Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **479,** 43–63 (2018)
Advance Access publication 2018 May 18

doi:10.1093/mnras/sty1318

The role of galaxies and AGN in reionizing the IGM – I. Keck spectroscopy of $5 < z < 7$ galaxies in the QSO field J1148+5251

Koki Kakiichi,[1★] Richard S. Ellis,[1★] Nicolas Laporte,[1★] Adi Zitrin,[2] Anna-Christina Eilers,[3] Emma Ryan-Weber,[4,5] Romain A. Meyer,[1] Brant Robertson,[6] Daniel P. Stark[7] and Sarah E. I. Bosman[1]

[1] *Department of Physics and Astronomy, University College London, London WC1E 6BT, UK*
[2] *Physics Department, Ben-Gurion University of the Negev, PO Box 653, Beer-Sheva 84105, Israel*
[3] *Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany*
[4] *Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, VIC 3122, Australia*
[5] *ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia*
[6] *Department of Astronomy and Astrophysics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA*
[7] *Steward Observatory, University of Arizona, Tucson, AZ 85719, USA*

Accepted 2018 May 16. Received 2018 May 15; in original form 2018 March 8

ABSTRACT

We introduce a new method for determining the influence of galaxies and active galactic nuclei (AGN) on the intergalactic medium (IGM) at high redshift and illustrate its potential via a first application to the field of the $z = 6.42$ QSO J1148+5251. Correlating spatial positions Lyman break galaxies (LBGs) with the Lyman alpha forest seen in the spectrum of a background QSO, we provide a statistical measure of the typical escape fraction of Lyman continuum photons. Using Keck DEIMOS spectroscopy to locate seven colour-selected LBGs in the range $5.3 \lesssim z \lesssim 6.4$ we examine the spatial correlation between this sample and Lyα/Lyβ transmission fluctuations in a Keck ESI spectrum of the QSO. Interpreting the statistical H I proximity effect as arising from faint galaxies clustered around the LBGs, we translate the observed mean Lyα transmitted flux into a constraint on the mean escape fraction $\langle f_{\rm esc}\rangle \geqslant 0.08$ at $z \simeq 6$. We also report individual transverse H I proximity effect for a $z = 6.177$ luminous LBG via a Lyβ transmission spike and two broad Lyα transmission spikes around the $z = 5.701$ AGN. We discuss the origin of such associations which suggest that while faint galaxies are primarily driving reionization, luminous galaxies and AGN may provide important contributions to the UV background or thermal fluctuations of the IGM at $z \simeq 6$. Although a limited sample, our results demonstrate the potential of making progress using this method in resolving one of the most challenging aspects of the contribution of galaxies and AGN to cosmic reionization.

Key words: galaxies: formation – galaxies: high-redshift – intergalactic medium – quasars: absorption lines – cosmology: observations – dark ages, reionization, first stars.

1 INTRODUCTION

Understanding how and when cosmic reionization occurred represents one of the most important challenges in observational cosmology and galaxy formation. Of particular interest is the nature of sources responsible, which was first discussed over 50 yr ago (Gunn & Peterson 1965). Although reionization is commonly assumed to be driven by the abundant population of intrinsically faint star-forming galaxies (e.g. Robertson et al. 2013, 2015, for a review see Stark 2016), a key assumption is that the average escape fraction of Lyman continuum (LyC) photons is $\sim 10-20$ per cent. Such high escape fractions are rarely encountered in lower redshift star-forming galaxies where direct measurements of the LyC leakage are possible (Mostardi et al. 2015; Naidu et al. 2017). On the other hand, recent observations of Lyα emission in the spectra of $z > 7$ galaxies (Oesch et al. 2015; Zitrin et al. 2015) might indicate that reionization is accelerated in the volumes around the most luminous galaxies (Stark et al. 2017), possibly as a result of their harbouring active galactic nuclei (AGN; Laporte et al. 2017). A significant contribution of ionizing photons from rare sources such as luminous galaxies and/or AGN (Giallongo et al. 2015, but see Parsa, Dunlop & McLure 2018) may also explain the significant scatter

★ E-mail: k.kakiichi@ucl.ac.uk (KK), richard.ellis@ucl.ac.uk (RSE), n.laporte@ucl.ac.uk (NL)

in the effective optical depth of Lyα absorption in the spectra of $z \gtrsim 5.5$ QSOs (Becker et al. 2015b; Chardin et al. 2015; Chardin, Puchwein & Haehnelt 2017; Bosman et al. 2018). However, both observationally and theoretically the relative ionizing contribution of galaxies and AGN is a subject of intense debate (Madau & Haardt 2015; D'Aloisio et al. 2017; Qin et al. 2017; Hassan et al. 2018; Mitra, Choudhury & Ferrara 2018).

