5.4 \times 10^{10} M_\odot$. Curves correspond to $\bar{x}_{\rm HI} = 0.26, 0.34, 0.42, 0.51, 0.61, 0.72, 0.83, 0.88, 0.93$, from left to right. The curves illustrate that the more complex topologies and large scatter among disparate LOSs at lower values of $\bar{x}_{\rm HI}$ significantly broaden the distributions of τ_D, in addition to decreasing their means. This is not because the bubble size distribution widens (in fact, it narrows as reionization progresses; Furlanetto et al. 2006) but because some LOSs pass almost entirely through neighboring ionized bubbles, while others pass through long skewers of neutral gas. A similar effect has been shown to have important consequences for the interpretation of quasar spectra (Lidz et al. 2007).

3.1 Comparison to an Analytic Model

Furlanetto et al. (2004) presented similar optical depth distributions calculated via the analytic Furlanetto et al. (2004) model for the ionized bubbles during reionization. Our "semi-numeric" simulation is also based on this model, but it includes additional effects such as the complicated, non-spherical geometry of the HII regions. It is therefore illuminating to compare our distributions with those of the purely analytic model. For the latter, we follow the calculation of Furlanetto et al. (2004) except that for simplicity we assume that all ionizing sources sit in the center of their bubble. The damping wing optical depth distributions then follow from arguments similar to the "extended Press-Schechter" formalism (Press & Schechter 1974; Bond et al. 1991; Lacey & Cole 1993).

Figure 4 shows our results for halos with $M = 5.4 \times 10^{10} M_\odot$ and a variety of neutral fractions. The solid curves are again taken from the simulations. The dashed curves show the predictions of the analytic model. There are two key differences with the simulated results. First, the analytic model predicts that the distribution peaks at larger τ_D than the simulation. This is not a surprise: Mesinger & Furlanetto (2007a) showed that, when defined in this way, the simulations have larger ionized bubbles early in reionization. This is because of neighboring bubbles that slightly overlap; the spherical geometry required by the analytic model does not ef-
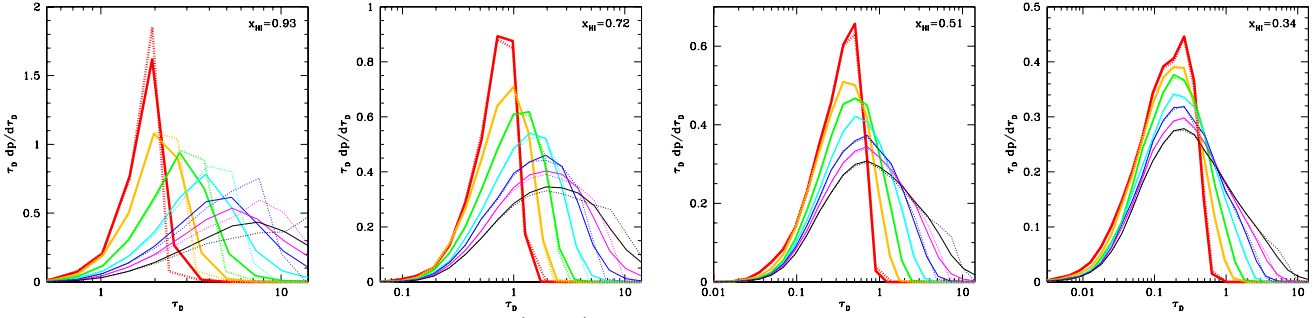
Figure 2. Damping wing optical depth distributions at $\bar{x}_{\mathrm{HI}}(z=9) = 0.93, 0.72, 0.51, 0.34$ (*left to right*). Curves correspond to LOSs originating from halos with masses $M = 2.6 \times 10^{11}, 1.2 \times 10^{11}, 5.4 \times 10^{10}, 2.5 \times 10^{10}, 1.1 \times 10^{10}, 5.1 \times 10^9$, and $2.3 \times 10^9 \, M_\odot$ (*left to right*). Solid curves include the peculiar velocity offset of the host halo; dotted curves do not.
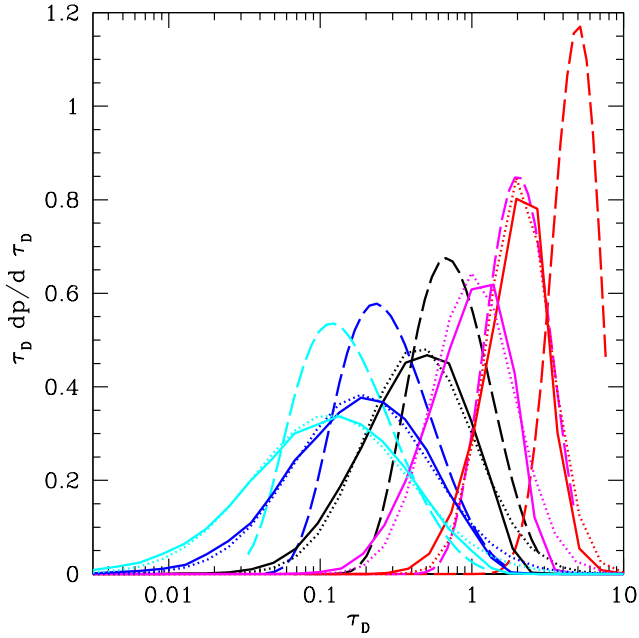
Figure 4. Damping wing optical depth distributions for halos with masses $M = 5.4 \times 10^{10} \, M_\odot$. The solid curves are computed from our simulations, and the dotted curves are the corresponding best-fit lognormal distributions. The dashed curves show the predictions of the analytic model. Within each set, the curves correspond to $\bar{x}_{\mathrm{HI}} = 0.26, 0.34, 0.51, 0.72, 0.88$, from left to right.

fectively capture such events. This can reduce the apparent optical depth by a factor of ~ 3 early on, so it is not a small effect.

Second, the analytic model also under-predicts the width of the distribution, particularly on the small-τ_D tail at the end of reionization. This is significantly larger than the differences in the distributions of bubble sizes. Instead, it likely results from the analytic model's assumption of a uniformly ionized medium outside of the source's host bubble: as described above, many lines of sight will pass through nearby, fully-ionized bubbles, allowing for the existence of many (nearly) clear lines of sight (see Mesinger & Furlanetto 2007b; McQuinn et al. 2007).

Because the analytic model does not provide a good fit, we searched for a better representation of the distributions. We find that our results are nearly always well-fit by log-normal distributions,

$$\frac{\partial p(> \tau_D)}{\partial \tau_D} = \frac{1}{\tau_D \sqrt{2\pi\sigma_D^2}} \exp\left[-\frac{(\ln \tau_D - \mu_D)^2}{2\sigma_D^2} \right], \qquad (2)$$

where μ_D and σ_D are parameters determined by fits to the simulated distributions. Some example fits are shown by the dotted curves in Figure 4; in general, the lognormal distribution overestimates the strength of the tail at large τ_D and underestimates its strength at small τ_D, with the skewness increasing for smaller halos and larger neutral fractions. The following simple bilinear fits for the parameters are accurate to $\lesssim 10\%$ over the entire mass range of our simulations and from $\bar{x}_{\mathrm{HI}} = 0.26$–$0.93$:

$$\mu_D = -3.37 + \log M_{10}(-0.115 - 0.587\bar{x}_{\mathrm{HI}}) + 5.30\bar{x}_{\mathrm{HI}} \quad (3)$$

$$\sigma_D = 1.68 + \log M_{10}(-0.155 - 0.265\bar{x}_{\mathrm{HI}}) - 1.08\bar{x}_{\mathrm{HI}}, \quad (4)$$

where $M = M_{10} \times 10^{10} M_\odot$. Note, however, that this fit only applies to halos at $z = 9$; the dependence on mass will no doubt change with redshift. It is possible, however, that the dependence on $\bar{x}_{\mathrm{HI}}$ is more robust, because the ionization pattern is nearly independent of the timing of reionization (Furlanetto et al. 2006; McQuinn et al. 2007).
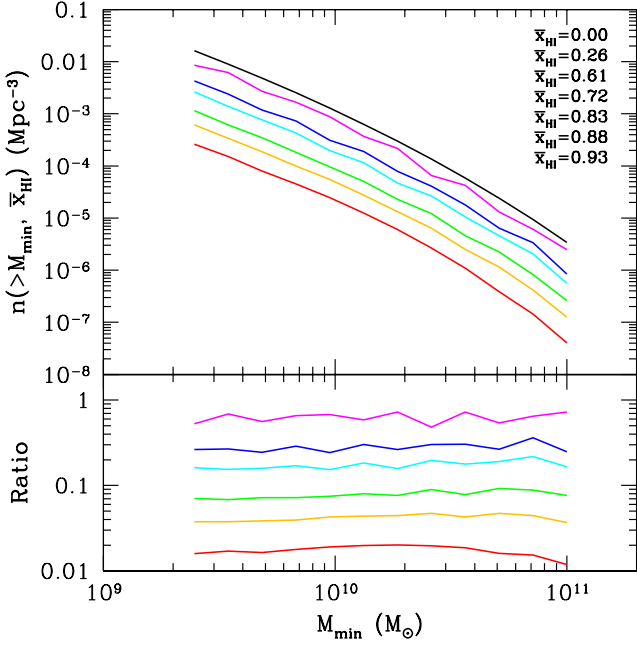
4 Z=9 LAE LUMINOSITY FUNCTIONS

