

# The inhomogeneous ionizing background following reionization

Andrei Mesinger<sup>1</sup>\*† and Steven Furlanetto<sup>2</sup>

<sup>1</sup>*Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA*

<sup>2</sup>*Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA*

Accepted 2009 August 13. Received 2009 July 23; in original form 2009 June 19

## ABSTRACT

We study the spatial fluctuations in the hydrogen-ionizing background in the epoch following reionization ( $z \sim 5\text{--}6$ ). The rapid decrease with redshift in the photon mean free path (mfp), combined with the clustering of increasingly rare ionizing sources, can result in a very inhomogeneous ionizing background during this epoch. We systematically investigate the probability density functions (PDFs) and power spectra of ionizing flux, by varying several parameters such as the mfp, minimum halo mass capable of hosting stars and halo duty cycle. In order to be versatile, we make use of analytic, seminumeric and numeric approaches. Our models show that the ionizing background indeed has sizable fluctuations during this epoch, with the PDFs being a factor of few wide at half of the maximum likelihood. The clustering of sources dominates the width of the PDFs, so analytic models must take large-scale clustering into account. The distributions also show marked asymmetries, with a high-value tail set by clustering on small scales, and a shorter low-value tail which is set by the mfp. The power spectrum of the ionizing background is much more sensitive to source properties than the PDF and can be well understood analytically with a framework similar to the halo model (usually used to describe dark matter clustering). Nevertheless, we find that Ly $\alpha$  forest spectra are extremely insensitive to the details of the ultraviolet background, despite marked differences in the PDFs and power spectra of our various ionizing backgrounds. Assuming a uniform ionizing background only underestimates the value of the mean ionization rate,  $\langle \Gamma_{12} \rangle$ , inferred from the Ly $\alpha$  forest by a few per cent. Instead, analysis of the Ly $\alpha$  forest is dominated by the uncertainties in the density field. Thus, our results justify the common assumption of a uniform ionizing background in Ly $\alpha$  forest analysis even during this epoch.

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

## 1 INTRODUCTION

The metagalactic hydrogen-ionizing ultraviolet background (UVB) is probably the most influential, pervasive radiation field throughout most of cosmic history. Much investigation has gone into understanding its properties and its impact on the intergalactic medium (IGM; see Meiksin 2007 for a recent review). At low redshifts ( $z \lesssim 3\text{--}4$ ), the ionizing photon mean free path (mfp) is very large, and so each gas element effectively ‘sees’ many sources of ionizing radiation. A uniform UVB seems to be an accurate approximation in this regime, with spatial fluctuations contributing only a few per cent of the mean (Zuo 1992; Fardal & Shull 1993; Gnedin & Hamilton 2002; Croft 2004; Meiksin & White 2004; Bolton et al. 2006).

However, the mfp decreases rapidly with redshift, with extrapolations predicting that it should approach values of  $\lambda_{\text{mfp}} \sim 30\text{--}40$  comoving Mpc by  $z \sim 5\text{--}6$  (e.g. Storrie-Lombardi et al. 1994; Miralda-Escudé 2003). Such modest values of the mfp at  $z \gtrsim 5$  can induce significant fluctuations in the UVB, sourced by cosmic variance in the number of ionizing sources inside each  $\lambda_{\text{mfp}}$  horizon. Additionally, the  $1/r^2$  flux attenuation, combined with the clustering of sources, can introduce additional scatter in the UVB on scales smaller than  $\lambda_{\text{mfp}}$  (Zuo 1992; Meiksin & White 2004). These fluctuations have been largely ignored in high-redshift analyses (though see section 5 in Bolton & Haehnelt 2007 and section 4 in Feng et al. 2008).

Here, we undertake a systematic study of the post-reionization metagalactic ionizing background at  $z \sim 5\text{--}6$ . We vary source prescriptions and  $\lambda_{\text{mfp}}$  in an attempt to *infer general conclusions about the theoretical properties and observational importance of spatial fluctuations in the UVB*. We focus on the simplest statistics, the probability density function (PDF) and power spectrum, which are

\*Hubble Fellow.

†E-mail: mesinger@astro.princeton.edu

most likely to be observed. Given the dearth of knowledge concerning the properties of absorbers and sources at high- $z$ , we make use of a range of methods that include analytic, seminumeric and numeric approaches in order to be comprehensive. We compare simple analytic models of these quantities to the numerical results in order to shed light on the origin and importance of the variations; we will show that analytic models are accurate in some, but not all, cases.

This paper is organized as follows. In Section 2, we outline the various modelling tools we use to study the fluctuations in the ionizing background. In Sections 3 and 4, we present the PDFs and power spectra of various UVBs, respectively. In Section 5, we study the impact of a spatially varying UVB on the Ly $\alpha$  forest. Finally, in Section 6 we summarize our findings and offer conclusions.

We quote all quantities in comoving units, with the exception of flux, and we denote proper units with a prefix ‘p’. We adopt the background cosmological parameters ( $\Omega_\Lambda$ ,  $\Omega_M$ ,  $\Omega_b$ ,  $n$ ,  $\sigma_8$ ,  $H_0$ ) = (0.72, 0.28, 0.046, 0.96, 0.82, 70 km s<sup>-1</sup> Mpc<sup>-1</sup>), matching the 5-year results of the *Wilkinson Microwave Anisotropy Probe* satellite (Dunkley et al. 2008; Komatsu et al. 2009).

## 2 GENERATING INHOMOGENEOUS FLUX FIELDS

### 2.1 Seminumerical simulations with flux tool

We generate our ionizing flux fields following the procedure described in Mesinger & Dijkstra (2008). We briefly outline the procedure below.

We begin with halo fields at  $z = 5$  and  $6$  generated with the seminumerical simulation DEXM,<sup>1</sup> which has been shown to reproduce the correct number density and clustering properties of haloes well into the quasi-linear and non-linear regimes (Mesinger & Furlanetto 2007; fig. 1 in Dijkstra et al. 2008; Mesinger et al., in preparation). Our simulation boxes are 150 Mpc on a side, with a 150/1800 Mpc halo grid cell size. The velocity fields used to perturb the halo field were generated on a lower resolution 900<sup>3</sup> grid; thus, our final halo field resolution is  $\Delta x = 150/900 = 0.17$  Mpc.

For each halo field, we create a corresponding UV flux field on a 150<sup>3</sup> grid (spatial resolution of 1 Mpc), by performing a halo mass/ $r^2$  weighted sum. Specifically, we compute the flux of ionizing photons (in units of ionizing photons s<sup>-1</sup> pcm<sup>-2</sup>) with

$$f(\mathbf{x}, z) = \frac{(1+z)^2}{4\pi} \sum_{M_i > M_{\min}} X_i(\text{DC}) \frac{M_i \times \epsilon_{\text{ion}}}{|\mathbf{x} - \mathbf{x}_i|^2} e^{-|\mathbf{x} - \mathbf{x}_i|/\lambda_{\text{mfp}}} \quad (1)$$

where  $\mathbf{x}$  is the location of the cell of interest,  $M_i$  is the total halo mass,  $\mathbf{x}_i$  is its location and the factor of  $(1+z)^2$  converts the factor  $|\mathbf{x} - \mathbf{x}_i|^2$  from comoving into proper units. We also include a duty cycle<sup>2</sup> parameter through the random variable  $X_i(\text{DC})$ , which has a value of 1 with likelihood DC or 0 with likelihood  $1 - \text{DC}$ . Finally, the factor  $\epsilon_{\text{ion}}$  in equation (1) denotes the rate at which ionizing photons are released into the IGM by a dark matter halo

per unit mass. Unless stated otherwise, we assume a fiducial value of  $\epsilon_{\text{ion}}(z) = 3.8 \times 10^{58} / \text{DC} [\frac{\Omega_b}{\Omega_M} \frac{1}{1+z}]$  photons  $M_\odot^{-1} \text{s}^{-1}$ , which provides a good fit to the observed luminosity functions of Ly $\alpha$  emitting galaxies (Kashikawa et al. 2006; Shimasaku et al. 2006; Dijkstra, Wyithe & Haiman 2007; McQuinn et al. 2007; Stark, Loeb & Ellis 2007), and the  $z = 6$  Lyman break galaxies (Bouwens et al. 2006). However, as we are mostly concerned with the shape of the flux PDF, where applicable we present results in units of the mean flux or intensity,  $J/\langle J \rangle$ .

In constructing the UVB, we also assume a minimum halo mass able to host stars,  $M_{\min}$ , and explore mainly two different values for this parameter:  $1.6 \times 10^8$  and  $1.4 \times 10^9 M_\odot$ . The former value corresponds to a virial temperature of  $10^4$  K at these redshifts ( $z \sim 5-6$ ), and represents the regime of ineffective feedback on atomically cooled haloes during reionization (Mesinger & Dijkstra 2008; Okamoto, Gao & Theuns 2008). The later value represents the regime of very inefficient star formation inside such small haloes, perhaps due to strong radiative and/or mechanical feedback (e.g. Yepes et al. 1997; Scannapieco et al. 2006; Pawlik & Schaye 2009). The exact value of the higher  $M_{\min}$  was also motivated by the fact that the number density of haloes with mass greater than  $1.4 \times 10^9 M_\odot$  is roughly 10 times smaller than the number density of haloes with masses greater than  $1.6 \times 10^8 M_\odot$  at  $z \approx 5$ , thus allowing us to compare results at fixed source number density with values of DC = 1.0 and 0.1 in the two models, respectively.