A fundamental impasse to progress is the absence of a reliable technique to measure the escape fraction $f_{\rm esc}$ of ionizing photons at high redshift where direct measures of the leaking LyC radiation become impractical due to foreground line-of-sight absorption. Indirect methods have been examined including absorption line measures of the covering fraction of low-ionization gas in the spectra of lensed galaxies (Jones et al. 2013; Leethochawalit et al. 2016) which suggest a modest increase in $f_{\rm esc}$ to $z \simeq 4$, but the method assumes low-ionization gas is a faithful tracer, geometrically and kinematically, of neutral hydrogen (Reddy et al. 2016; Vasei et al. 2016). Other methods such as the analysis of recombination lines (Zackrisson, Inoue & Jensen 2013; Zackrisson et al. 2017), requires access to Balmer lines seen beyond 2 µm at high redshift and also necessitates an accurate knowledge of the nature of the stellar population.

In this paper, we propose a new method for estimating $f_{\rm esc}$ at high redshift which is based on examining the cross-correlation between star-forming galaxies and the Lyα absorption spectrum of a background QSO probed in the same cosmic volume. Such an approach (Adelberger et al. 2003, 2005) has been productive at $z \simeq 2-3$ in exploring associations between galaxies and QSOs and their immediate environments (Rudie et al. 2012; Prochaska et al. 2013; Turner et al. 2014), as well as in studies of the reionization of He II (Schmidt et al. 2017). However, the idea is largely unexploited in the H I reionization era other than studies by Díaz et al. (2011, 2014, 2015) (also García et al. 2017; Cai et al. 2017) which focused on the environs of C IV absorption systems at $z \sim 5.7$. In this first paper in the series, we develop the method which exploits the statistical association between star-forming galaxies proximate to the QSO sightline and fluctuations in the Lyα forest in the QSO spectrum. We illustrate the potential via an application to a cosmic volume spanning the redshift range $5.3 \lesssim z \lesssim 6.4$ in the field of the $z = 6.42$ SDSS QSO J1148+5251.

To test the influence of star-forming galaxies and AGN on reionization we propose to establish a direct connection between the distribution of galaxies of known redshift and luminosity and the physical state of the intergalactic medium (IGM) in the same cosmic volume. In this paper we introduce the methodology of how the population-averaged LyC escape fraction can thus be determined. High-resolution spectroscopy of a $z > 6$ QSO provides the redshift-dependent Lyα forest transmission of the IGM and the photoionization rate $\Gamma_{\rm H\,I}$ of the UV background, with the aid of cosmological simulations (Bolton & Haehnelt 2007; Faucher-Giguère et al. 2008; Becker & Bolton 2013). Additionally, spectroscopic follow up of colour-selected Lyman break galaxies (LBGs) provides UV luminosities $L_{\rm UV}$ and precise redshifts in the same volume probed by the background QSO. By predicting the number of ionizing photons emitted from survey galaxies, we can evaluate the contribution of galaxies to the observationally measured UV background. Using the spectroscopically detected luminous LBGs as signposts (e.g. using the host-halo mass) it is possible to estimate the abundance of (unseen) fainter galaxies clustered around them. For this we utilize the results of deeper imaging data which has established the galaxy–halo connection from joint analyses of the well-established luminosity function down to $M_{\rm UV} \simeq -15$ (Atek et al. 2015; Bouwens et al. 2015, 2017; Livermore, Finkelstein & Lotz 2017; Ishigaki et al. 2018; Ono et al. 2018) and clustering measurements (McLure et al. 2009; Barone-Nugent et al. 2014; Harikane et al. 2016, 2018b) in the context of ΛCDM cosmology (e.g. Springel et al. 2005; van den Bosch et al. 2013). The population-averaged LyC escape fraction, $\langle f_{\rm esc}\rangle$ is then obtained by equating the total ionizing output from the combined population of luminous and fainter galaxies with the photoionization rate of the IGM.

A plan of the paper follows. In Section 2, we introduce the necessary observations for our programme which includes broad-band photometry necessary for colour-selection of $z > 5$ LBGs, Keck spectroscopy using the wide-field DEep Imaging Multi-Object Spectrograph (DEIMOS) which yields precise redshifts essential for accurate mapping, and the archival ESI spectrum of QSO J1148+5251. We use these data to produce a catalogue of star-forming galaxies as well as the Lyα transmission spectrum in the same redshift range. We analyse our observations in Section 3, calculating the correlation between our spectroscopically confirmed galaxies and the fluctuations in the Lyα forest which gives us the mean Lyα transmitted flux around galaxies. In Section 4, we discuss the physical origin of the observed Lyα transmitted flux around LBGs and introduce our methodology which takes into account the associated but fainter galaxies which are undetected in our imaging survey thereby deriving a mean escape fraction of LyC photons at $z \simeq 6$. The result is presented in Section 5. In Section 6, we examine two specific cases where sources can be directly associated with features in the Lyα forest which provides insight into the possible contribution of rarer, luminous sources including AGN. In Section 7, we discuss the promise and challenges of our new method and the prospects with further data.