With our optical depth distributions in-hand, we can proceed to generate $z = 9$ LAE luminosity functions. To do this, we make the standard simplifying assumption (e.g. Furlanetto et al. 2006; Dijkstra et al. 2006; Stark et al. 2007; McQuinn et al. 2007) of a deterministic, linear mapping of halo mass to Lyα luminosity, $M \propto L$. As mentioned above, this assumption is often justified by the narrow mass range probed by existing instruments. Using this ansatz, the observed Lyα luminosity of a LAE, $L_{\mathrm{obs}} = Le^{-\tau_D}$ (where the "intrinsic" Lyα luminosity, L, includes resonant attenuation), can be written in terms of the halo mass: $M_{\mathrm{obs}} = Me^{-\tau_D}$, where M_{obs} is the apparent mass of the halo under this mapping, after attenuation by the IGM. Given the minimum observable luminosity, $L(M_{\min})$ (where $M_{\min}$ is the mass that would produce a luminosity L without any IGM absorption), one could then detect halos with $\tau_D < -\ln(M_{\min}/M)$. In order to maintain generality, we postpone the discussion of the $L(M)$ mapping until §4.1, where we compare the luminosity functions to observations.

Hence, the cumulative number density of observable halos can be written as:

$$n(> M_{\min}, \bar{x}_{\mathrm{HI}}) = \int_{M_{\min}}^{\infty} dM \frac{dn(>M)}{dM} \qquad (5)$$

$$\times \int_0^{\ln(M/M_{\min})} d\tau_D \frac{\partial p(> \tau_D, M, \bar{x}_{\mathrm{HI}})}{\partial \tau_D},$$

where $dn(>M)/dM$ is the number density of halos per unit mass,

Figure 5. Luminosity functions at $z = 9$. *Top:* Number density of objects brighter than some observed luminosity $L_{\rm obs}(M_{\rm min})$, i.e. with $M \exp[-\tau_D] > M_{\rm min}$. Curves correspond to $\bar{x}_{\rm HI} \approx 0$, 0.26, 0.61, 0.72, 0.83, 0.88, 0.93 (*top to bottom*). *Bottom:* Ratio of luminosity functions: $n(> M_{\rm min}, \bar{x}_{\rm HI})/n(> M_{\rm min}, 0)$. Curves correspond to $\bar{x}_{\rm HI} = 0.26$, 0.61, 0.72, 0.83, 0.88, 0.93 (*top to bottom*).

and the second integral is the fraction of halos of mass M which can be detected above the threshold $M_{\rm min}$.

"Luminosity" functions generated with this procedure are shown in the top panel of Figure 5. The curves correspond to $\bar{x}_{\rm HI} \approx$ 0, 0.26, 0.61, 0.72, 0.83, 0.88, 0.93 (*top to bottom*). The bottom panel shows the ratio of the apparent and intrinsic luminosity functions: $n(> M_{\rm min}, \bar{x}_{\rm HI})/n(> M_{\rm min}, 0)$. Curves correspond to $\bar{x}_{\rm HI}$ = 0.26, 0.61, 0.72, 0.83, 0.88, 0.93 (*top to bottom*).

The luminosity functions in Fig. 5 represent the first such predictions for $z = 9$ which include the effect of inhomogeneous reionization. As in analytic models and other simulations, the decline in the number density of observable LAE in a partially neutral universe when compared to a fully ionized universe (see bottom panel of Fig. 5) is nearly independent of halo mass (Furlanetto et al. 2006; McQuinn et al. 2007).[4] McQuinn et al. (2007) examined this question in detail and showed that the scale-independence is robust to other assumed mappings between halo mass and luminosity. One explanation is that a lognormal distribution of optical depths, acting on a power-law intrinsic luminosity function, produces a nearly power-law apparent luminosity function. The bright end, where the intrinsic luminosity function falls exponentially, would also fall exponentially except that the broad distribution of optical depths nearly erases the change in slope. Our results are consistent with this explanation (see also McQuinn et al. 2007), especially because we have shown that the optical-depth distribution is broadest for

massive objects. The (nearly) scale-independent suppression appears to be a generic feature of reionization, requiring some other explanation for the decline in number density for bright LAEs observed by Kashikawa et al. (2006), and possibly that of Ota et al. (2007) as well.

4.1 Existing Observational Constraints

By taking advantage of the strong magnification provided by gravitational lensing through galaxy clusters, Stark et al. (2007) found six candidate ($> 5\sigma$) LAEs in the redshift range $z = 8.7$–10.2, with all but one falling within $z = 9 \pm 0.35$. The candidates have unlensed luminosity estimates ranging from $10^{41.2}$ to $10^{42.7}$ ergs s^{-1}. A search for additional emission lines expected if the candidates were low-z interlopers has been completed and has not found evidence for a low-z scenario for any of the six candidates. If genuine, these detections can provide an invaluable first glimpse into the $z \sim 9$ universe.

In order to compare our cumulative number densities in Fig. 5 with observations, one needs a mapping of Lyα luminosity, L, to halo mass, M. Even if one assumes a deterministic, linear relation, there are still many unknowns. The simplest version follows a well-trod path. We begin by assuming that about 2/3 of the ionizing photons absorbed within the galaxy are converted into Lyα photons (Osterbrock 1989). One can then write the conversion as $L = 0.67h\nu_\alpha(1 - f_{\rm esc})\dot{\rho}_*\epsilon_\gamma T_{\gamma,\rm res}$, where ν_α is the rest-frame Lyα frequency, $f_{\rm esc}$ is the escape fraction of ionizing photons, $\dot{\rho}_*$ is the star formation rate (SFR), ϵ_γ is the ionizing photon efficiency per stellar mass, and $T_{\gamma,\rm res}$ is the fraction of Lyα photons which escape from the galaxy without getting *resonantly* absorbed.[5] Assuming that galaxies steadily convert a fraction, f_*, of their gas into stars over some mean time-scale, t_*, and that $f_{\rm esc} \ll 1$, one can write the above relation as:

$$
\begin{aligned}
L &= 0.67h\nu_\alpha f_* \frac{\Omega_b}{\Omega_{\rm M}} M \frac{1}{t_*} \epsilon_\gamma T_{\gamma,\rm res} \quad (6) \\
&= 1.88 \times 10^{-12} {\rm erg} \left(\frac{\epsilon_\gamma f_* T_{\gamma,\rm res}}{t_*} \right) M .
\end{aligned}
$$

The free parameters in equation (6) are all almost unconstrained at high redshifts. Hence, we are interested in exploring a wide range of possibilities. From an astrophysical standpoint, it is much easier to use observed number densities to set robust *upper limits* on $\bar{x}_{\rm HI}$ than it is to set robust lower limits, since in the theoretical model one at least has the hard upper limit of using all available gas in every galaxy above the detection threshold. Setting conservative upper limits on $\bar{x}_{\rm HI}$ translates to maximizing $(\epsilon_\gamma f_* T_{\gamma,\rm res}/t_*)$ in the $L \leftrightarrow M$ mapping. Keeping this in mind, we apply four different $L \leftrightarrow M$ choices to the Stark et al. (2007) sample, with values of ϵ_γ all taken from Schaerer (2003) assuming metallicities of $Z \sim 0.04$ Z_* for Pop II and $Z \sim 10^{-7}$ Z_* for Pop III star formation:

(i) $z \sim 6$ parameters with Pop II stars. Here we use values which have been shown to fit $z = 5.7$ LAE luminosity functions fairly well (Dijkstra et al. 2006; Stark et al. 2007; McQuinn et al. 2007). We set $f_* T_{\gamma,\rm res} = 0.1$, $t_* = 2/3$ of the Hubble time = 1.87×10^{16} s, and $\epsilon_\gamma = 6.3 \times 10^{60}$ ionizing photons $M_\odot^{-1}$ as might be expected from a Pop II IMF. This translates to the relation

---

[4] The slightly "wavy" features in some of the curves in Fig. 5 result from the discrete halo masses returned by our halo filtering procedure. The effect is more noticeable at small $\bar{x}_{\rm HI}$, when the τ_D distributions become weaker functions of M (see Fig. 2 and eq. 5).

[5] As we have seen, our simulations model the damping wing component of the IGM absorption, τ_D. Hence, in order to compare with our simulated number densities, we only need worry about the fraction of Lyα photons that are *resonantly* absorbed. Defined as such, $T_{\gamma,\rm res}$ ranges approximately from 0.5 to 1.