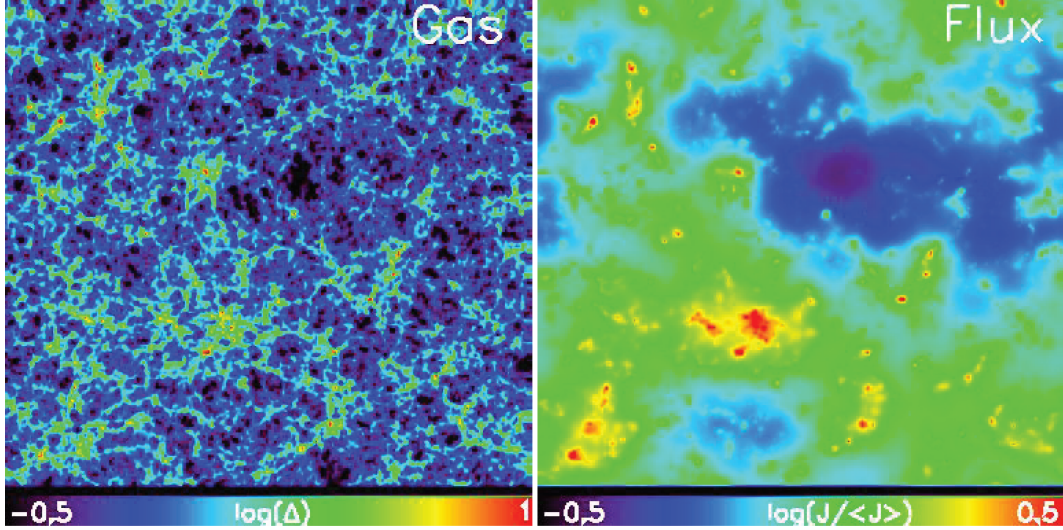
### 2.2 Hydrodynamical simulations with flux tool

Additionally, we generate a flux field corresponding to the  $z = 5.71$  source field from the cosmological hydrodynamical simulation presented in Trac, Cen & Loeb (2008) (their  $z = 6$  reionization model). This simulation is fixed grid, 143 Mpc on a side, and includes prescriptions for modelling dark matter, baryons and ionizing photons (for details see Trac & Cen 2007 and Trac et al. 2008). The density field was calculated on a grid of 0.19 Mpc cells, which resolves the Jeans length in the mean density, ionized IGM by a factor of few and then smoothed to a cell size of 0.74 Mpc. Each halo’s instantaneous star formation rate (SFR) is proportional to the instantaneous gas accretion rate, and enters our equation (1) in place of the halo’s mass,  $M_i$ , with the normalization adjusted accordingly so as to match the mean value of the UVB. We make use of this flux field, combined with the  $z = 5.71$  gas density fields from the hydro simulation, to generate more accurate mean Ly $\alpha$  forest flux decrement statistics in Section 5. At these redshifts and scales, our seminumerically generated density fields somewhat overpredict the rare voids which dominate this statistic (Mesinger et al., in preparation). Furthermore, the ray-tracing algorithm from the numerical simulation overpredicts the fluctuations in the flux field, due to an insufficient number of ray splittings (Trac 2008, private communication). This problem only becomes severe following reionization, but since this is the epoch we study here we use this hybrid prescription in some of the results below (i.e. we do not use the radiative transfer field from the simulation.).

In Fig. 1, we present a 0.74 Mpc thick slice through the density and flux fields based on this  $z = 5.71$  simulation output from Trac et al. (2008). The flux fields were calculated assuming  $\lambda_{\text{mfp}} = 30$  Mpc, and DC = 1. Even for such a moderately high choice of  $\lambda_{\text{mfp}}$ , we can qualitatively see that there is significant inhomogeneity in the flux fields. Furthermore, it is evident that the flux and density fields are highly correlated, an issue we will return to in Section 5.

<sup>1</sup> <http://www.astro.princeton.edu/~mesinger/DexM/>

<sup>2</sup> Note that there are several definitions of the term ‘duty cycle’ in the literature, some of which include this same parameter in the star formation time-scale of each halo (an assumption only valid in extremely idealized scenarios). We adopt the simple definition above of the duty cycle, namely the fraction of dark matter haloes which are ‘on’ (i.e. hosting active galaxies emitting ionizing photons into the IGM; cf. Lee et al. 2009).



**Figure 1.** Slices through the  $z = 5.71$  gas density (left-hand panel) and UVB (right-hand panel) fields using the hydrodynamical simulation of Trac et al. (2008) and our simple flux generation tool. Each slice is 143 Mpc on a side and 0.74 Mpc thick. The UV flux is generated using the SFRs obtained from the simulation and the prescription outlined in Section 2, assuming  $\lambda_{\text{mfp}} = 30$  Mpc,  $\text{DC} = 1$  and  $M_{\text{min}} \sim 3 \times 10^8 M_{\odot}$ .

### 2.3 Analytic model

We complement these numerical techniques with a relatively simple analytic model. While this approach cannot incorporate all of the physics provided by the seminumeric and numeric models, we will find it to be helpful in elucidating the physics of the background radiation field. As in the remainder of this paper, we will focus on computing two statistical descriptions of the flux field: the PDF and the power spectrum.

#### 2.3.1 Flux PDF

The PDF of the radiation background  $J$ , normalized to its mean value  $\langle J \rangle$ , can be computed exactly for randomly distributed sources, provided that we assume a constant attenuation length  $\lambda_{\text{mfp}}$  for ionizing photons much smaller than the Hubble length (Zuo 1992; Meiksin & White 2004). In this case, the crucial parameter is the number of sources within each attenuation length,  $\bar{N}_0 = (4\pi/3)n_i\lambda_{\text{mfp}}^3$ , where  $n_i$  is the source number density and

$$\frac{dp(> X|\bar{N}_0)}{dX} = \frac{1}{\pi} \int_0^{\infty} ds \exp \left[ -s\bar{N}_0 \int dx x\phi(x) \text{Im} G(sx) \right] \times \cos \left[ -sX + s\bar{N}_0 \int dx x\phi(x) \text{Re} G(sx) \right], \quad (2)$$

where  $X = J/J_*$ ,  $\phi(x)$  is the source luminosity function, in units of the mean luminosity  $L_*$  and normalized to have unit integral,  $J_* = L_*/(4\pi\lambda_{\text{mfp}})^2$ ,

$$G(t) = \int_0^{\infty} du \tau^3(u)e^{iut}, \quad (3)$$

and  $u = e^{-\tau}/\tau^2$ . In this limit, the mean background is  $\langle X \rangle = 3\bar{N}_0$ , or  $\langle J \rangle \propto \bar{N}_0 J_* \propto L_* r_0$ . For more details, see equation (20) in Meiksin & White (2003) and associated discussion.

At the end of reionization,  $\lambda_{\text{mfp}} \gtrsim 20$  Mpc and (typically)  $\bar{N}_0 \gtrsim 10^5$ . Thus, the Poisson fluctuations in the galaxy counts are very small and one might expect the ionizing background to be extremely uniform. However, equation (2) assumes that the source counts within each attenuation volume are Poisson distributed

around the fixed value  $\bar{N}_0$ ; in actuality, deterministic halo clustering causes fluctuations in the *expected* number of haloes within each attenuation volume, with a fractional amplitude  $\sim \bar{b}\sigma(\lambda_{\text{mfp}})$  where  $\bar{b}$  is the mean bias of the star-forming galaxies. For our fiducial model at  $z = 6$ , we find  $\bar{b} \sim 2.5$ , which yields fractional variations of the order of  $\sim 2$  for  $\lambda_{\text{mfp}} = 30$  Mpc.

To crudely account for this deterministic clustering, we approximate the underlying density field as linear on scales  $\sim \lambda_{\text{mfp}}$  [so that the PDF of density,  $p(\delta)$ , is a Gaussian with variance  $\sigma^2(\lambda_{\text{mfp}})$ ] and vary the expected number of sources within each attenuation volume according to linear theory, so that  $\bar{N}_0 \rightarrow \bar{N}_0(\delta, \lambda_{\text{mfp}})$ , which depends on the conditional mass function for a region with a fractional overdensity  $\delta$  on a scale  $\lambda_{\text{mfp}}$  (Lacey & Cole 1993). Thus,

$$\frac{dp(> X)}{dX} = \int d\delta p(\delta) \frac{dp(X|\bar{N}_0[1 + \bar{b}\delta])}{dX}. \quad (4)$$

We will see that this is only a crude approximation, for two reasons. First, our prescription for the expected source density is not terribly accurate, since the typical fractional fluctuation in the galaxy counts is not too much smaller than unity, and because we effectively assign the modified source density to all neighbouring regions as well. More importantly, non-linear clustering is actually quite important to the PDF.

#### 2.3.2 Flux power spectrum

Similar techniques also allow one to compute the correlation function or power spectrum of the intensity field due to randomly distributed sources (Zuo & Phinney 1993). However, in this case it is possible to include (linear) clustering exactly by using arguments similar to the ‘halo’ model that describes dark matter and galaxy clustering.

The intensity field around a single galaxy of mass  $m$  is

$$J(r|m_i) = \frac{\epsilon_{\text{ion}} m_i \exp(-r/\lambda_{\text{mfp}})}{(4\pi r)^2}. \quad (5)$$

In our calculations, the normalization factor is set as in the seminumeric model. The total intensity field is simply the sum of the profiles around each galaxy in the Universe. By analogy to the halo

model, the power spectrum can then be written as the sum of two terms, one arising from correlations within a single source's envelope and one arising from correlations between the locations of two sources (Scherrer & Bertschinger 1991; Cooray & Sheth 2002). These two terms both depend on the normalized Fourier transform of the single source profile (equation 5):

$$u_J(k) = \frac{\arctan(k\lambda_{\text{mfp}})}{k\lambda_{\text{mfp}}}, \quad (6)$$

which is independent of galaxy mass (because it only describes the shape, not the amplitude, of the intensity profile). Note that  $u_J \rightarrow 1$  for  $k\lambda_{\text{mfp}} \ll 1$  and  $u_J \rightarrow \pi/(2k\lambda_{\text{mfp}})$  for  $k\lambda_{\text{mfp}} \gg 1$ .