Throughout this paper we adopt the Planck 2015 cosmology $(\Omega_m, \Omega_\Lambda, \Omega_b, h, \sigma_8, n_s) = (0.3089, 0.6911, 0.04860, 0.6774, 0.8159, 0.9667)$ (Planck Collaboration XIII 2016). We use pkpc and pMpc (ckpc and cMpc) to indicate distances in proper (comoving) units. All magnitudes in this paper are quoted in the AB system (Oke & Gunn 1983).

2 OBSERVATIONS

Our choice of the SDSS QSO J1148+5251 at $z = 6.4189$ (RA = 11h48m16.7s +52deg51m50.39s, J2000) for the illustration of our new method was based on the availability of its ESI high signal-to-noise spectrum and deep ground and space-based imaging from which we can photometrically select galaxies in the relevant redshift range. For this QSO the uncontaminated Lyα forest spans the redshift range $5.26 < z < 6.42$. Archival data from the *Spitzer* and *Chandra* Space Telescopes provides additional information on the stellar mass and AGN activity of selected sources in the QSO field (e.g. Jiang et al. 2006; Gallerani et al. 2017).

2.1 Imaging data and photometric catalogue

Deep archival Large Binocular Telescope (LBT) images of the Q1148 field in the SDSS *r*-, *i*-, and *z*-band filters taken by the Large Binocular Camera (LBC) were used to construct a photometric catalogue of *r*- and *i*-dropout candidates for Keck spectroscopic follow-up. LBC pipeline-reduced images reported by Morselli et al. (2014) (PI: R. Gilli)[1] were downloaded from the LBT archive The exposure times were ∼3 hrs in *r* and ∼1.5 hrs in *i* and *z*. This panoramic data

[1] http://www.oabo.inaf.it/ LBTz6/

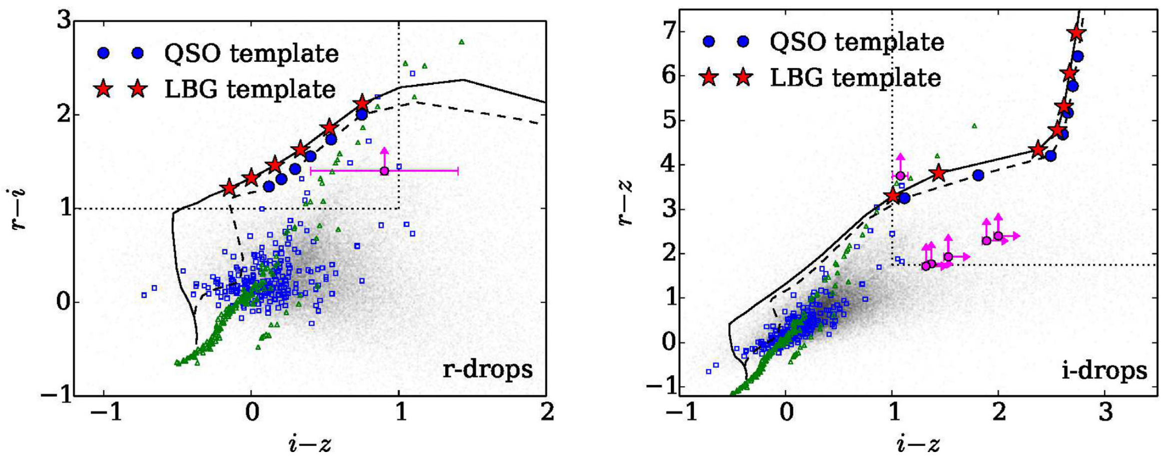

Figure 1. Colour–colour diagram for *r*- (left) and *i*-dropouts (right). The locii of LBGs with $W_{\mathrm{Ly}\alpha}$= 50 Å(red star symbols) and a QSO (filled blue symbols) template spectrum from $z = 5.2, \ldots, 5.7$ (left) and $5.8, \ldots, 6.4$ (right) by 0.1 interval are shown. The magenta points are the spectroscopically confirmed *r*, *i*-dropouts in the Q1148 field. The small black points represent candidates from the photometric catalogue identified by SExtractor. Typical colours for $0 < z < 3$ interlopers (open blue squares) from VUDS-DR1 samples in COSMOS field (Le Fèvre et al. 2015; Tasca et al. 2017) and for Galactic stars (open green triangles) (Gunn & Stryker 1983) are overlaid. Our adopted selection criteria for dropout candidates are indicated by dotted lines.

set covers a field of 23 × 25 arcmin (∼39.5 × 42.5 h^{-1} cMpc at $z = 6$) which covers a substantial fraction of the expected mean free path of ionizing photons at this epoch, $\lambda_{\mathrm{mfp}} \simeq 6.0[(1 + z)/7]^{-5.4}$ pMpc (Worseck et al. 2014) or 17arcmin in radius. From the processed data, we constructed our own photometric source catalogue using SExtractor (Bertin & Arnouts 1996). The limiting magnitudes in each bandpass were estimated by randomly placing fixed 2 arsec apertures in blank regions. We derived 5σ limiting magnitudes of $r = 26.3$, $i = 25.9$, and $z = 25.0$ (and at 2σ, $r = 27.3$, $i = 26.9$, and $z = 26.0$) in agreement with the values reported by Morselli et al. (2014).