$L = 6.3 \times 10^{31} \mathrm{erg\ s^{-1}} (M/M_\odot)$, and to roughly 0.24 $(f_\mathrm{esc}/0.02)$ ionizing photons per H atom.[6]

(ii) $z \sim 6$ parameters with Pop III stars. Here we again take values of f_*, $T_{\gamma,\mathrm{res}}$, and t_* which fit $z = 5.7$ LAE luminosity functions but use a Pop III IMF for ϵ_γ at $z = 9$. Stark et al. (2007) show that a similar model can fit the Stark et al. (2007) constraints moderately well. Specifically, we set $f_* T_{\gamma,\mathrm{res}} = 0.1$, $t_* = 2/3$ of the Hubble time $= 1.87 \times 10^{16}$ s, and $\epsilon_\gamma = 8.1 \times 10^{61}$ ionizing photons $M_\odot^{-1}$ as might be expected from a Pop III IMF. This translates to the relation $L = 8.2 \times 10^{32} \mathrm{erg\ s^{-1}} (M/M_\odot)$, and to 3.1 $(f_\mathrm{esc}/0.02)$ ionizing photons per H atom.

(iii) "maximally" conservative with Pop II stars. Here we push the limits of a Pop II model by setting $f_* T_{\gamma,\mathrm{res}} = 1$, $t_* =$ the dynamical time $= 4.6 \times 10^{15}$ s, and $\epsilon_\gamma = 6.3 \times 10^{60}$ ionizing photons $M_\odot^{-1}$ as might be expected from a Pop II IMF. In other words, we assume that every baryon inside every halo is converted into stars in one dynamical time (with this star formation episode synchronized across all halos) and every Lyα photon escapes the environs of the galaxy. This translates to the relation $L = 2.6 \times 10^{33} \mathrm{erg\ s^{-1}} (M/M_\odot)$, and to 2.4 $(f_\mathrm{esc}/0.02)$ ionizing photons per H atom.

(iv) "maximally" conservative with Pop III stars. Here we completely push the limits by setting $f_* T_{\gamma,\mathrm{res}} = 1$, $t_* =$ the dynamical time $= 4.6 \times 10^{15}$ s, *and* $\epsilon_\gamma = 8.1 \times 10^{61}$ ionizing photons $M_\odot^{-1}$ as might be expected from a Pop III IMF. This translates to the relation $L = 3.3 \times 10^{34} \mathrm{erg\ s^{-1}} (M/M_\odot)$, and to 31 $(f_\mathrm{esc}/0.02)$ ionizing photons per H atom.[7]

The relations implied by (ii) or (iii) are probably the most reasonably conservative estimates, as any stronger sources would have produced many more photons than are required to reionize the IGM. Relations with more efficient star formation [such as the extreme (iv)] would result in an early reionization, contrary to the evidence suggested by *WMAP* (Page et al. 2006) and the SDSS quasar spectra (Fan et al. 2006; Mesinger & Haiman 2004, 2007).

Number densities derived from the $z \sim 9$ candidates in the Stark et al. (2007) survey, transformed to mass units using the $L \leftrightarrow M$ relations above are shown in Figure 6, with the (i), (ii), (iii), (iv) mappings shown right to left in the figure [8]. Note that extending the $\bar{x}_\mathrm{HI} > 0$ luminosity functions from our model to the low scales required by the most conservative $L \leftrightarrow M$ relation is computationally prohibitive. Luckily, the suppression of the luminosity function is almost independent of M (see the bottom panel of Fig. 5), allowing us to extend the $\bar{x}_\mathrm{HI} > 0$ estimates by using the mean suppression ratios obtained over the mass range $10^9 - 10^{11}$ $M_\odot$. Solid curves thusly generated correspond to $\bar{x}_\mathrm{HI} \approx 0, 0.26, 0.61, 0.72, 0.83, 0.88, 0.93$ (*top to bottom*).

To estimate the size of the total (Poisson and cosmic variance) uncertainties, we have performed Monte-Carlo simulations mod-
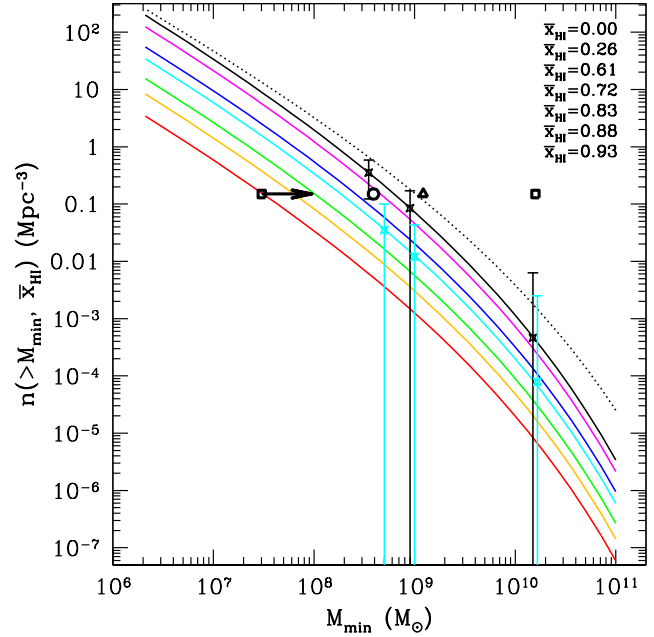
Figure 6. Luminosity functions at $z = 9$. Points correspond to number densities derived from the six $z \sim 9$ candidates in the Stark et al. (2007) survey, mapped using the $L \leftrightarrow M$ relations in the text, with (i), (ii), (iii), (iv) mappings shown right to left in the figure. Solid curves correspond to $\bar{x}_\mathrm{HI} \approx 0, 0.26, 0.61, 0.72, 0.83, 0.88, 0.93$ (*top to bottom*). The dotted curve corresponds to $\bar{x}_\mathrm{HI} \approx 0$, but with $\sigma_8 = 0.86$. Total error bars (Poisson and cosmic variance), generated with the Monte-Carlo procedure described in text, are shown for $\bar{x}_\mathrm{HI} \approx 0$ and 0.72, at several choices of M_min.

eling the Stark et al. (2007) observations, which used the lensing signature of 9 clusters to probe a combined volume of 13.5 Mpc^3, given $L > 10^{42}$ erg s^{-1}. Specifically, we tile our simulation box with cells whose volume is equal to the mean volume probed by each cluster, 1.5 Mpc^3.[9] We then repeatedly randomly select 9 such sample volumes in our simulation box, keeping track of the total number of LAEs in each group of 9 sample volumes.[10] The total error bars thus generated for $\bar{x}_\mathrm{HI} \approx 0$ and 0.72, at several choices of M_min are now shown in Fig. 6.

Assuming that the Stark et al. (2007) candidates are genuine and using the maximally conservative relation in (iv), $\bar{x}_\mathrm{HI}(z \sim 9)$ cannot be constrained: such a scenario would permit the IGM to be almost entirely neutral. However, it is very unlikely that *all* the factors in eq. (6) conspire to provide such efficient star formation and Lyα photon transmission (and even if they did, reionization would likely have ended long before). The more reasonably conservative models, (ii) and (iii), prefer $\bar{x}_\mathrm{HI} \lesssim 0.2$ and $\bar{x}_\mathrm{HI} \lesssim 0.7$, respectively. The $L \leftrightarrow M$ relation from (i), which has been shown to provide

---

[6] We estimate the number of ionizing photons per H atom by $(\epsilon_*/n_\mathrm{H}) f_* f_\mathrm{esc} (\Omega_b/\Omega_\mathrm{M}) \int M dn(> M)/dM\ dM$, where n_H is the number density of hydrogen atoms, the integration extends over all halos at $z = 9$ above the cooling threshold $T_\mathrm{vir} = 10^4$ K, and we set $T_{\gamma,\mathrm{res}} = 1$ in order to be conservative.

[7] We will see that this extreme mapping implies that the Stark et al. (2007) galaxies are below the $T_\mathrm{vir} = 10^4$ K atomic cooling threshold. If we instead allow stars to form down to the H2 cooling threshold, $T_\mathrm{vir} \sim 300$ K, this mapping would imply a whopping 140 $(f_\mathrm{esc}/0.02)$ ionizing photons per H atom!

[8] Of the three luminosity bins presented in Stark et al. (2007), we only show the most tightly constrained one at $L = 10^{42}$ erg s^{-1} This is the only number density with less than 100% Poisson errors.

[9] Note that the estimated geometry of this lensed long-slit spectroscopic survey can be approximated with a long parallelepiped with a 0.02 Mpc^2 FOV and a LOS distance of 500 Mpc per cluster (Stark et al. 2007). Since our box is 250 Mpc on a side and our halo field resolution scale is $\sim$0.17 Mpc, we must content ourselves with a flatter cell of the same volume with which to tile our box: a 0.03 Mpc^2 FOV with a LOS distance of 50 Mpc.