The 'single-source' term, describing correlations within randomly distributed sources' profiles, is

$$P_{1s}(k) = |u_J(k)|^2 \int_{M_{\text{min}}}^{\infty} dm n(m) \left( \frac{m}{\bar{\rho}_{\text{gal}}} \right)^2, \quad (7)$$

where  $n(m)$  is the mass function of the sources (or the halo mass function in our calculations) and  $\bar{\rho}_{\text{gal}} = n_i \bar{m}_{\text{gal}}$  is the mean mass density of galaxies (and  $\bar{m}_{\text{gal}}$  is the mean mass of the ionizing sources). Note that the single-source correlation function can also be derived from the two-point probability density (Zuo & Phinney 1993; see also Furlanetto 2009a). Furthermore, the amplitude is determined by the luminosity function, but the shape is entirely determined by our imposed attenuation profile.

Similarly, the term describing the correlations between two sources may be written as

$$P_{2s}(k) = |u_J(k)|^2 \bar{b}^2 P_{\text{lin}}(k, z) \left[ \int_{M_{\text{min}}}^{\infty} dm n(m) \left( \frac{m}{\bar{\rho}_{\text{gal}}} \right) \right]^2, \quad (8)$$

where  $P_{\text{lin}}(k, z)$  is the linear matter power spectrum. Here, we have assumed that the power spectrum of two haloes may be written as  $P_{\text{hh}}(k, z) \approx b^2(z) P_{\text{lin}}(k, z)$ , or in other words the bias remains linear on the relevant scales. This is generally a good approximation in the halo model at lower redshift, because the approximation only breaks down where the one-halo term already dominates. Again, this two-point function has a shape fixed by the source profiles, but an amplitude that depends on the nature of the sources. In fact, the term in square brackets is simply equal to unity; we have left it in only for comparison to the halo model, where  $u(k)$  is a function of  $m$  and so must be left inside the integral. Because  $|u_J|^2 \rightarrow 1$  for small  $k$ , the two-source term traces the linear matter power spectrum on these scales, before being truncated at  $k\lambda_{\text{mfp}} \sim 1$ .

If we take a simple model in which all sources have a fixed mass, we find that

$$\frac{P_{2s}}{P_{1s}} \approx \bar{b}^2 n_i P_{\text{lin}}(k). \quad (9)$$

Thus, the two-source term becomes more and more important as the number density of sources, their bias and the underlying linear clustering increase: this is simply because all of these increase the importance of deterministic clustering relative to the correlations inherent to the flux profile itself.

### 3 THE FLUX PDF

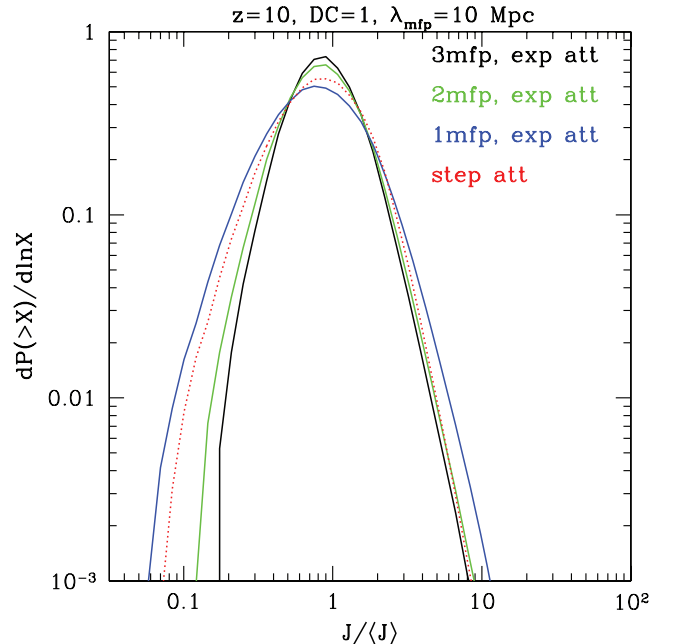
#### 3.1 Attenuation

As discussed above, we account for flux attenuation through the factor  $e^{-r/\lambda_{\text{mfp}}}/(4\pi r^2)$  in equation (1). Thus, flux attenuation is taken into account in a spatially homogeneous way through the free parameter  $\lambda_{\text{mfp}}$ . Although such an approach misses the obvious

correlation of absorbers [Lyman limit systems (LLSs) and damped Lyman  $\alpha$  systems (DLAs)] with sources, it had several benefits over using the typical approximate radiative transfer algorithms in cosmological hydrodynamical simulations. First, our approach is versatile, easy to calibrate and to run quickly for large-scale boxes. This allows us to easily explore the effects of varying  $\lambda_{\text{mfp}}$ , which is important, given the large uncertainties in this value at these redshifts (e.g. Faucher-Giguère et al. 2008). Secondly, our formalism allows for easy comparisons with analytic models, and provides direct insight into the features of the UVB. Finally, and most importantly, large-scale cosmological hydrodynamical simulations *do not directly resolve LLSs and DLAs*; thus, they too must employ analytic prescriptions to populate their box with absorbers. We also note that the mfps at these redshifts are so large ( $\gtrsim 20$  Mpc; Storrie-Lombardi et al. 1994; Madau, Haardt & Rees 1999; Miralda-Escudé 2003; Faucher-Giguère et al. 2008) that flux received by each cell is sourced by tens of thousands of sources, likely averaging out detailed radiative transfer effects. We thus do not expect our results to be significantly affected by the fact that we neglect the details of radiative transfer and the clustering of absorbers.

Nevertheless, stochastic fluctuations in the absorber population will help to increase the variance in the radiation field: if LLSs are primarily responsible for absorption, each attenuation length will contain only a single absorber on average and the absorption length will vary quite strongly. However, in practice  $\gtrsim 50$  per cent of the absorption is typically due to lower column density absorbers (see e.g. the appendix in Furlanetto & Oh 2005), so this is probably not a large factor for us.

In Fig. 2, we plot the flux PDFs for a  $L = 100$  Mpc box with 0.5 Mpc cells at  $z = 10$ ,  $\lambda_{\text{mfp}} = 10$  Mpc and  $M_{\text{min}} = 10^8 M_{\odot}$ . We show the effects of truncating the sum over contributing sources in equation (1) at distances of  $1\lambda_{\text{mfp}}$ ,  $2\lambda_{\text{mfp}}$  and  $3\lambda_{\text{mfp}}$  (left to right solid curves). The distributions appear to converge at  $2\lambda_{\text{mfp}} \rightarrow 3\lambda_{\text{mfp}}$ . The

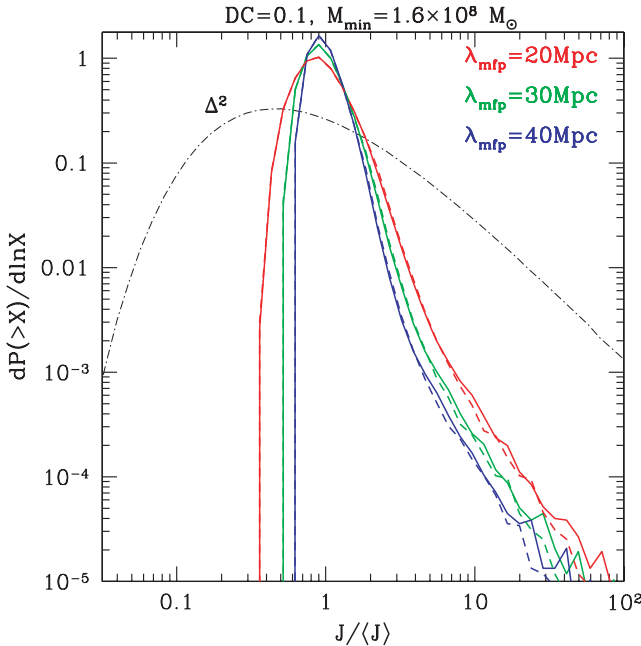


**Figure 2.** Flux PDFs at  $z = 10$ , assuming  $\lambda_{\text{mfp}} = 10$  Mpc,  $M_{\text{min}} = 10^8 M_{\odot}$  and  $DC = 1.0$ , calculated according to equation (1) but truncating the sum over contributing sources at distances of  $1\lambda_{\text{mfp}}$ ,  $2\lambda_{\text{mfp}}$  and  $3\lambda_{\text{mfp}}$  (left to right solid curves). The dotted curve assumes a step-function attenuation, essentially  $\tau = 0$  for distances  $r < \lambda_{\text{mfp}}$  and  $\tau = \infty$  at distances  $r > \lambda_{\text{mfp}}$ .

high-end tails of the distribution are dominated by nearby ionizing sources and are fairly insensitive to the effective horizon (i.e. the high-value tails are roughly constant as more distant sources are included in the UVB calculation). When plotted in unnormalized units (i.e.  $J$  instead of  $J/\langle J \rangle$ ), this overlap is more pronounced and the leftmost (blue) curve shifts to even smaller values. In contrast, the low-end tails of the PDFs correspond to locations far away from ionizing sources and so are very sensitive to the inclusion of more distant sources, which roughly translates to changing the amplitude of an effective homogeneous UVB. If sources were homogeneously distributed, truncating the flux distribution at distances greater than  $n \lambda_{\text{mfp}}$  would underestimate the ionizing flux by a factor of  $(1 - e^{-n})$ . The clustering of sources results in the more complicated behaviour seen in Fig. 2.