In order to select our candidate LBGs in the desired redshift range, we imposed a 5σ detection limit of $z = 25.0$ for our primary selection with fainter secondary candidates at the 3σ limit of $z = 25.6$. We selected candidate LBGs in the sought-after redshift range $5.26 \lesssim z \lesssim 6.42$ according to the following criteria:

$$r - i > 1.0 \text{ and } i - z < 1.0 \tag{1}$$

for *r*-dropouts and

$$i - z > 1.0 \text{ and } [r > r(2\sigma) \text{ or } r - z > 1.75] \tag{2}$$

for *i*-dropouts.

We can visualize the *i*-dropout criteria by considering template spectra for target LBGs and AGN in Fig. 1. Here, a strong Lyα emission line could produce bluer $i - z$ colours and thus a traditional $i - z > 1.3$ colour cut (e.g. Bouwens et al. 2006, 2007) would miss a substantial fraction of objects at $5.3 < z < 5.7$ and ∼20−30 per cent at $z > 5.7$ (Malhotra et al. 2005; Díaz et al. 2011). Likewise Type II QSOs could have a very blue $i - z < 0$ colour at $z > 5.3$ due to the strong Lyα emission line (Meiksin 2006b; Díaz et al. 2011). In Fig. 1, we consider both *r*- and *i*-dropout criteria in the context of the locus of a BPASS galaxy model (version 2.0, Stanway, Eldridge & Becker 2016; Eldridge et al. 2017) with continuous star formation at 100 Myr age, $Z = 0.20\,\mathrm{Z}_{\odot}$ metallicity, and Lyα equivalent width $W_{\mathrm{Ly}\alpha}$= 50 Å, and that of a mean QSO template (Telfer et al. 2002) from redshift 5.3 to 6.4 at 0.1 redshift interval in the context of LBT filters.[2] The IGM transmission is computed using IGMTRANSMISSION code (Harrison, Meiksin & Stock 2011) based on the transmission curves of Meiksin (2006a). The adopted selection criteria, equations (1) and (2), are marked. After applying these criteria, two authors (KK and NL) visually inspected all candidates removing sources contaminated with artefacts, diffraction spikes of nearby stars and sources close to the boundaries of the detector mosaic. There are 124 objects in the final photometric catalogue of *r*- and *i*-drop candidates.

2.2 Galaxy spectroscopy

The photometric candidates were spectroscopically observed through an ongoing survey undertaken with the DEIMOS at the Nasmyth focus of the 10-m Keck II telescope (Faber et al. 2003) on 2017 March 26–27 (PI: Zitrin). Conditions were clear and seeing was typically between 0.9–1.5 arcsec on 26th and 0.7–1.0 arcsec on 27th. We placed one slitmask of 16.7 × 5.0 arcmin^2 field of view so as to maximize the number of dropout targets from the LBT photometric catalogue and encompassing a large volume within the mean free path of ionizing photons at this epoch (Fig. 2). In selecting targets for the mask, greater priority was given to *i*-dropouts to increase the likelihood of detecting Lyα emission in redshift range sampled by the Lyα forest, yielding 45 dropout targets in the mask. A 1.0 arcsec slitwidth was used with the 600 line mm^{-1} grating (600ZD) providing spectroscopic coverage between 4950 Å and 10 000 Å with a spectral resolution of 3.5 Å . The mask was observed for 4.3 h. All data were reduced using the SPEC2D IDL pipeline (Cooper et al. 2012; Newman et al. 2013). The wavelength calibration was done using the afternoon arc lamp. The final reduction provides two-dimensional (2D) spectra and variance arrays. The spectra were visually inspected for emission lines independently by the four of the authors (KK, RSE, NL, and AZ). Two authors (RSE and NL) were blinded from the locations of transmission features in the QSO spectrum (see below) to avoid unconscious biases.

[2]The filter bandpasses were derived from http://abell.as.arizona.edu/ lbtsci/Instruments/LBC/lbc.html