[10] Note that this tiling procedure does not precisely mimic the survey, because we should ideally select volumes without tiling beforehand. Obtaining accurate statistics by randomly sampling volumes of our simulation box would be computationally prohibitive at small source number densities.
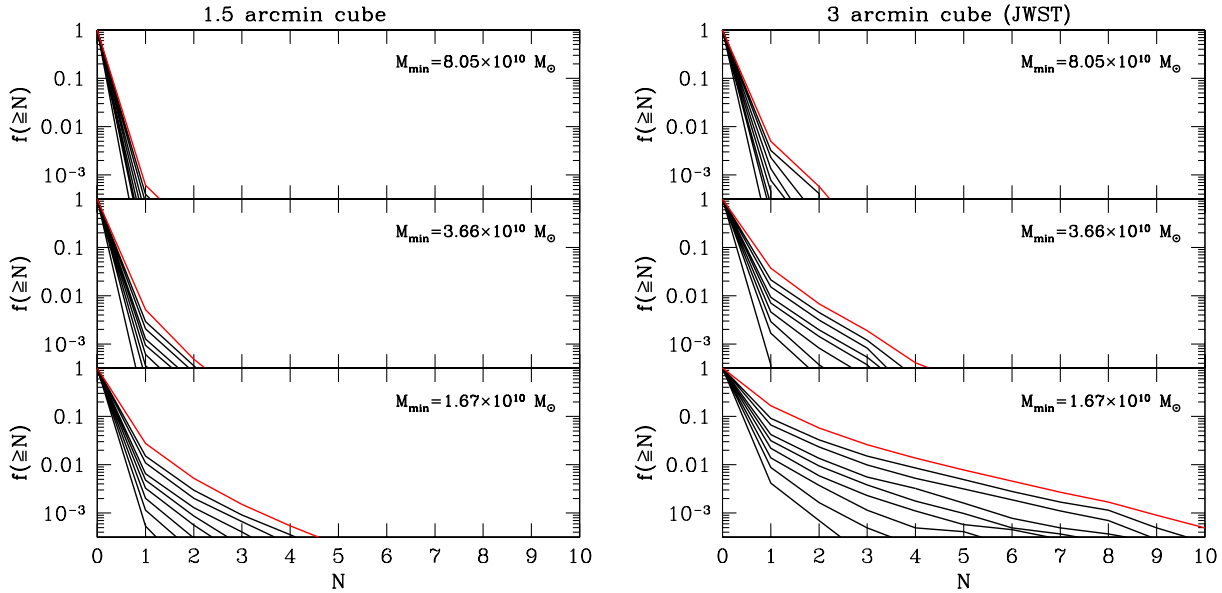
Figure 7. Fraction of mock-survey fields containing $\geq N$ number of LAE. Curves correspond to $\bar{x}_{HI} \approx 0$, 0.46, 0.56, 0.67, 0.72, 0.77, 0.83, 0.88, 0.93 (*top to bottom*), assuming $M_{\min} = 8.06 \times 10^{10}$, 3.68×10^{10}, and 1.67×10^{10} $M_\odot$ in the sub-panels (*top to bottom*). Panels assume FoVs of: $1.5' \times 1.5'$ (*left panel*) and $3' \times 3'$ (*right panel*).

a decent fit to $z = 5.7$ LAE luminosity functions, requires quite massive halos and drastically overestimates the observed $z \sim 9$ abundances (overestimating the mass function by a factor of ~ 400 at the required masses). It is therefore inconsistent with even a completely ionized universe. Note however that since our estimates are based on the limited volume of our 250 Mpc^3 simulation box, we might be slightly underestimating the cosmic variance contribution to these error bars.

One possible remedy is if our background cosmology is wrong. The dotted curve shows the mass function if $\bar{x}_{HI} \approx 0$, but with $\sigma_8 = 0.86$, as the combined three–year *WMAP* and Lyman–α forest data prefer (Lewis 2006). This higher value of $\sigma_8 = 0.86$ increases the abundance by a factor of several at the high-mass end, but it has only a modest effect in the range of interest. The observed abundances in our model (i) are still inconsistent with a $\bar{x}_{HI} \approx 0$ universe, overestimating the mass function by a factor of ~ 100.

On a more fundamental level, the Stark et al. (2007) LAE sample, if genuine, *requires* substantial star formation in halos with $M \lesssim 10^9 M_\odot$. This is only a factor ~ 10 greater than the atomic cooling threshold at $z \sim 9$, and these objects would correspond to some of the smallest galaxies observed at *any* redshift. They must also be fundamentally different from sources at $z = 5.7$, with either much higher star formation rates or stars that are much more efficient at producing ionizing photons. Unless the escape fraction of UV photons is extremely small, any scenario that is consistent with the observations would produce several ionizing photons per IGM baryon, so we would expect reionization to have been well underway (if not complete) by $z = 9$. Needless to say, such a scenario would be difficult to reconcile with reionization at $z \sim 6$–7.

To this point, we have assumed that *all* galaxies above $M_{\min}$ are LAEs. However, at moderate redshifts, we know that a majority of Lyman-break galaxies (LBGs) have weak or absent Lyαemission lines (e.g. Shapley et al. 2003), though there is some evidence of an increase in the fraction of LAEs among LBGs at higher redshifts (Dawson et al. 2004; Hu et al. 2004; Shimasaku et al. 2006).

Incorporating this into Figure 6 would be equivalent to shifting the curves downward by an amount equal to the fraction of all galaxies that are LAEs. Thus this only strengthens our arguments, because it *widens* the disparity between the observed number density of sources and the theoretical curves.

5   COUNTS-IN-CELLS STATISTICS

LAE abundances and luminosity functions have certainly proven to be very useful in studying the $z \sim 6$ universe. However, their interpretation is inevitably controversial, because knowledge of the $M \leftrightarrow L$ mapping is required for a meaningful estimate of $\bar{x}_{HI}$. We have already seen how difficult it is to place meaningful constraints with this method.

It is therefore worth considering other, more robust, signatures. Reionization modulates the observed LAE field, increasing the clustering of sources (Furlanetto et al. 2004, 2006; McQuinn et al. 2007). Galaxies inside large HII bubbles are more likely to be seen than those inside smaller bubbles. Thus, during reionization, we expect a field of view (FoV) that contains a LAE to have a *higher* probability of containing *another* LAE, than would be the case after reionization. This reionization-induced clustering is illustrated in Figure 1 and can be quantified in any number of ways. To date, studies have focused on the linear bias and power spectrum (Furlanetto et al. 2006; McQuinn et al. 2007). However, the modulation is non-gaussian, so other statistics may be as powerful.

Here we explore simple, statistical estimates of "counts-in-cells" that can be easily applied to future surveys. These simple number counts essentially represent an "integrated" measurement of the clustering and hence can be easier to generate with a limited observational sample than other statistical indicators such as the power spectrum/correlation function, which try to measure the detailed scale dependence. They also place less stringent require-

ments on the survey strategy than, for example, the power spectrum, because they are easier to interpret with non-uniform survey coverage (for example, following up bright sources to search for fainter neighbors; see below). They are therefore most likely to be powerful early in reionization or at extremely high redshifts, when sources are rare.

To study this reionization-induced clustering, we perform a mock-survey in our simulation box. We tile our simulation box and tabulate "in-cell" number counts, given $\bar{x}_{\rm HI}$ and $M_{\rm min}$. For simplicity, we assume the survey FoVs are cubical volumes, but our results can easily be extended to more detailed survey specifications, once they are available. Note also that cell sizes are fairly arbitrary, as surveys can be broken down into small cells for analysis, and the optimal choice will depend on the survey depth and volume. In Figure 7, we present the fraction of our mock-survey fields containing $\geq N$ LAEs. The panels are constructed assuming FoVs of $1.5' \times 1.5'$ (*left panel*) and $3' \times 3'$ (*right panel*), corresponding to comoving sizes of 4.23^3 and 8.45^3 Mpc^3, respectively. The latter (in angular coordinates) is the FoV of the Near-Infrared Spectrograph (NIRSpec) on *JWST* (Gardner et al. 2006); other future IR instruments subtend comparable areas (for example, the Infrared Multi-Object Spectrograph (IRMOS) on the proposed TMT has a $5' \times 5'$ FoV[11]). Curves correspond to $\bar{x}_{\rm HI} \approx 0$, 0.46, 0.56, 0.67, 0.72, 0.77, 0.83, 0.88, 0.93 (*top to bottom*), assuming $M_{\rm min} = 8.06 \times 10^{10}$, 3.68×10^{10}, and 1.67×10^{10} $M_\odot$ in the sub-panels (*top to bottom*). In Figure 8, we show the ratio of the PDF curves from the bottom right panel in Fig. 7 to those expected from a pure Poisson distribution. Note the relatively small change in the probabilities during the later stages of reionization, from $\bar{x}_{\rm HI} = 0.5$ to $\bar{x}_{\rm HI} \approx 0$, compared to the rapid evolution during the initial stages. This is, of course, primarily because the strong damping-wing absorption causes most of the sources to fall below the survey detection threshold before $\bar{x}_{\rm HI} \sim 0.5$.