The dotted curve in Fig. 2 neglects the exponential attenuation term in equation (1), and instead only includes the contributions of sources at distances  $r < \lambda_{\text{mfp}}$ . This is analogous to assuming a step-function optical depth so that  $\tau = 0$  for distances  $r < \lambda_{\text{mfp}}$  and  $\tau = \infty$  at distances  $r > \lambda_{\text{mfp}}$ . Such a prescription can simplify computation in analytic models (e.g. Zuo 1992) and numerical calculations (e.g. Bolton et al. 2006) and seems to result in fairly accurate flux distributions, though it slightly overestimates the number of faint regions. This overestimate worsens somewhat if one integrates sources only out to  $1\lambda_{\text{mfp}}$  but also includes exponential attenuation (e.g. Santos et al. 2008).

In Fig. 3, we explore the impact of varying  $\lambda_{\text{mfp}}$ . Flux PDFs at  $z = 5$  (solid curves) and  $z = 6$  (dashed curves) are shown assuming  $M_{\text{min}} = 1.6 \times 10^8 M_{\odot}$  and  $\text{DC} = 0.1$ . Note that the evolution of structure from  $z = 10$  in Fig. 2 to  $z = 5$  and  $z = 6$  in Fig. 3 preferentially suppresses the low-end tail of the PDFs as regions containing few ionizing sources become extremely rare. This results in narrower and more asymmetric distributions.



**Figure 3.** Flux PDFs at  $z = 5$  (solid curves) and  $z = 6$  (dashed curves) generated assuming  $M_{\text{min}} = 1.6 \times 10^8 M_{\odot}$  and  $\text{DC} = 0.1$ . Starting from the left-hand side of the figure, curves correspond to values of  $\lambda_{\text{mfp}} = 20, 30, 40$  Mpc, respectively. For comparison, the dot-dashed curve shows the PDF of the squared overdensity,  $\Delta^2 \equiv (\rho/\bar{\rho})^2$ , taken from the hydrodynamical simulation at  $z = 5.71$  with cell size  $0.74$  Mpc.

As expected, increasing the mfp also results in narrower flux distributions, as the gas cells see increasingly representative volumes of space inside their horizon. We note, however, that the  $\lambda_{\text{mfp}} = 40$  Mpc distributions are still fairly broad. Understandably, when plotted in unnormalized units (i.e.  $J$  instead of  $J/\langle J \rangle$ ), the high-value tails overlap as they are dominated by nearby haloes. Perhaps surprisingly, we see very little variation of the flux PDF with redshift, over the range  $z = 5-6$ .

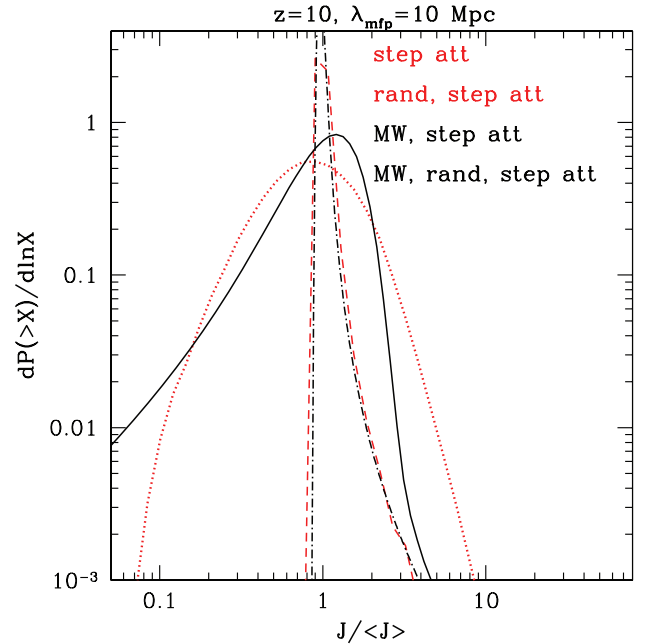
### 3.2 Source clustering

The clustering of sources broadens the flux PDF and creates a high-value tail in the distribution (e.g. Dijkstra et al. 2008; Mesinger & Dijkstra 2008). How strong is this effect and how much do the properties of the underlying sources affect these trends?

In Fig. 4, we plot flux PDFs at  $z = 10$ , assuming  $\lambda_{\text{mfp}} = 10$  Mpc. The dotted curve in the figure was generated from the seminumeric calculation assuming a step-function attenuation at  $\lambda_{\text{mfp}} = 10$  Mpc, shown to be an adequate approximation in Fig. 2. The dashed curve is obtained in the same manner, but with randomized source locations. We see that ignoring source clustering can severely underestimate the widths of the flux PDF. Even the shape of the curves is quite different.

Fig. 4 also compares the PDF to our analytic model. The dot-dashed line uses the analogue of equation (2) but for step-wise attenuation (see Zuo 1992; Furlanetto 2009b), again assuming Poisson distributed sources. It closely matches the seminumeric results for the same assumptions.

The solid line includes linear clustering over the scale  $\lambda_{\text{mfp}}$  as described in equation (4). The clustering length of  $M \sim M_{\text{min}} \sim 10^8 M_{\odot}$  haloes is  $\sim 2$  Mpc at this redshift, so the clustering on scales



**Figure 4.** Flux PDFs at  $z = 10$ , assuming  $\lambda_{\text{mfp}} = 10$  Mpc,  $M_{\text{min}} = 10^8 M_{\odot}$  and  $\text{DC} = 1.0$ . The dotted curve was generated assuming a step-function attenuation, essentially  $\tau = 0$  for distances  $r < \lambda_{\text{mfp}}$  and  $\tau = \infty$  at distances  $r > \lambda_{\text{mfp}}$ . The dashed curve is obtained in the same manner, but with randomized source locations (i.e. ignoring clustering). The solid and dot-dashed curves correspond to the analytical model described in the text, including and ignoring clustering on scales  $> \lambda_{\text{mfp}}$  as described in equation (4), respectively.

larger than the mfp should be well into the linear regime and accurately predicted by the analytic model. Obviously, this simple prescription provides a relatively poor match to the full seminumeric results: although the large-scale clustering does broaden the distribution by a comparable amount, the shapes still disagree.

Part of this difference is easy to understand and is a consequence of the analytic model's simple prescription for clustering: the model not only fixes the *local* attenuation volume to have a given density but it also effectively assumes that surrounding regions have the same number of sources. This artificially suppresses the contribution from distant galaxies (which provide 2/3 of the mean ionizing background in these models). In practice, points inside low-density regions should be exposed to higher flux values from neighbouring higher density regions. Such long-wavelength effects are missed in the simple analytic treatment, resulting in the spuriously long small-value tail in the PDF.

However, by the same logic our prescription places high-density regions near high-density neighbours, so one might expect us to overestimate the extent of the variation at high  $J$  as well – but in fact the prescription *underestimates* the width there as well. This is because, in contrast to the low- $J$  points, the radiation field at such high levels is not dominated by the background of distant sources but by nearby clusters of objects, for which smaller scale clustering and non-linear effects are important.

This is not particularly surprising in light of the analogy we have drawn to the halo model: although that simple prescription does a remarkably good job of matching the dark matter power spectrum, it cannot easily be used to predict the actual density distribution of the IGM. Indeed, Scherrer & Bertschinger (1991) have shown that the PDF of the fractional deviation of  $J$  from the mean value,  $1 + \delta_J = J/\langle J \rangle$ , is

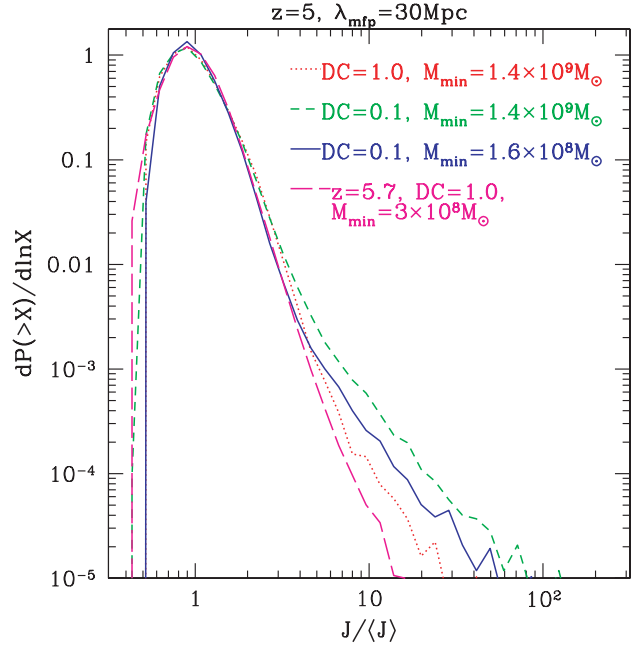
$$\frac{dp(> \delta_J)}{d\delta_J} = \int_{-\infty}^{\infty} \frac{ds}{2\pi} e^{-i\delta_J s} \exp \left[ \sum_{p=1}^{\infty} \frac{(is)^p}{p!} \xi_J^{(p)}(0, 0, \dots, 0) \right], \quad (10)$$

where  $\xi_J^{(p)}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_p)$  is the  $p$ -point correlation function of the radiation field. Thus, the PDF depends on *all* higher order correlations at zero-lag, which are evidently strongly affected by the non-linear clustering on small scales. The simple addition of large-scale linear clustering is not sufficient to match the seminumeric results.