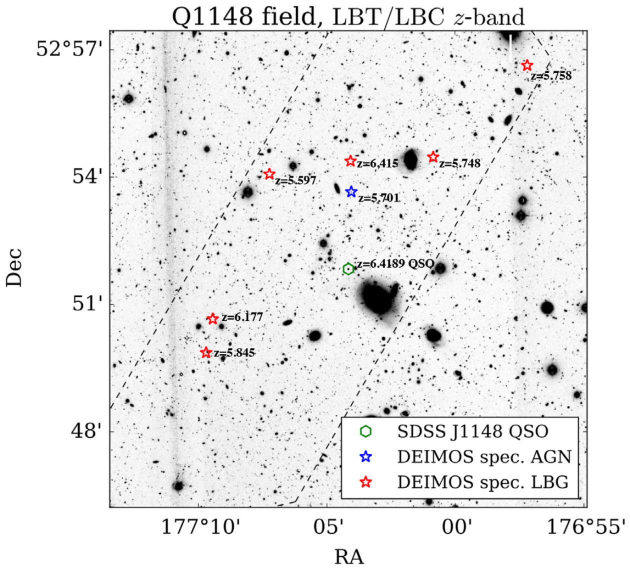

Figure 2. LBT/LBC z-image of the Q1148 field overlaid with spectroscopically identified dropouts (star symbols, red: LBG, blue: AGN) and the background SDSS J1148 QSO (diamond). Each symbol is annotated with the spectroscopic redshift. The DEIMOS footprint is marked (dashed).

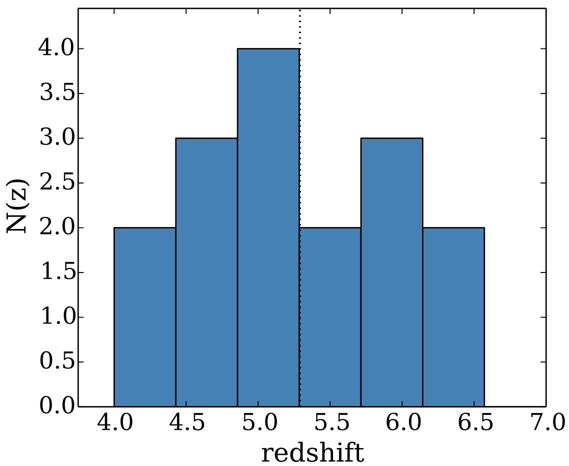

Figure 3. Redshift distribution of the spectroscopically confirmed LBGs. The dashed line indicates the lower limit at $z = 5.3$ for which the Lyα forest can be examined in the ESI spectrum of the QSO.

In total we secured spectroscopic redshifts for 16 sources including a previously identified AGN (Mahabal et al. 2005), corresponding to a $\simeq$35 per cent success rate of spectroscopic confirmation. All emission lines in each 2D spectrum coincide with the expected location of the dropout target on the slit. The overall redshift distribution of the spectroscopic sample is shown in Fig. 3. However, due the limited three bands photometry for the Q1148 field, the photometric redshifts were fairly approximate. Within the $5.3 < z < 6.4$ redshift range which overlaps the volume where the IGM transmission can be traced in the absorption line spectrum of SDSS J1148+5251, we have a sample of six spectroscopically confirmed LBGs plus the AGN (excluding one LBG at $z_{\mathrm{Ly}\alpha} = 6.415$ lying in the proximity zone of the Q1148). Thus, the final success rate of finding galaxies in the Lyα forest region was $\simeq$13 per cent. Spectra of the LBGs and AGN are shown in Figs 4 and 5. The properties of the sources in the relevant redshift range for this study are listed in Table 1.

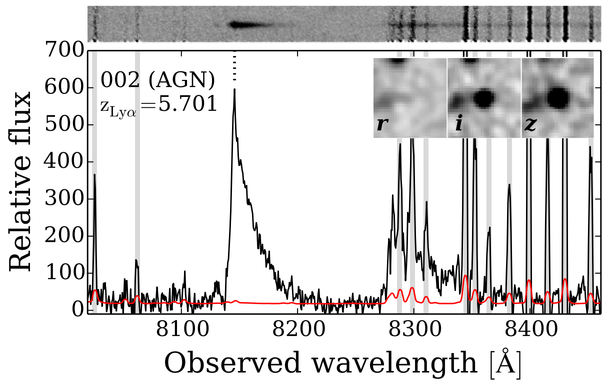

Figure 4. The spectroscopically confirmed faint AGN at $z_{\mathrm{Ly}\alpha} = 5.701$ in the Q1148 field (black: flux, red: noise) with DEIMOS. Skylines (Osterbrock et al. 1996) are marked (grey shaded). The 2D spectrum (top panel) and the postage stamp riz image (inset) are shown. This source (RD J1148+5253) was previously identified by Mahabal et al. (2005).

2.3 QSO spectroscopy and Lyα transmission features

To examine the structure in the Lyα forest of SDSS J1148+5251 QSO (Fan et al. 2003), we used a spectrum taken with the Echellette Spectrograph and Imager (ESI) at the Keck II telescope from a large sample of QSOs uniformly reduced by Eilers et al. (2017) (Fig. 6). The systemic redshift of J1148+5251 is taken from the CO redshift presented in Carilli et al. (2010). The spectral resolution is $R \approx 5000$ sampled with $\sim$5 pixels ($\simeq$10 km s^{-1} per pixel) within one resolution element.