In other words, because the curves shown in Fig. 7 correspond to raw number counts, they include two distinct consequences of higher values of $\bar{x}_{\rm HI}$: (i) the enhanced clustering footprint of reionization; and (ii) the decrease in the absolute number of observed LAEs. As is the case for luminosity functions, (ii) is difficult to disentangle from astrophysical uncertainties and hence is not an ideal probe of reionization. The first offers more promise, as the enhanced clustering of LAEs from reionization could only be mimicked by a large shift in the masses of underlying halos hosting LAEs, as discussed below. Thus we wish to isolate the first effect.

To this end, we note that the variance of the in-cell count distributions can be represented as (e.g. Peebles 1980):

$$\sigma_N^2 = \langle N \rangle + \left(\frac{\langle N \rangle}{V}\right)^2 \int_V \int_V dV_1 dV_2 \, \xi_{12} \, , \qquad (7)$$

where $\langle N \rangle$ is the mean number of LAEs in a cell (i.e. field), and the integrals over the two-point correlation function, ξ_{12}, are performed over the cell volume, V. Keeping this in mind, in Figure 9, we plot:

$$\bar{\xi}_{12} \equiv \left(\sigma_N^2 - \langle N \rangle\right) / \langle N \rangle^2 \, , \qquad (8)$$

essentially the average value of the correlation function over a cell's volume. Curves correspond to $M_{\rm min} = 3.68 \times 10^{10}$ $M_\odot$ (*dotted*), and 1.67×10^{10} $M_\odot$ (*short-dashed*).

If sources are uncorrelated (i.e. Poisson distributed), $\sigma_N^2 =$

---

[11] http://www.tmt.org
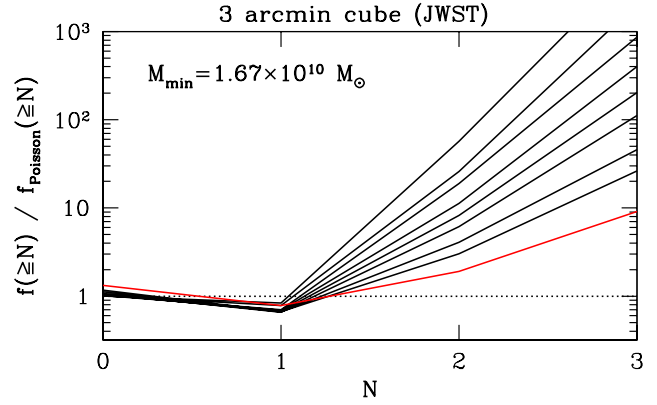
Figure 8. Ratio of the PDF curves from the bottom right panel in Fig. 7 (in reverse order, *bottom to top*) to those expected from a pure Poisson distribution.

$\langle N \rangle$, and $\bar{\xi}_{12} = 0$. The biased nature of structure formation manifests as a constant, positive $\bar{\xi}_{12}$ (because we hold redshift, and hence the halo population, fixed), and reionization manifests as an increase in $\bar{\xi}_{12}$. Shallower surveys with larger $M_{\rm min}$ have a larger $\bar{\xi}_{12}$, because these rarer objects are more highly clustered intrinsically. Interestingly, the curves are fairly flat for small neutral fractions; this is because nearly all of the objects are visible, so the observed clustering nearly matches its intrinsic value. Reionization induces a sharp rise in $\bar{\xi}_{12}$ at $\bar{x}_{\rm HI} \sim 0.5$, with the location of the rise being a very weak function of the cell size and $M_{\rm min}$. This increase is a result of the "small-scale" clustering enhancement described by Furlanetto et al. (2006), which occurs because only the sources inside rare large bubbles remain visible early in reionization, but at the same time a large fraction of sources inside such bubbles are visible. The enhancement is nearly independent of the underlying halo mass.

The long-dashed curve in the right panel of Fig. 9 shows the same quantity as the short-dashed curve, but in a scenario in which $\langle N \rangle$ is held constant regardless of the neutral fraction (at the number density found with $\bar{x}_{\rm HI} = 0.77$) by randomly selecting halos above the $\bar{x}_{\rm HI}$-dependent mass threshold. The fact that the long and short dashed curves overlap illustrates explicitly that we have accurately removed the Poisson component of the fluctuations from our statistic.

The dot-dashed curve in the right panel of Fig. 9 is generated by selecting only the most massive halos with number density fixed by the value at $\bar{x}_{\rm HI} = 0.77$ for $M_{\rm min} = 1.67 \times 10^{10}$ $M_\odot$ (unlike the random selection performed for the long-dashed curve). The curve is flat at small $\bar{x}_{\rm HI}$ since it corresponds to the same set of massive sources. The curve then increases at $\bar{x}_{\rm HI} \gtrsim 0.6$, when ionized regions are small enough to make some of these highly-biased, massive sources fall below the the $M_{\rm min} = 1.67 \times 10^{10}$ $M_\odot$ detection threshold. We see that the reionization-induced rise in $\bar{\xi}_{12}$ surpasses the intrinsic clustering of even the most massive sources with the same number density at $\bar{x}_{\rm HI} \gtrsim 0.6$. Thus there is no question that reionization can be observed through clustering measurements, at least sufficiently early in the process. Hence, counts-in-cells can provide a simple, robust probe of reionization.

Detecting the sharp rises evident in Fig. 9 could be a "smoking-gun" signature of reionization; however, at any particular redshift we have only one measurement and (as with the luminosity function) can only compare with measurements at dif-

ferent redshifts.[12]. The fact that the curves in Fig. 9 are strong functions of $M_{\rm min}$ complicates the interpretation of any such detection, because any increase in $\bar{\xi}_{12}$ with redshift could be because of galaxy evolution alone; for example, even if $M_{\rm min}$ remains constant with redshift, the bias of the objects would still increase with redshift. This degeneracy can be overcome if reionization progresses rapidly compared to the underlying structure, or if one correlates the LAE field with a Lyman-break galaxy (LBG) field, since the detectability of LBGs should not be affected by changes in $\bar{x}_{\rm HI}$ (see McQuinn et al. 2007 for discussion of similar issues with reference to the power spectrum).

Unfortunately, such a process is always uncertain. Instead we favor measuring the excess probability (over that in a completely ionized universe) that a cell will contain N or more LAEs, *given that* it already contains at least one LAE:

$$\Delta P_{\bar{x}_{\rm HI}}(\geq N | \geq 1) = P_{\bar{x}_{\rm HI}}(\geq N | \geq 1) - P_{\bar{x}_{\rm HI} \approx 0}(\geq N | \geq 1), (9)$$

where $P(\geq N | \geq 1)$ and $P_{\bar{x}_{\rm HI} \approx 0}(\geq N | \geq 1)$ are the probabilities that a cell containing at least one LAE will contain N or more LAEs, in a partially ionized and a fully ionized universe, respectively, *normalized to have the same number density of LAEs*. We normalize our $\bar{x}_{\rm HI} \approx 0$ field by randomly choosing LAEs above $M_{\rm min}$, until we obtain the same number of LAEs as in the partially ionized box. Note, however, that the proper normalization procedure is not well-defined so long as the mapping between mass and luminosity remains unknown. However, we checked that selecting only the most massive halos, until we obtain the same number density, did not appreciably change the results for the $N = 2$ curves.

The basic idea of this measure is to remove the overall normalization of the galaxy number density, whose evolution is uncertain, but to retain the enhanced probability of observing groups of sources inside the rare large bubbles. Also note that while the $\bar{\xi}_{12}$ statistic in eq. (8) only uses second-order correlations and can be obtained by integrating over the power spectrum, the counts-in-cells method takes advantage of higher-order correlations and the excess probability from eq. (9) shows this explicitly for $N > 2$.

In Figure 10, we plot the excess probability from eq. (9), for $1.5' \times 1.5'$ (*left panel*) and $3' \times 3'$ (*right panel*) cells. Solid (dotted) curves assume $M_{\rm min} = 1.67 \times 10^{10}$ (3.68×10^{10}) $M_\odot$. Thick and thin curves correspond to $N = 2$ and 3, respectively.

The enhanced clustering footprint of reionization can easily be seen from Fig. 10. For $\bar{x}_{\rm HI} \gtrsim 0.5$ and $N = 2$, $\Delta P \gtrsim 10\%$, and it increases by a factor of $\gtrsim 6$ from $\bar{x}_{\rm HI} = 0.2$ to 0.8. Furthermore, the $N = 2$ curves are much more sensitive to $\bar{x}_{\rm HI}$ than to $M_{\rm min}$, suggesting that this statistic of the reionization-induced clustering *cannot* be mimicked by a change in the mass threshold of the survey, or analogously by a shift in the underlying halo masses hosting LAEs. Hence, the excess probability that a cell containing LAEs contains more than one LAE is a robust indicator of changes in $\bar{x}_{\rm HI}$. The $N = 3$ curves do fall significantly for the most massive objects (as do all the curves for $\bar{x}_{\rm HI} \approx 1$); this is simply because the sources are too rare (see Fig. 10). The apparent drop-off at $\bar{x}_{\rm HI} \approx 0.9$ is an artifact of our finite box size.