To see the impact of the source properties on the clustering signature in the flux PDF, in Fig. 5, we plot several flux PDFs assuming  $\lambda_{\text{mfp}} = 30$  Mpc. The dotted, short-dashed and solid curves were all generated at  $z = 5$ , but assuming various values of DC and  $M_{\text{min}}$ . These two parameters drive the correlation function of sources and are crucial in constraining the UV luminosity to halo mass relation at high- $z$  (e.g. Lee et al. 2009). From Fig. 5, we see that the low-end cut-off is generally unaffected by these parameters. This is to be expected since, as we mention above, the low-end tail of the flux distribution corresponds to voids distant from UV sources, and are sensitive to the effective horizon  $\lambda_{\text{mfp}}$  (which is identical in all these curves). Thus, most of the clustering imprint is in the high-value tails of the PDFs.

The figure also confirms that the widest flux PDF results from the model in which sources are the most biased and rare: having the highest  $M_{\text{min}}$  and lowest DC. As the duty cycle is increased, the high-value tail retracts. A similar trend is seen as one lowers  $M_{\text{min}}$ .

As mentioned above, the choice of  $M_{\text{min}} = 1.4 \times 10^9 M_{\odot}$  was motivated by the fact that (DC,  $M_{\text{min}}/M_{\odot}$ ) = (1.0,  $1.4 \times 10^9$ ) and (0.1,  $1.6 \times 10^8$ ) result in the same number density of sources, and



**Figure 5.** Flux PDFs assuming  $\lambda_{\text{mfp}} = 30$  Mpc. Dotted, short-dashed and solid curves correspond to  $z = 5$  and (DC,  $M_{\text{min}}/M_{\odot}$ ) = (1.0,  $1.4 \times 10^9$ ), (0.1,  $1.4 \times 10^9$ ) and (0.1,  $1.6 \times 10^8$ ), respectively. The long-dashed curve is generated from the  $z = 5.71$  SFR output from the numerical results of Trac et al. (2008), as described in the text.

thus are useful in isolating the imprint of each these two parameters. Even in our simple model where the duty cycle is not a function of mass, these corresponding curves are quite similar. Thus, it is unlikely that the flux PDF, even if it could be measured, could provide a useful constraint on these parameters.

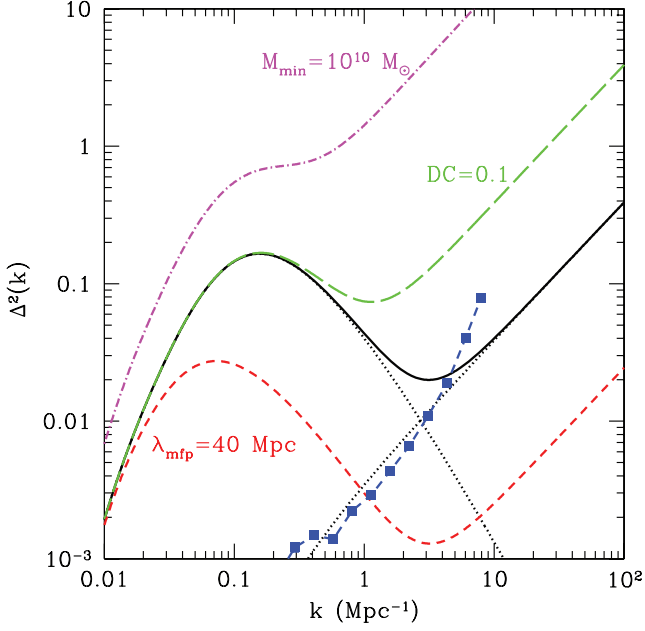
The long-dashed curve in Fig. 5 shows the flux PDF obtained from the  $z = 5.71$  SFR output from the hydrodynamical simulation of Trac et al. (2008), as described in Section 2.2. Despite differences in the star formation prescription, this curve agrees well with the above trends. Note that the slight offset at low  $J$  is attributable to the fact that these fluxes were calculated including sources within  $1.5\lambda_{\text{mfp}}$ , instead of within  $2\lambda_{\text{mfp}}$ , as was the case for the seminumerical boxes.

## 4 THE FLUX POWER SPECTRUM

### 4.1 Analytic results

We begin our discussion of the flux power spectrum by examining the ‘halo model’ analytic results (which we expect to be more successful than the PDF). Fig. 6 shows some examples at  $z = 10$ ; here  $\Delta^2(k, z) = k^3/(2\pi^2) P(k, z)$ . The solid curve takes  $M_{\text{min}} = 10^8 M_{\odot}$ , DC = 1 and  $\lambda_{\text{mfp}} = 10$  Mpc; the dotted curves show the corresponding one- and two-source contributions. Clearly, these two terms are cleanly separated in scale for this set of parameters. The two-source term dominates on large scales; for  $k\lambda_{\text{mfp}} \ll 1$ , it traces the linear power spectrum with an amplitude determined by the mean galaxy bias. It is then damped at  $k\lambda_{\text{mfp}} \sim 1$ : the short-dashed curve, which takes  $\lambda_{\text{mfp}} = 40$  Mpc, shows how the turnover scale shifts with the attenuation length.

On small scales, the one-source term dominates: it has  $\Delta_{\text{1s}}^2 \propto k^3 u_J(k)^2 \propto k/\lambda_{\text{mfp}}^2$ . Thus, its amplitude decreases rapidly with the attenuation length – although this dependence is hard to separate



**Figure 6.** Flux power spectra at  $z = 10$  computed with our analytic model. The solid black line takes  $M_{\min} = 10^8 M_{\odot}$ ,  $DC = 1$  and  $\lambda_{\text{mfp}} = 10$  Mpc; the dotted curves show the corresponding one- and two-source contributions. The curve with solid squares shows the corresponding one-source power spectrum computed from the seminumeric simulations. The long-dashed line takes  $DC = 0.1$  instead and the short-dashed line takes  $\lambda_{\text{mfp}} = 40$  Mpc. The dot-dashed line takes  $M_{\min} = 10^{10} M_{\odot}$ .

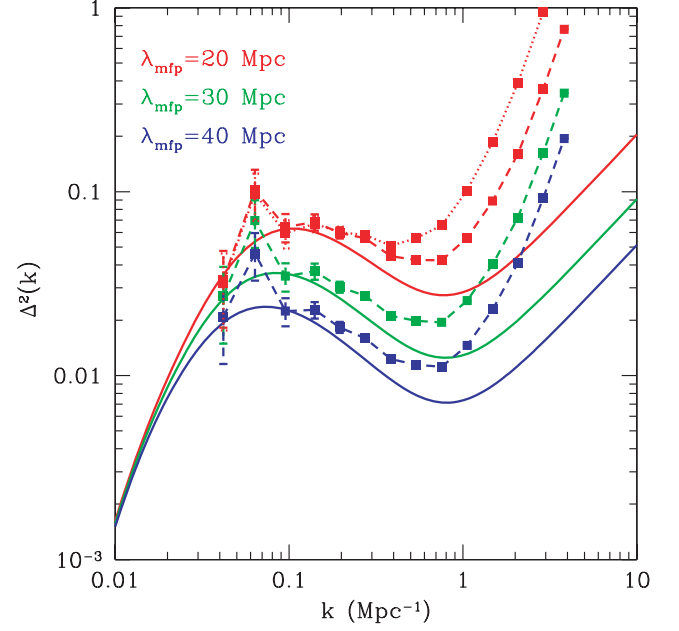
from the dependence on overall source density. The dashed curve with solid squares shows the one-source term computed from the seminumeric simulations (with sources randomly distributed, so that the two-source term disappears). It provides an excellent match to the analytic model, except on scales near the resolution limit,  $k \sim 10 \text{ Mpc}^{-1}$ , where the quantization of values (predominantly separations which are quantized to  $1/3$  of the resolution limit) in our grid-based, seminumerical approach slightly increases the small-scale power (i.e. shot noise).

The dot-dashed curve takes a larger  $M_{\min}$ . Note that the power spectrum increases with the halo bias on very large scales. Also, the decreased source number density strongly increases the importance of the one-source term, giving a much less clean separation between the one-source and two-source regime.

## 4.2 Numerical results

We now compare these analytic predictions to the seminumerical simulations. For the numerical calculations, we use the convention  $\Delta^2(k, z) = k^3/(2\pi^2 V) \langle |\delta_J(\mathbf{k}, z)|^2 \rangle_k$ . In Fig. 7, we plot the flux power spectra at  $z = 6$  for both the analytic (solid curves) and seminumeric (dashed curves) models, generated assuming  $M_{\min} = 1.6 \times 10^8 M_{\odot}$  and  $DC = 0.1$ . Curves correspond to values of  $\lambda_{\text{mfp}} = 20, 30, 40$  Mpc, from top to bottom.

The agreement between the two calculations is quite impressive, especially at large scales. The analytic model clearly shows the separation between the two-source and single-source terms; with the seminumeric results the difference is less clear because the boxes are not large enough to sample the regime  $k\lambda_{\text{mfp}} \gg 1$ , where  $\Delta^2$  traces the matter power spectrum. However, the amplitudes of the analytic and seminumeric calculations match reasonably well on these scales: the differences in the two-source regime are attributable to



**Figure 7.** Flux power spectra at  $z = 6$  from the analytic model (solid curves) and the seminumeric simulations (dashed curves, with Poissonian error bars). All curves assume  $M_{\min} = 1.6 \times 10^8 M_{\odot}$  and  $DC = 0.1$ ; the three sets correspond to values of  $\lambda_{\text{mfp}} = 20, 30, 40$  Mpc, from top to bottom. The dotted curve shows the seminumeric result at  $z = 5$  for  $\lambda_{\text{mfp}} = 20$  Mpc.