To estimate the wavelength-dependent continuum level, we use a principal component analysis (PCA) as described by Eilers et al. (2017). This PCA-based continuum estimate C_λ is used to calculate the Lyα transmitted flux $F_\alpha = e^{-\tau_\alpha}$,

$$F_\alpha = f_\lambda / C_\lambda + n_\lambda / C_\lambda, \tag{3}$$

where f_λ is the observed flux and n_λ is the noise in the Q1148 ESI spectrum.

To estimate the uncertainty, we also employed an empirical technique based on *HST*/COS spectra of $z \lesssim 1$ UV-bright AGN (Danforth et al. 2016).[3] The continuum level was then estimated for the subset of 17 *HST*/COS continuum spectra classified as type 'QSO'. We compared the continuum redward of the Lyα emission line of the *HST*/COS spectra with the Q1148 ESI spectrum and derived the best-fitting continuum by minimizing the chi-square for $>$1270 Å. Although Q1148 has a weak Lyα emission line, unlike those in the set of the 17 *HST*/COS spectra, this only affects the derived Lyα absorption properties in the vicinity of the QSO, which is not used in the subsequent analysis. Comparing the Lyα transmitted flux between the PCA-based and *HST*/COS-based methods, the difference in the continuum level is $\simeq$ 20 per cent level at median over the redshift range $5.5 < z_{\mathrm{Ly}\alpha} < 6.3$. This is sufficiently small not to affect the subsequent analysis and results in this paper.

We identify Lyα and Lyβ transmission spikes using an automated wavelet-based algorithm. We correlate (i.e. wavelet transform) the continuum-normalized Lyα forest spectrum with a 'Mexican hat' wavelet $\psi_\sigma(x) \propto \sigma^{-1/2}(1 - (x/\sigma)^2)\exp(-x^2/2\sigma^2)$ (normalized with

[3] Publicly available online: https://archive.stsci.edu/prepds/igm/

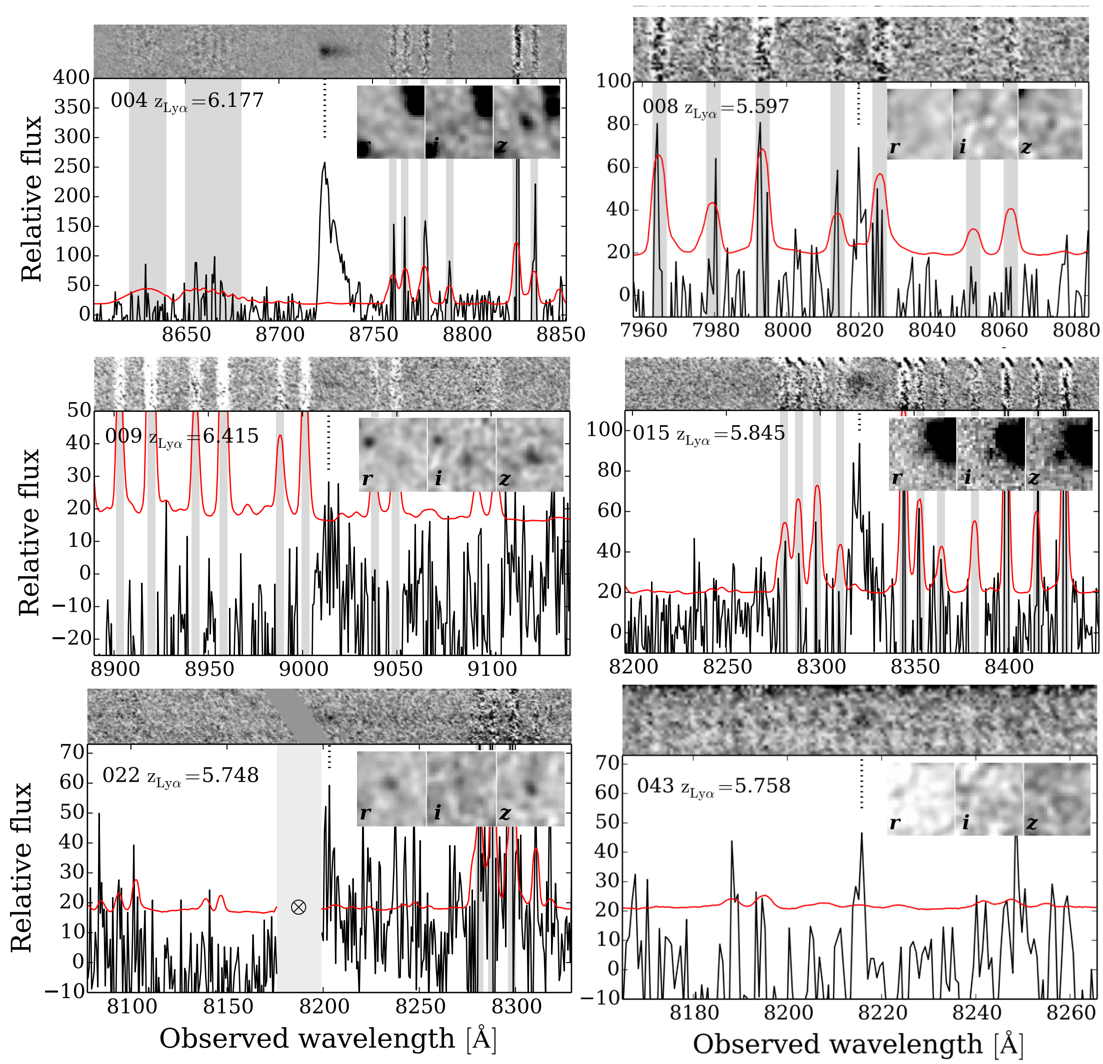

Figure 5. Same as Fig. 4, but for spectroscopically confirmed $z > 5.3$ LBGs in the Q1148 field.