---

[12] We remind the reader that our statistics are generated with the same intrinsic source field, i.e. at a fixed redshift $z = 9$. Since present-day simulations cannot accurately simulate the redshift evolution of $\bar{x}_{\rm HI}$, this is the cleanest way of extracting statistics on reionization, especially given that the ionization topology at fixed $\bar{x}_{\rm HI}$ is almost independent of redshift in this range (McQuinn et al. 2007). Unfortunately, the real Universe is uncooperative on this point, and thus observations of the different stages of reionization must necessarily be from different redshifts.

We attempt to approximately remove the Poisson component of the clustering statistics from the excess probability by subtracting the second term in eq. (9). However, it is not immediately obvious that this statistic is completely independent of the LAE number density. Hence it is intriguing to probe the robustness of the seeming overlap of the $N = 2$ curves of different mass scales in Fig. 10. To this end, we recreated the $M_{\rm min} = 1.67 \times 10^{10}$ $M_\odot$, $N = 2$ curve from Fig. 10 for $\bar{x}_{\rm HI} \leq 0.77$, by randomly choosing halos in order to keep the number density constant (i.e. equal to the number density at $\bar{x}_{\rm HI} = 0.77$, as we had for the long-dashed curve in the right panel of Fig. 9). We find that the excess probability remains fairly unchanged at $\bar{x}_{\rm HI} \gtrsim 0.5$, seemingly suggesting that our ΔP statistic is robust (not very sensitive to the LAE number density). The excess probability falls off more rapidly at $\bar{x}_{\rm HI} \lesssim 0.5$ (where the excess clustering signature is weaker) than in our Figure 10, though it is not clear if this trend is statistically significant since the error bars become large in this regime. Thus, our proposed ΔP statistic does appear to be relatively robust to the intrinsic number density, although the particularly close match in the $N = 2$ curves of Figure 10 may be somewhat coincidental.

In Figure 11, we show the survey characteristics required to detect this excess probability, assuming $N = 2$ and $3'$ cubical cells. Specifically, we require $\Delta P - n\sigma > 0$ for a n-σ detection, where ΔP is our derived value from Fig. 10 and σ is the uncertainty on the measured value of ΔP in the given survey volume. Solid and dotted curves correspond to $M_{\rm min} = 1.67 \times 10^{10}$ and 3.68×10^{10} $M_\odot$, respectively. In the bottom panel, we show the total number of cells containing galaxies that must be observed in order to detect the effect at the 3-σ (*top curve*) and 2-σ (*bottom curve*) level. It is nearly independent of halo mass at higher neutral fractions, because this integrated clustering measure depends only weakly on the characteristics of the underlying halo population (see the discussion above). Interestingly, a reasonably strong detection requires only several tens of galaxy detections, with the required source count decreasing as $\bar{x}_{\rm HI}$ increases. This is because ΔP is large compared to the raw probability of detecting two neighboring galaxies and increases as the ionized regions get smaller. This indicates the power of the counts-in-cell approach: the actual number density of objects decreases by ~ 5 over this range, but this is compensated by the increased probability. Many other approaches to clustering, such as the power spectrum, would lose sensitivity in this range because of the rarity of the sources.

The top panel shows the corresponding survey volumes that must be observed to detect this many galaxies. (Of course, this increases rapidly with the mass threshold, even though ΔP does not, because the probability of having $N = 1$ is much smaller for rarer sources.) We can see that a survey volume of 6 (12) $\times 10^5$ Mpc^3 is required to detect the enhanced clustering at $\bar{x}_{\rm HI} \sim 0.5$–$0.8$, with a 2-$\sigma$ (3-σ) accuracy and $M_{\rm min} = 1.67 \times 10^{10}$ $M_\odot$ sensitivity. If the survey depth is only one cell deep, this corresponds to a several square degree survey at least. A deep, blind survey subtending such a region may be difficult in the foreseeable future, given the modest FoVs of forthcoming near-IR instruments. However, the modest number of cells that actually contain sources suggests that a shallow but wide survey to identify extremely bright candidates, followed by deep followup with a large telescope to search for fainter neighbors, may be a viable strategy. Moreover, if the Stark et al. (2007) candidates are truly at $z = 9$ and correspond to halos with $M \lesssim 10^9 M_\odot$, detecting large numbers of sources may be much easier than our conservative estimates suggest. This is one advantage of counts-in-cells over the power spectrum, which is more model-dependent in such unconventional survey strategies.
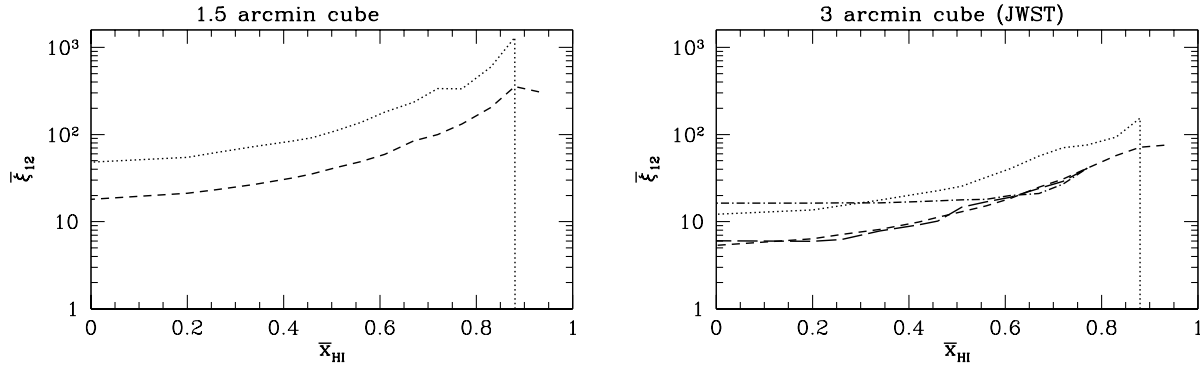
Figure 9. The average two-point correlation function from eq. (8) as functions of $\bar{x}_{HI}$. Curves correspond to $3.68 \times 10^{10}\ M_\odot$ (*dotted*), and $1.67 \times 10^{10}\ M_\odot$ (*short-dashed*). The long-dashed curve in the right panel shows the same quantity as the short-dashed curve, but in a scenario in which $\langle N \rangle$ is held constant at the number density found with $\bar{x}_{HI} = 0.77$. The dot-dashed curve in the right panel is generated assuming $M_{min} = 1.67 \times 10^{10}\ M_\odot$, but by selecting only the most massive halos with number density fixed by the value at $\bar{x}_{HI} = 0.77$.
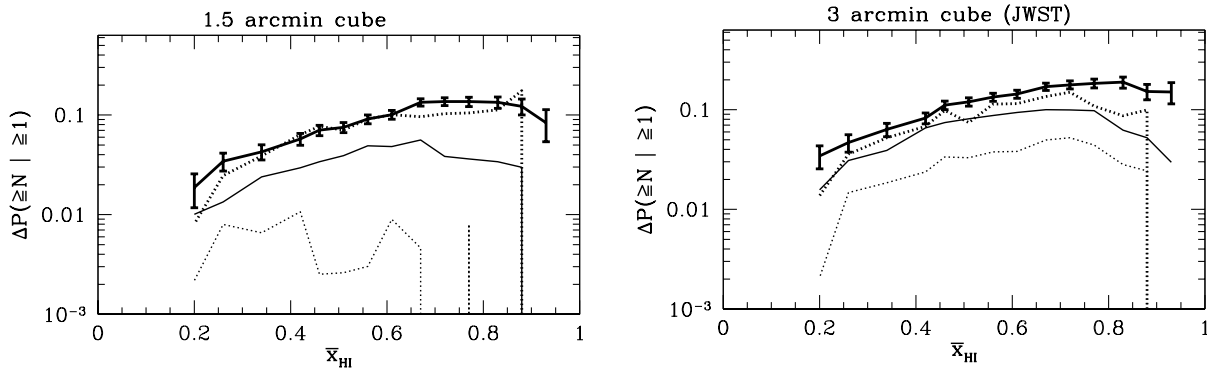
Figure 10. Excess probability (over that in an ionized universe normalized to the same number density of LAEs) that a FoV containing at least one LAE will contain N or more LAEs. Panels correspond to FoV of $1.5' \times 1.5'$ (*left*) and $3' \times 3'$ (*right*). Solid and dotted curves assume $M_{min} = 1.67 \times 10^{10}$ and $3.68 \times 10^{10}\ M_\odot$, respectively. Thick and thin curves correspond to $N = 2$ and 3, respectively. Error bars indicate the 1-σ Poisson uncertainty on the $N = 2$, $M_{min} = 1.67 \times 10^{10}\ M_{min}$ curves.