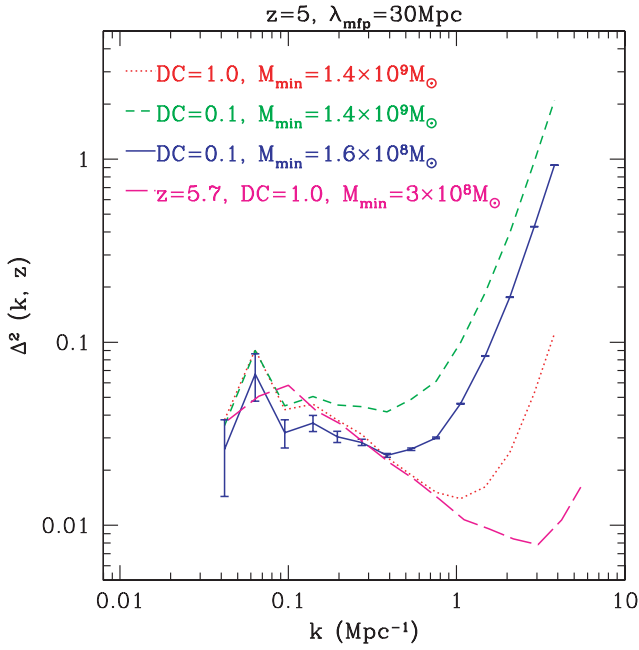
slightly different mean halo biases in the analytic and seminumeric calculations.

However, the seminumeric results apparently have a steeper slope for the one-source component. This is most likely due to non-linear clustering, which is not included in either of our two-source or one-source terms (see the discussion following equation 8). In fact, although the density field itself remains linear to quite small scales, high-redshift haloes are highly biased, so non-linear effects modify the halo clustering on  $\sim \text{Mpc}$  scales (e.g. fig. 1 in Dijkstra et al. 2008). One way to think of this is as a steepening of the one-source term, if one treats nearby sources as a single cluster; another is as a boost in the two-source term at moderately small scales. Interestingly, however, the crossover scale between the one-source and two-source regimes is accurately predicted by the analytic model, because it occurs at  $r_{\text{cross}} \sim 2\pi/k_{\text{cross}} \sim 6$  Mpc, where non-linear clustering is not so important.

The three sets of curves in Fig. 7 show that the mfp has little impact on the shape of the spectrum over the scales probed by the seminumeric simulations: it primarily serves to decrease power fairly evenly on all scales with  $k\lambda_{\text{mfp}} \lesssim 1$ . As described above, for smaller wavenumbers the attenuation length has no effect: effectively, a larger attenuation length causes more damping relative to the linear power spectrum.

Finally, the dashed curve shows the flux power spectrum from the seminumeric simulations at  $z = 5$  and  $\lambda_{\text{mfp}} = 20$  Mpc. By comparing Figs 7 and 3, we see that the flux power spectra are more sensitive to redshift than the flux PDFs. The differences are similar for other values of  $\lambda_{\text{mfp}}$ , and in the analytic model: in all cases, the amplitude of the one-source component is somewhat larger at the lower redshift.

This is easily understood from the analytic formalism. The two-source term is proportional to  $\bar{b}^2 P_{\text{lin}}(k, z)$ ; each of these terms evolves with redshift, but in the product the evolution nearly



**Figure 8.** Flux power spectra analogous to the models shown in Fig. 5.

cancels at high redshift (Oh 2001). The one-source term is proportional to the second moment of the mass function (when integrated over mass – equation 7), which increases towards lower redshifts as more and more massive haloes are assembled.

A similar argument holds for the power spectra presented in Fig. 8, where we vary the DC and  $M_{\min}$  parameters. The curves correspond to the same models presented in Fig. 5. Here, we see drastic differences on moderate to small scales, because models in which the sources are rare (i.e. with low DC and high  $M_{\min}$ ) transition to the single-source regime on larger scales. If observed, this statistic could be a powerful tool in constraining these parameters. The flux field has the benefit of encoding information about ionizing sources which are too faint to be detected by galaxy surveys. Unfortunately, such flux power spectra will be nearly impossible to extract from the Ly $\alpha$  forest, even given a very large number of sightlines (see Fig. 11 and associated discussion).

## 5 LYMAN $\alpha$ FOREST

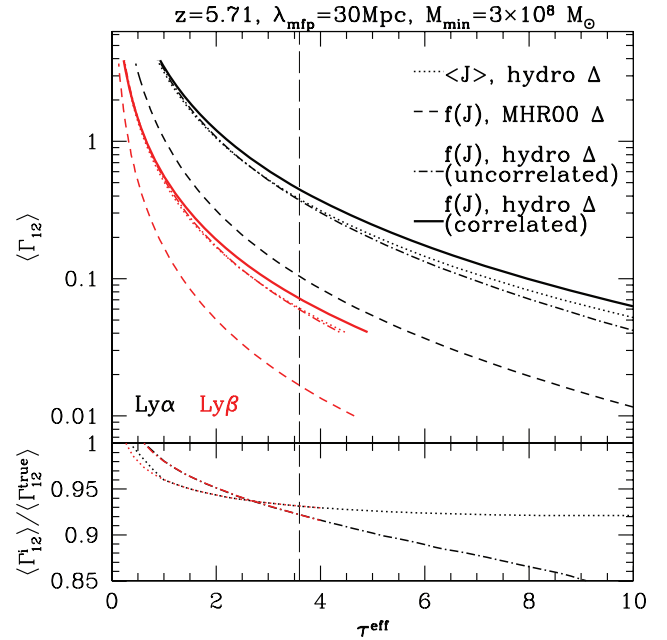
The most obvious application of our flux field analysis is in the interpretation of the Ly $\alpha$  forest spectra following reionization. Thus far, almost all estimates of the photoionization rate per hydrogen atom,  $\Gamma_{12} \equiv \Gamma/10^{-12} \text{ s}^{-1}$ , obtained from the Ly $\alpha$  forest have assumed a uniform UV background (e.g. Fan et al. 2006). However, in Section 3 we showed that UV background actually has sizable fluctuations in this regime. These fluctuations could potentially introduce bias and stochasticity in the determinations of  $\Gamma_{12}$  from Ly $\alpha$  forest attenuation measurements.

Bolton & Haehnelt (2007) briefly addressed this issue by including a similar UV background model to our seminumeric approach. They confirm the results of lower redshift studies (Gnedin & Hamilton 2002; Meiksin & White 2004; Bolton et al. 2006) that assuming a homogeneous UVB underestimates the mean ionization rate obtained from Ly $\alpha$  forest absorption,  $\langle \Gamma_{12} \rangle$ . They find that this effect is of the order of  $\sim 1$ –10 per cent, depending on  $\lambda_{\text{mfp}}$  and  $M_{\min}$  (see their fig. 3). However, for this analysis they used a single realization

of a  $30 h^{-1}$  Mpc simulation box. This box size is comparable to a single ionizing photon mfp at these redshifts (Storrie-Lombardi et al. 1994; Madau et al. 1999; Miralda-Escudé 2003; Faucher-Giguère et al. 2008) and is several times smaller than the length of the Gunn–Peterson (GP) troughs used to estimate  $\langle \Gamma_{12} \rangle$ . Thus, their analysis was unable to fully take into account cosmic variance.

Using our large simulation boxes, we wish to systematically investigate this effect. As a first step, we do not create detailed mock absorption spectra. Instead, for a given flux field, we merely calculate the Ly $\alpha$  optical depth,  $\tau_{\alpha}$ , for each cell, assuming ionization equilibrium. Then, we average the flux decrements over the entire box to obtain a corresponding effective GP optical depth:  $\tau_{\alpha}^{\text{eff}} \equiv -\ln(\exp[-\tau_{\alpha}])_x$ .

In the top panel of Fig. 9, we plot  $\langle \Gamma_{12} \rangle$  for each corresponding effective Ly $\alpha$  (black curves) and Ly $\beta$  (red curves) GP optical depth, for the  $z = 5.71$  hydrodynamical simulation box, assuming  $\lambda_{\text{mfp}} = 30$  Mpc. The cell size for this simulation is 0.74 Mpc. As mentioned above, the density field was calculated on a much finer grid of 0.19 Mpc, which resolves the Jeans length in the mean density, ionized IGM by a factor of few. The smoothed cell size of 0.74 Mpc corresponds to the Jeans mass in the mean density ionized IGM, and is roughly comparable to the Doppler width of the Ly $\alpha$  line at  $T = 10^4$  K, as well as to the Keck Echelle Spectrometer and Imager (ESI) resolution, when binned to increase the dynamic range, at this redshift. Thus, we expect our  $\tau^{\text{eff}} \leftrightarrow \langle \Gamma_{12} \rangle$  mapping to be adequate



**Figure 9.** Top panel: mean ionization rate corresponding to an effective Ly $\alpha$  (black curves) and Ly $\beta$  (red curves) GP optical depth, for the  $z = 5.71$  hydrodynamical simulation box of Trac et al. (2008), assuming  $\lambda_{\text{mfp}} = 30$  Mpc. Dotted curves use the density field from the simulation, but assume a uniform UVB. Dashed curves assume a fluctuating UVB from our simulations, but convolved with the Miralda-Escudé et al. (2000) density distribution. Dot-dashed curves use the fluctuating UVB and density field from the simulations, but randomly correlate the two. Finally, solid curves use the correlated fluctuating UVB field and underlying density field from the simulations. Bottom panel: ratio of the mean ionization rate for each model  $i$  and the ‘true’ model shown as the solid curves in the top panel. The vertical dashed line demarcates the average value of  $\tau_{\alpha}^{\text{eff}} \sim 3.6$ , roughly corresponding to the value measured from the SDSS QSOs at this redshift (taken from Fan et al. 2006).