Table 1. DEIMOS spectroscopic catalogue.

ID	$z_{\mathrm{Ly}\alpha}{}^{a}$	RA (J2000)	Dec. (J2000)	r (mag)	i (mag)	z (mag)	$M_{\mathrm{UV}}{}^{b}$	Note
002	5.701	11h48m16.20s	52d53m39.55s	>27.3	24.63 ± 0.06	23.55 ± 0.04	−23.11 ± 0.04	AGN (Lyα+N v)
004	6.177	11h48m37.80s	52d50m39.60s	>27.3	>26.9	25.01 ± 0.14	−21.78 ± 0.14	LBG (Lyα)
008	5.597	11h48m28.98s	52d54m04.50s	>27.3	>26.9	25.53 ± 0.13	−21.10 ± 0.13	LBG (Lyα)
009	6.415	11h48m16.32s	52d54m22.94s	>27.3	>26.9	24.90 ± 0.12	−21.95 ± 0.12	LBG (near Q1148)
015	5.845	11h48m38.83s	52d49m51.97s	26.76 ± 0.69	25.36 ± 0.39	24.46 ± 0.31	−22.24 ± 0.31	LBG (Lyα)
022	5.748	11h48m03.42s	52d54m28.56s	>27.3	>26.9	25.37 ± 0.15	−21.30 ± 0.15	LBG (Lyα)
043	5.758	11h47m48.72s	52d56m37.98s	>27.3	>26.9	25.58 ± 0.16	−21.10 ± 0.16	LBG (Lyα)

[a] By interpreting the peak of the line as Lyα redshift (measured in this work).
[b] Based on the apparent z magnitude, assuming the k-correlation $2.5(\alpha - 1)\log_{10}(1 + z_{\mathrm{Ly}\alpha})$ with a spectral slope $\alpha = 2$.

$\int \psi_\sigma(x)\mathrm{d}x = 0$),

$$w_\sigma(\lambda) = \int F_\alpha(\lambda)\psi_\sigma(\lambda - \lambda')\mathrm{d}\lambda'. \quad (4)$$

The width of the wavelet was varies according to $\sigma = 10, \ldots, 250$ km s^{-1} with 10 km s^{-1} interval. At each wavelength pixel, we record the maximum wavelet coefficient $w_{\max}(\lambda) = \max_{\sigma\in\mathrm{all}} w_\sigma(\lambda)$ for all width choice. Robust transmission spikes are chosen as the local maxima of the wavelet coefficients, $w_{\max}(\lambda)$, whose signal-to-noise ratio at a peak pixel is larger than 5σ. The wavelet-based estimate of the widths of the transmission spikes are recorded as the width at which gives the local maxima of the wavelet coefficients. The method successfully identifies the previous known Lyα transmission spike at $z = 6.083$ (White et al. 2003, 2005; Oh & Furlanetto 2005). The list of the identified Lyα and Lyβ transmission spikes is tabulated in Table 2.

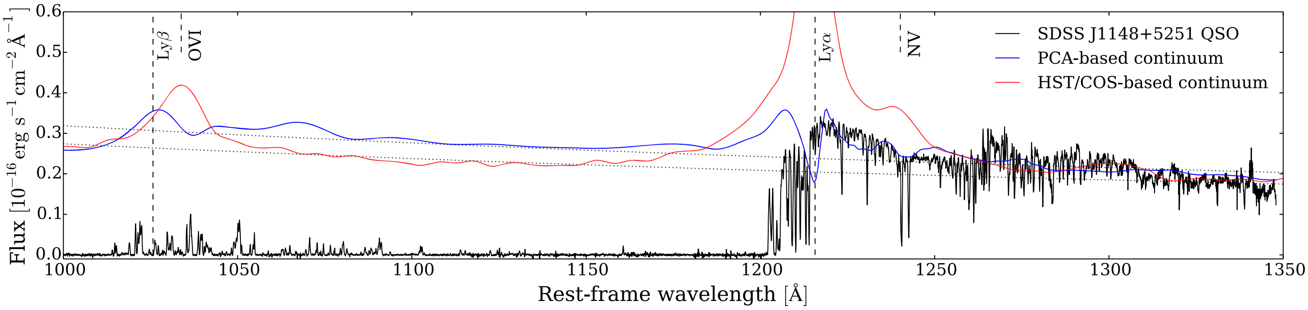

Figure 6. ESI spectrum and estimated continuum level for the SDSS J1148+5251 $z = 6.4189$ QSO (black). The latter is based on the PCA spectrum (blue) and *HST*/COS spectrum (red: continuum of SDSS J0929+4644 $z = 0.24$ QSO). The dotted lines indicate a power-law continuum with $\alpha_\nu = -0.5$.