Throughout our discussion, we have used somewhat arbitrary cell sizes, as surveys can be broken down into small cells for analysis, and the optimal choice will depend on the characteristics of the particular survey. However, in the absence of spectroscopic follow-up, the LAE redshift might only be localized to the width of a narrow band filter. This sets a minimum cell size in the LOS direction. With this in mind, in Fig. 12, we plot the same in-cell statistics as in Fig. 11, but extending the LOS axis of the cells to correspond to a narrow band filter with $R = \lambda/\Delta\lambda \sim 100$. Comparing to Fig. 11, the required number of cells and survey volume increases by less than a factor of ~ 2. Essentially, so long as the cells are not so long that random, distant neighbors overwhelm the reionization clustering, cells of any radial size can be used without a substantial increase in the survey requirements.

As mentioned previously, counts-in-cells (and specifically our ΔP statistic) can make use of higher-order correlations. In Fig. 13, we plot the same statistics as in Fig. 11, but using clustering of the $N = 3$ term. Interestingly, at least for low-mass halos, measuring this term is no more difficult than measuring the two-point statistic. These higher-order correlations have not yet been quantified in the context of reionization, so we also show (with the dot-dashed curve) the requirements (assuming $M_{min} = 1.67 \times 10^{10}\ M_\odot$) to

detect three-point clustering *without* reionization, assuming that a survey probes all halos above a given mass threshold, normalized to the same number density. (In other words, this is the requirement to detect three point clustering from just gravitational instability; for this nearly gaussian random field it contains little more information than the power spectrum itself.) The $N = 3$ term is much weaker in this case and requires a considerably larger survey to measure. Thus, detection of this three-point term may provide powerful support for any detection of reionization-induced clustering.

Note that we have ignored foreground contamination here. This is likely a serious issue in any survey, but the details of foreground removal depend on the survey specifics. Hence we defer it to future works, more focused on particular surveys.

As mentioned previously, stochasticity in the $L \leftrightarrow M$ mapping, and secondary effects on the detectability of the Lyα line, such as galactic winds and gas infall, can in principle affect the details of our counts in-cell estimates above. We defer a thorough study of such effects to future works involving numerical simulations.
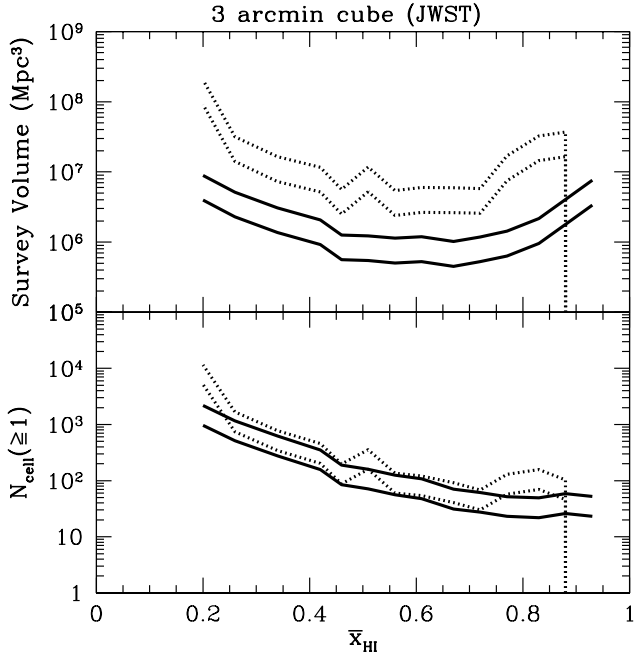
Figure 11. Survey characteristics required to detect the excess clustering probability due to reionization (eq. 9) assuming $M_{\min} = 1.67 \times 10^{10}$ $M_\odot$ (*solid curves*) and 3.68×10^{10} $M_\odot$ (*dotted curves*), and $N = 2$. The top panel shows the survey volume required for 3-σ (*top curve*) and 2-σ (*bottom curve*) detections; the bottom panel shows the corresponding number of cells that actually contain galaxies.
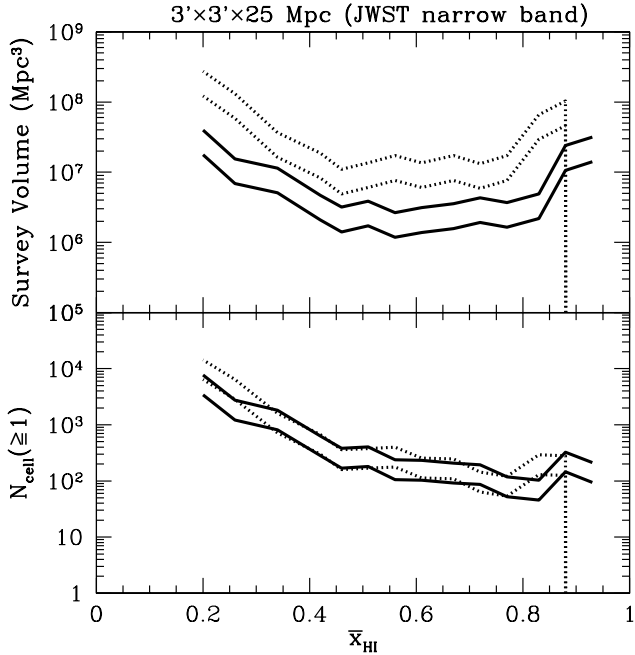
Figure 12. Same as Fig. 11, but with a cell size typical of narrow band LAE surveys, with $R = \lambda/\Delta\lambda \sim 100$.
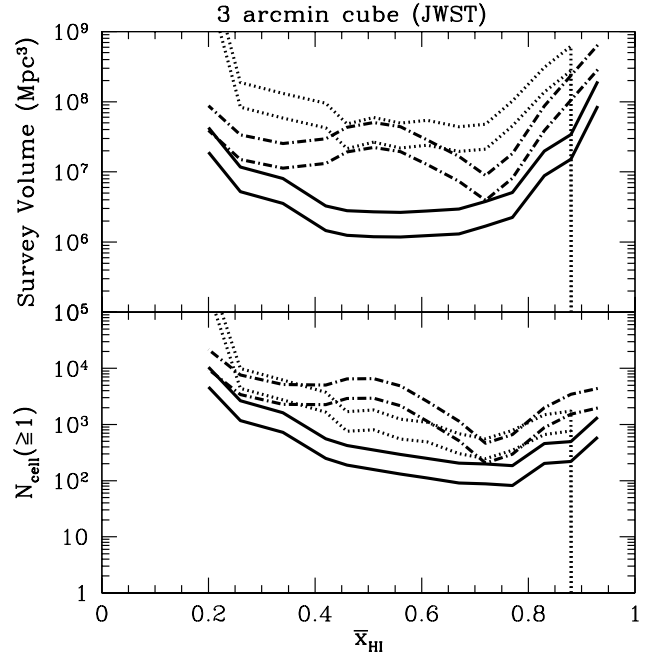
Figure 13. Same as Fig. 11, but with $N = 3$. The dot-dashed curve is generated assuming $M_{\min} = 1.67 \times 10^{10}$ $M_\odot$, but by selecting only the most massive halos, normalized to the same number density at each $\bar{x}_{\mathrm{HI}}$.

6 CONCLUSIONS

In this work, we investigate the use of LAEs in constraining reionization at $z \sim 9$. We generate $z = 9$ halo, velocity and density fields in a 250 Mpc "semi-numerical" simulation box (Mesinger & Furlanetto 2007a). Our excursion-set approach allows us to resolve halos with masses $M \gtrsim 2.2 \times 10^8 M_\odot$. We construct ionization topologies corresponding to various values of $\bar{x}_{\mathrm{HI}}$.

As a first step, we generate damping wing Lyα optical depth distributions for the halos in our simulations. As expected, the more massive halos have narrower distributions with smaller mean absorption, though the distributions become weaker functions of halo mass as reionization progresses. We show that these distributions are roughly lognormal but are broader and have a smaller mean than purely analytic predictions.

Using the τ_D distributions, we generate $z = 9$ LAE luminosity functions for various values of $\bar{x}_{\mathrm{HI}}$. Constraining $\bar{x}_{\mathrm{HI}}$ with luminosity functions is difficult due to the many uncertainties inherent in the host halo mass $\leftrightarrow$ Lyα luminosity mapping. However, we show that applying a very conservative mapping to the number densities of the Stark et al. (2007) sample yields $\bar{x}_{\mathrm{HI}}(z = 9) \lesssim 0.7$. More fundamentally, these LAE number densities, if genuine, *require* substantial star formation in halos with $M \lesssim 10^9 M_\odot$, making them unique among the current sample of observed high-z objects.