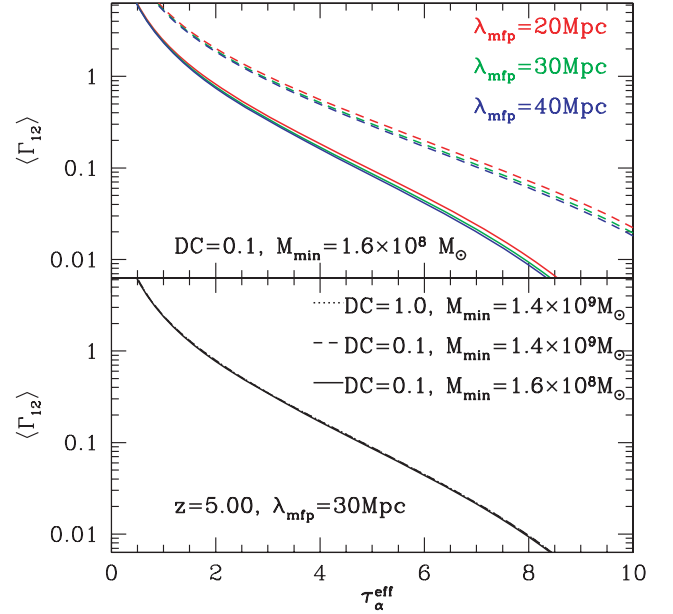
[indeed, our results are consistent with the mock spectra analysis of Bolton & Haehnelt (2007)], but we stress that we did not create detailed mock spectra.

The curves in Fig. 9 show the effects of various simplifications on the value of  $\langle \Gamma_{12} \rangle$  inferred from the  $\tau^{\text{eff}}$  measured in the spectra. The most accurate model, using both our fluctuating flux fields and underlying density fields, is shown by the solid curves. By comparing the solid and dotted curves (or their ratios in the bottom panel), we see that assuming a uniform UVB makes very little impact on the estimated mean ionization rate per hydrogen atom. At low values of  $\tau^{\text{eff}}$ , the two curves are practically indistinguishable, with the difference increasing with  $\tau^{\text{eff}}$ . At  $\tau^{\text{eff}} \sim 3.6$  (roughly corresponding to the observed value at this redshift; e.g. Fan et al. 2006), assuming a uniform UVB results in a  $\sim 5$  per cent underestimate in the value of  $\langle \Gamma_{12} \rangle$ .

It is fairly easy to interpret this behaviour by looking back at Fig. 3. The optical depth approximately scales as  $\tau \propto \Delta^2/\Gamma$ , where  $\Delta \equiv \rho/\bar{\rho}$  is the gas density in units of the mean. Thus, the optical depth distribution can be approximately obtained by convolving these two quantities. From Fig. 3, we see that the distribution of  $\Delta^2$  is much wider than all of the flux distributions. Thus, a uniform UVB, which treats the flux distributions as a delta function, is an accurate assumption. Because of the exponential attenuation involved, the  $\tau^{\text{eff}}$  statistic is sensitive to the low-end tail of the density distribution. As the ionizing background gets fainter (i.e. increasing  $\tau^{\text{eff}}$ ), the  $\tau^{\text{eff}}$  statistic is sensitive to an increasingly narrow range in the low-end tail of the optical depth distribution. It is in this regime that the non-zero width of the ionizing background PDF becomes evident, as the convolution with  $\Delta^2$  is mostly evident in smearing out this low-end tail. Thus, assuming a uniform UVB underestimates the true  $\langle \Gamma_{12} \rangle$ , but only slightly, with this underestimate increasing with increasing  $\tau^{\text{eff}}$ .

Fig. 9 also shows the impact of other common assumptions in the  $\tau^{\text{eff}} \leftrightarrow \langle \Gamma_{12} \rangle$  mapping. The dot-dashed curve is generated by using the same UVB and density fields as in the solid curve, but randomly pairing the two. This also results in a slight underestimate of  $\langle \Gamma_{12} \rangle$ , though a little more notable than the uniform UVB assumption. Since underdense pixels are preferentially farther away from most sources, where the UVB is lower (cf. Fig. 1), not correlating the two fields results in underdense pixels receiving fluxes which are too strong, thus lowering  $\langle \Gamma_{12} \rangle$ . Nevertheless, the effect is still small because of the relative width of the PDFs argument in the above paragraph.

Finally, the most drastic effect, resulting in an underestimate of a factor of  $\sim 2$  in our models, comes from using the analytic density PDF from Miralda-Escudé, Haehnelt & Rees (2000). This offset is normally attributed to errors in the fluctuating GP approximation, which ignores effects such as the line wings, peculiar velocities and blending. However, our numeric results make a similar approximation, so that cannot be the source of the error here. Instead, it may be differences between the assumed density distribution (generated to fit numerical simulations at redshifts  $z \sim 2-4$ , using a simulation box of only  $10 h^{-1}$  Mpc) and that of the simulation. At higher redshifts (i.e.  $\tau^{\text{eff}}$  values), the transmission becomes dominated by rare voids (Oh & Furlanetto 2005; Lidz, Oh & Furlanetto 2006), and so cosmic variance is very important. Proper analysis requires large simulation boxes with accurate hydrodynamical treatment and direct higher redshift outputs. However, if the semi-analytic method is renormalized using a simulation (as is typical), the *shapes* of the curves are similar enough that the errors will be relatively small, so the approximation is still useful for understanding the evolution of  $\langle \Gamma_{12} \rangle$ . This offset is most likely dominated by differences in the



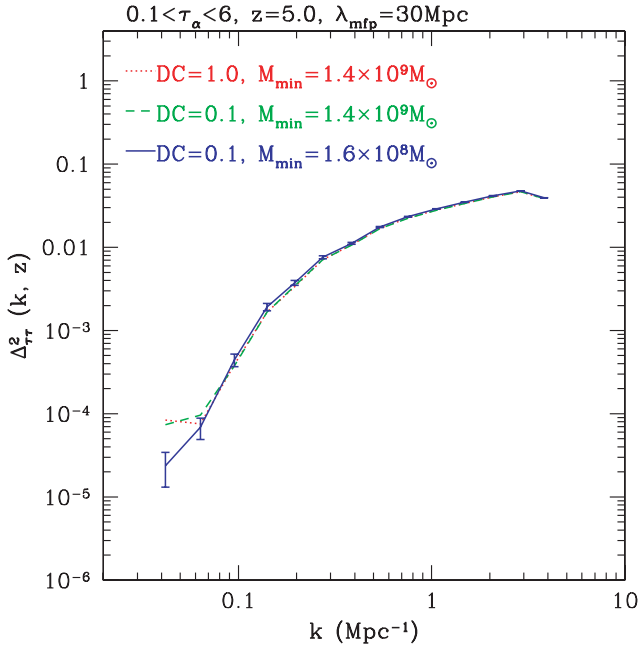
**Figure 10.**  $\tau^{\text{eff}} \leftrightarrow \langle \Gamma_{12} \rangle$  mappings using our seminumerical simulations, assuming various flux models. In the top panel, solid curves correspond to  $z = 5$ , while dashed curves correspond to  $z = 6$ .

assumed thermal history of the gas, which sets the effective local smoothing scale, as well as the effective resolution of the simulation, which we smooth to match our seminumerical simulation and crudely mimic instrument resolution. A detailed exploration of this is beyond the scope of this work, and merits its own study.

In Fig. 10, we show the  $\tau^{\text{eff}} \leftrightarrow \langle \Gamma_{12} \rangle$  mappings using our seminumerical simulations, assuming various flux models. Note that the seminumerical density fields slightly overpredict the presence of very rare voids which dominate this statistic (Mesinger et al., in preparation), thereby over-predicting the transmission for a given UVB. It is remarkable how insensitive this mapping is to the details of the UVB, despite the notable differences in the flux PDFs and power spectra shown in the previous section. Again, this is due to the fact that the fluctuations in the UVB, and the differences among various models of the UVB, are overshadowed by the fluctuations in the density field. The only notable differences among the curves is due to the density field evolution from  $z = 6$  to 5.

We note that our results are in general agreement with Bolton & Haehnelt (2007), who found that assuming a homogeneous UVB underestimates  $\langle \Gamma_{12} \rangle$ , by a few per cent over this range of  $\lambda_{\text{mfp}}$  values. However, they find a somewhat stronger dependence on  $\lambda_{\text{mfp}}$  than we do. This is most likely a result of cosmic variance attributable to the small simulation box they used to study this effect.

Finally, we wish to quantify how easy is it to distinguish various UVB models from observations of the Ly $\alpha$  forest. One can see from Fig. 8 that the shape of the flux power spectrum encodes information about the clustering properties of the sources. As mentioned above, the flux field (as potentially extracted from the Ly $\alpha$  forest) has a great benefit in being able to probe the impact of ionizing sources below the detection thresholds of galaxy surveys. Hence, in Fig. 11 we show the optical depth power spectra at  $z = 5$  and  $\lambda_{\text{mfp}} = 30$  Mpc for some of the same models as in Fig. 8. Unfortunately, the optical depth power spectra of these models are practically indistinguishable from each other, despite the large differences shown in Fig. 8. This result could have been foreshadowed from Fig. 10



**Figure 11.** Optical depth power spectra at  $z = 5$  and  $\lambda_{\text{mfp}} = 30$  Mpc, within the range  $0.1 < \tau_{\alpha} < 6$ , approximately corresponding to the range probed by current instruments.

and is again due to the fact that the density field and its fluctuations swamp the UVB signal.