The topology of reionization increases the apparent clustering of the observed LAEs, aside from merely suppressing their number densities. We investigate the detectability of this signature using "counts-in-cell" statistics, which are more robust than the power-spectrum at studying such non-gaussian fields and (as integrated measures) are more straightforward to interpret in the few source limit. We find that the likelihood of observing *more than one* LAE among the subset of fields which contain LAEs is $\gtrsim$10% greater in a universe with $\bar{x}_{\mathrm{HI}} \gtrsim 0.5$ than in an ionized universe with the

same LAE number density. We show that this effect can be detected at $z \sim 9$ with just a few tens of $3'$ cubical cells containing galaxies, regardless of the underlying host halo mass.

Counts-in-cells is only one approach to clustering, and it has advantages and disadvantages compared to the more common power spectrum (or correlation function) approach (McQuinn et al. 2007). The latter accounts for the detailed scale dependence of the clustering enhancement (and so in principle provides information on the ionization field; Furlanetto et al. 2006) but does not include higher-order corrections to the clustering. Moreover, robust power spectrum measurements require a large number of sources to be detected and are relatively unforgiving of ad hoc survey strategies, such as deep followup, which may be required when sources are rare. The small FoVs of planned near-IR instruments therefore make such measurements difficult at high redshifts.

By contrast, our counts-in-cells approach offers very little information on the scale dependence of the ionization field. However, it does include non-gaussianities, which become important early in reionization ($\bar{x}_{HI} \gtrsim 0.5$). We explicitly showed, for the first time, that reionization induces non-gaussianities in the galaxy distribution that should be separable from structure formation, at least in the deep survey limit. As a result, our signature becomes more powerful earlier in the reionization process, more than compensating for the declining apparent number density of LAEs. It also places very few constraints on the survey geometry, because the fields need not be contiguous (and in fact, to compensate for cosmic variance, probably should not be). An ideal strategy may be to identify particularly bright candidate objects with a wide, shallow survey. Then, one can follow up these candidates with deeper integrations to confirm their identity and search for fainter neighbors. Given that only a few tens of sources must be followed-up to provide interesting constraints, such a strategy would require relatively modest telescope resources.

REFERENCES

Barton E. J., Davé R., Smith J.-D. T., Papovich C., Hernquist L., Springel V., 2004, ApJ, 604, L1

Becker G. D., Rauch M., Sargent W. L. W., 2007, ApJ, 662, 72

Bolton J. S., Haehnelt M. G., 2007, ArXiv e-prints, 705

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

Bond J. R., Myers S. T., 1996, ApJS, 103, 1

Cuby J.-G., Hibon P., Lidman C., Le Fèvre O., Gilmozzi R., Moorwood A., van der Werf P., 2007, A&A, 461, 911

Dawson S., Rhoads J. E., Malhotra S., Stern D., Dey A., Spinrad H., Jannuzi B. T., Wang J., Landes E., 2004, ApJ, 617, 707

Dawson S., Rhoads J. E., Malhotra S., Stern D., Wang J., Dey A., Spinrad H., Jannuzi B. T., 2007, ArXiv e-prints, 707

Dijkstra M., Wyithe S., Haiman Z., 2006, ArXiv Astrophysics e-prints, astro-ph/0611195

Efstathiou G., 1992, MNRAS, 256, 43P

Fan X., Carilli C. L., Keating B., 2006, ARAA, 44, 415

Fan X., Strauss M. A., Becker R. H., White R. L., Gunn J. E., Knapp G. R., Richards G. T., Schneider D. P., Brinkmann J., Fukugita M., 2006, AJ, 132, 117

Fernandez E. R., Komatsu E., 2007, ArXiv e-prints, 706

Furlanetto S. R., Hernquist L., Zaldarriaga M., 2004, MNRAS, 354, 695

Furlanetto S. R., McQuinn M., Hernquist L., 2006, MNRAS, 365, 115

Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, ApJ, 613, 1

Furlanetto S. R., Zaldarriaga M., Hernquist L., 2006, MNRAS, 365, 1012

Gardner J. P., et al., 2006, Space Science Reviews, 123, 485

Gnedin N. Y., 2000, ApJ, 542, 535

Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633

Haiman Z., 2002, ApJ, 576, L1

Haiman Z., Cen R., 2005, ApJ, 623, 627

Haiman Z., Spaans M., 1999, ApJ, 518, 138

Hu E. M., Cowie L. L., Capak P., McMahon R. G., Hayashino T., Komiyama Y., 2004, AJ, 127, 563

Hu E. M., et al., 2002, ApJ, 568, L75

Iye M., et al., 2006, Nature, 443, 186

Kashikawa N., Shimasaku K., Malkan M. A., Doi M., Matsuda Y., Ouchi M., Taniguchi Y., Ly C., Nagao T., Iye M., Motohara K., Murayama T., Murozono K., Nariai K., Ohta K., Okamura S., Sasaki T., Shioya Y., Umemura M., 2006, ApJ, 648, 7

Kobayashi M. A. R., Totani T., Nagashima M., 2007, ArXiv e-prints, 705

Kodaira K., et al., 2003, PASJ, 55, L17

Lacey C., Cole S., 1993, MNRAS, 262, 627

Le Delliou M., Lacey C. G., Baugh C. M., Morris S. L., 2006, MNRAS, 365, 712

Lewis A., 2006, ArXiv Astrophysics e-prints

Lidz A., McQuinn M., Zaldarriaga M., Hernquist L., Dutta S., 2007, ArXiv Astrophysics e-prints

Lidz A., Oh S. P., Furlanetto S. R., 2006, ApJ, 639, L47

Madau P., Rees M. J., 2000, ApJ, 542, L69

Malhotra S., Rhoads J. E., 2004, ApJ, 617, L5

Malhotra S., Rhoads J. E., 2006, ApJ, 647, L95

Maselli A., Gallerani S., Ferrara A., Choudhury T. R., 2007, MNRAS, 376, L34

McQuinn M., Hernquist L., Zaldarriaga M., Dutta S., 2007, ArXiv e-prints 0704.2239, 704

McQuinn M., Lidz A., Zahn O., Dutta S., Hernquist L., Zaldarriaga M., 2007, MNRAS, 377, 1043

McQuinn M., Lidz A., Zaldarriaga M., Hernquist L., Dutta S., 2007, submitted to MNRAS [arXiv.org/0710.1018 (astro-ph)], 710

Mesinger A., Furlanetto S., 2007a, ApJ, in press, ArXiv e-prints 0704.0946, 704

Mesinger A., Furlanetto S., 2007b, MNRAS, in press [arXiv.org/0710.0371 (astro-ph)], 710

Mesinger A., Haiman Z., 2004, ApJ, 611, L69

Mesinger A., Haiman Z., 2007, ApJ, 660, 923

Miralda-Escudé J., 1998, ApJ, 501, 15

Miralda-Escude J., 1998, ApJ, 501, 15

Osterbrock D. E., 1989, Astrophysics of gaseous nebulae and active galactic nuclei. Research supported by the University of California, John Simon Guggenheim Memorial Foundation, University of Minnesota, et al. Mill Valley, CA, University Science Books, 1989, 422 p.

Ota K., et al., 2007, ArXiv e-prints, 707

Page L., et al., 2006, ArXiv Astrophysics e-prints astro-ph/0603450

Partridge R. B., Peebles P. J. E., 1967, ApJ, 147, 868

Peebles P. J. E., 1980, The large-scale structure of the universe. Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p.

Press W. H., Schechter P., 1974, ApJ, 187, 425

Rhoads J. E., et al., 2004, ApJ, 611, 59

Santos M. R., 2004, MNRAS, 349, 1137

Santos M. R., et al., 2004, ApJ, 606, 683

Schaerer D., 2003, A&A, 397, 527

Shapiro P. R., Giroux M. L., Babul A., 1994, ApJ, 427, 25

Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., 2003, ApJ, 588, 65

Shimasaku K., Kashikawa N., Doi M., Ly C., Malkan M. A., Matsuda Y., Ouchi M., Hayashino T., Iye M., Motohara K., Murayama T., Nagao T., Ohta K., Okamura S., Sasaki T., Shioya Y., Taniguchi Y., 2006, PASJ, 58, 313

Spergel D. N., et al., 2006, ApJ, submitted, preprint astro-ph/0603449

Stanway E. R., et al., 2004, ApJ, 604, L13

Stark D. P., Ellis R. S., Richard J., Kneib J.-P., Smith G. P., Santos M. R., 2007, ApJ, 663, 10

Stark D. P., Loeb A., Ellis R. S., 2007, ArXiv Astrophysics e-prints, astro-ph/0701882

Taniguchi Y., et al., 2005, PASJ, 57, 165

Thoul A. A., Weinberg D. H., 1996, ApJ, 465, 608

Willis J. P., Courbin F., 2005, MNRAS, 357, 1348

Zahn O., Lidz A., McQuinn M., Dutta S., Hernquist L., Zaldarriaga M., Furlanetto S. R., 2007, ApJ, 654, 12

5 arcmin cube (TMT)