However, we note that the density fluctuations decrease as redshift increases, while fluctuations in the UVB increase. The UVB becomes more inhomogeneous not only through the evolution of  $\lambda_{\text{mfp}}$  and the bias of sources, but it is also modulated by the ionized bubbles throughout reionization (whenever that epoch may be). Thus, inhomogeneities in the flux field become increasingly important at higher redshifts, although whether this can be observed in the highly saturated forest at  $z \gtrsim 6$ , where the density fluctuations themselves cause substantial fluctuations in  $\tau^{\text{eff}}$  (Lidz et al. 2006), is questionable.

## 6 CONCLUSIONS

We systematically study the expected fluctuations in the hydrogen-ionizing background in the epoch following reionization ( $z \sim 5-6$ ). Unlike at lower redshifts ( $z \lesssim 4$ ), the UVB can be quite inhomogeneous in this regime. This is due to the smaller mfp of ionizing photons, as well as the clustering of increasingly rare ionizing sources.

We confirm that there is a sizable spread (the widths of the PDFs at half of the maximum likelihoods spanning factors of  $\sim 2-4$ ) and asymmetry in the distribution of ionizing fluxes in a cosmological volume. Expected values of the mfp are large enough ( $\lambda_{\text{mfp}} \gtrsim 20$  Mpc), and sources are ubiquitous enough, that the PDF of the ionizing background is asymmetric, with a tail extending to higher values. This high-value tail is set by clustering on small scales, and is insensitive to the contributions of distant sources (i.e.  $\lambda_{\text{mfp}}$ ) (Mesinger & Dijkstra 2008). We also find that the PDFs are most sensitive to the mfp, with higher values of  $\lambda_{\text{mfp}}$  increasingly truncating the low-value tails. Furthermore, we note that the analytic formalism of Meiksin & White (2004) severely underestimates the width of the flux PDF at this epoch, since it ignores the clustering of sources.

The power spectra of the ionizing background are even more sensitive to the details of the source model. We have shown that an approach analogous to the halo model, in which contributions from the proximity zones of single sources (which dominate fluctuations on small scales and are independent of the large-scale galaxy distribution) and multiple sources (which dominate on large scales and depend on the large-scale clustering of the ionizing sources) are separated, provides a good match to the simulated results, although non-linear clustering provides even more small-scale power than the analytic model predicts. The single-source component is particularly sensitive to the number density of galaxies; unfortunately, these power spectra will be nearly impossible to observe with the Ly $\alpha$  forest.

We also model the imprint of various UVBs in the Ly $\alpha$  forest at  $z \sim 5-6$ . We find that the Ly $\alpha$  forest spectra are extremely insensitive to the details of the UVB, despite marked differences in the PDFs and power spectra of our various UVBs. This is attributable to the fact that the Ly $\alpha$  optical depth scales as  $\tau \propto \Delta^2/\Gamma$ , and the squared overdensity distribution,  $f(\Delta^2)$ , is much wider than all of our  $\Gamma$  distributions. In fact, even the extreme scenario of a uniform background (i.e. a delta function  $\Gamma$  distribution) only underestimates the value of  $\langle \Gamma_{12} \rangle$  inferred from the Ly $\alpha$  forest by a few per cent. Thus, our results justify the common assumption of a uniform UVB in Ly $\alpha$  forest analysis.

We also confirm that accurate modelling of the density field is very important in the  $\tau^{\text{eff}} \leftrightarrow \langle \Gamma_{12} \rangle$  relation. We find significant differences (factor of  $\sim 2$ ) in this relation among the numerical and extrapolated Miralda-Escudé et al. (2000) density distributions (though we stress again that we did not create detailed mock spectra in this work). At higher redshifts, the  $\tau^{\text{eff}}$  statistic increasingly depends on a narrower range in the low-value density PDF. Large-scale hydrodynamical modelling of the statistics of these rare voids is therefore essential in interpreting the high-redshift Ly $\alpha$  forest. Pawlik, Schaye & van Scherpenzeel (2009) have recently studied the IGM density field during reionization, and do find significant differences from the Miralda-Escudé et al. (2000) fitting formula, although their simulation boxes are much too small to study the rare voids pertinent to these efforts. Differences in the density fields are most likely dominated by differences in the assumed thermal history of the gas, which sets the effective local smoothing scale, as well as the effective resolution of the simulation, which we smooth to match our seminumerical simulation and crudely mimic instrument resolution.

An interesting comparison can be made between our analysis of hydrogen reionization and similar effects during and after helium reionization (Dixon & Furlanetto 2009). In the latter case, the PDF of  $J$  is much broader (spanning over an order of magnitude), and the fluctuations have a much larger effect on the He II Ly $\alpha$  forest, introducing a factor of  $\sim 2$  bias in estimates of  $\langle \Gamma \rangle$  when the background is assumed uniform. This has important implications for the He II Ly $\alpha$  forest during reionization: unlike the case we have studied here, the large fluctuations in the ionizing background lead to substantial opacity gaps even during the middle phases of helium reionization.

## ACKNOWLEDGMENTS

We thank Hy Trac and Renyue Cen for the use of the  $z = 5.71$  density and star formation rate output of their simulations, as well as for stimulating conversations. We also thank James Bolton for helpful comments on the manuscript. Support for this work was provided by NASA through Hubble Fellowship grant #HF-01222.01 to AM,

awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. SF was partially supported by the NSF through grant AST-0607470 and by the David and Lucile Packard Foundation.

## REFERENCES

- Bolton J. S., Haehnelt M. G., 2007, *MNRAS*, 382, 325
- Bolton J. S., Haehnelt M. G., Viel M., Carswell R. F., 2006, *MNRAS*, 366, 1378
- Bouwens R. J., Illingworth G. D., Blakeslee J. P., Franx M., 2006, *ApJ*, 653, 53
- Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1
- Croft R. A. C., 2004, *ApJ*, 610, 642
- Dijkstra M., Wyithe J. S. B., Haiman Z., 2007, *MNRAS*, 379, 253
- Dijkstra M., Haiman Z., Mesinger A., Wyithe J. S. B., 2008, *MNRAS*, 391, 1961
- Dixon K. L., Furlanetto S. R., 2009, *ApJ*, submitted (arXiv:0906.4116)
- Dunkley J. et al., 2008, preprint (arXiv e-prints, 803)
- Fan X. et al., 2006, *AJ*, 132, 117
- Fardal M. A., Shull J. M., 1993, *ApJ*, 415, 524
- Faucher-Giguère C.-A., Lidz A., Hernquist L., Zaldarriaga M., 2008, *ApJ*, 688, 85
- Feng L., Bi H., Liu J., Fang L.-Z., 2008, *MNRAS*, 383, 1459
- Furlanetto S., 2009a, *ApJ*, 700, 1666
- Furlanetto S., 2009b, *ApJ*, 703, 702
- Furlanetto S. R., Oh S. P., 2005, *MNRAS*, 363, 1031
- Gnedin N. Y., Hamilton A. J. S., 2002, *MNRAS*, 334, 107
- Kashikawa N. et al., 2006, *ApJ*, 648, 7
- Komatsu E. et al., 2009, *ApJS*, 180, 330
- Lacey C., Cole S., 1993, *MNRAS*, 262, 627
- Lee K., Gialalisco M., Conroy C., Wechsler R. H., Ferguson H. C., Somerville R. S., Dickinson M. E., Urry C. M., 2009, *ApJ*, 695, 368
- Lidz A., Oh S. P., Furlanetto S. R., 2006, *ApJ*, 639, L47
- McQuinn M., Hernquist L., Zaldarriaga M., Dutta S., 2007, *MNRAS*, 381, 75
- Madau P., Haardt F., Rees M. J., 1999, *ApJ*, 514, 648
- Meiksin A. A., 2007, preprint (arXiv e-prints)
- Meiksin A., White M., 2003, *MNRAS*, 342, 1205
- Meiksin A., White M., 2004, *MNRAS*, 350, 1107
- Mesinger A., Dijkstra M., 2008, *MNRAS*, 390, 1071
- Mesinger A., Furlanetto S., 2007, *ApJ*, 669, 663
- Miralda-Escudé J., 2003, *ApJ*, 597, 66
- Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, *ApJ*, 530, 1
- Oh S. P., 2001, *ApJ*, 553, 499
- Oh S. P., Furlanetto S. R., 2005, *ApJ*, 620, L9
- Okamoto T., Gao L., Theuns T., 2008, *MNRAS*, 390, 920
- Pawlik A. H., Schaye J., 2009, *MNRAS*, 396, L46
- Pawlik A. H., Schaye J., van Scherpenzeel E., 2009, *MNRAS*, 394, 1812
- Santos M. G., Amblard A., Pritchard J., Trac H., Cen R., Cooray A., 2008, *ApJ*, 689, 1
- Scannapieco C., Tissera P. B., White S. D. M., Springel V., 2006, *MNRAS*, 371, 1125
- Scherrer R. J., Bertschinger E., 1991, *ApJ*, 381, 349
- Shimasaku K. et al., 2006, *PASJ*, 58, 313
- Stark D. P., Loeb A., Ellis R. S., 2007, *ApJ*, 668, 627
- Storrie-Lombardi L. J., McMahon R. G., Irwin M. J., Hazard C., 1994, *ApJ*, 427, L13
- Trac H., Cen R., 2007, *ApJ*, 671, 1
- Trac H., Cen R., Loeb A., 2008, *ApJ*, 689, L81
- Yepes G., Kates R., Khokhlov A., Klypin A., 1997, *MNRAS*, 284, 235
- Zuo L., 1992, *MNRAS*, 258, 36
- Zuo L., Phinney E. S., 1993, *ApJ*, 418, 28

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.