Table 2. Transmission features at $z > 5.5$ in the Lyα and Lyβ forest regions in the Q1148 ESI spectrum.

z	S/N	z	S/N
	Lyα transmission spikes		
5.527	6.1	5.641	20.0
5.534	30.8	5.647	7.0
5.547	18.1	5.650	16.0
5.551	5.5	5.657	21.5
5.558	20.1	5.675	7.4
5.570	18.5	5.729	8.4
5.588	12.3	5.798	10.0
5.593	32.3	5.806	8.5
5.599	15.2	5.850	9.4
5.624	6.4	5.862	8.9
5.631	22.6	5.901	6.6
5.638	12.8	6.083	15.1
	Lyβ transmission spikes		
6.056	6.0		
6.086	6.6		
6.185	8.7		

3 GALAXY–LYα FOREST CROSS-CORRELATIONS

We now introduce the observed correlation between galaxies and Lyα transmission features in the J1148 QSO field. We focus initially on the 3D mapping of galaxies as it relates to identifiable Lyα transmission spikes and absorption troughs. We then examine the statistical correlation between spectroscopically confirmed galaxies and the Lyα transmitted flux. Later, in Section 4 we discuss the physical basis of this cross-correlation signal and develop a methodology in order to derive a constraint on the mean LyC escape fraction at $z \sim 6$ in Section 5.

3.1 The observed distribution of galaxies around Lyα transmission spikes and absorption troughs

In Fig. 7, we show the spatial distribution of spectroscopically confirmed galaxies from our DEIMOS survey in the context of Lyα forest transmission spikes and absorption troughs in the ESI spectrum of QSO J1148+5251. The continuum normalized QSO spectrum of the transmitted Lyα flux, $e^{-\tau_\alpha}$, is shown with the Lyα redshifts $z_{\mathrm{Ly}\alpha}$ and the physical separation $r_\perp$ of the galaxies relative to the QSO sightline. This 3D mapping of galaxies around the varying Lyα transmission gives us our first glimpse of how galaxies influence the physical state of the IGM at the end of reionization. Three out of our 6 LBGs (at $z_{\mathrm{Ly}\alpha} = 5.597, 5.845, 6.177$) lie close to the vicinity of Lyα and/or Lyβ transmission spikes in the QSO spectrum, while 2 LBGs at $z_{\mathrm{Ly}\alpha} = 5.748, 5.758$ are located close to deep absorption troughs. One of our LBGs at $z_{\mathrm{Ly}\alpha} = 6.415$ resides within the proximity zone of the J1148+5251 QSO (indicated by the blue shaded region). The source at $z_{\mathrm{Ly}\alpha} = 5.701$ is a previously known AGN (Mahabal et al. 2005) and its location is bracketed by two broad Lyα transmission spikes (Gallerani et al. 2008).

It is noteworthy that $\simeq$40 per cent of our spectroscopic sample is found close to Lyα transmission spikes, particularly since the redshift distribution of r, i-dropout selection is quite broad (Vanzella et al. 2009; Stark et al. 2010). However, there may well be selection effects biasing the visibility of Lyα emission in the galaxy sample, e.g. in wavelength regions unaffected by strong skylines. In order to quantify the relative spatial distribution of LBGs and Lyα absorption more rigorously, it is necessary to adopt a statistical approach.

3.2 Statistical H I proximity effect: the mean Lyα transmitted flux around galaxies

To examine the cross-correlation between the location of spectroscopically confirmed galaxies and Lyα forest absorption features, we compute the mean Lyα transmitted flux, $\langle \exp(-\tau_\alpha(r)) \rangle$, around the spectroscopically confirmed LBGs as a function of physical distance r from a galaxy to Lyα forest pixels,

$$\langle \exp(-\tau_\alpha(r)) \rangle = \frac{\sum_{i<\mathrm{pair}(r)} w_i F_{\alpha,i}}{\sum w_i}, \quad (5)$$

where $F_{\alpha,i} = \mathrm{e}^{-\tau_{\alpha,i}}$ is the Lyα transmitted flux at a physical distance r_i from a galaxy of interest. w_i is the weight for galaxy–Lyα forest flux pair, by which we down-weight noisy pixels as $w_i = 1/\sigma^2_{N,i}$. The physical radial distance is computed from $r = \sqrt{r_\perp^2 + r_\parallel^2}$ where $r_\perp = \theta D_A(z_{\mathrm{LBG}})$ and $r_\parallel = \int_{z_{\mathrm{pixel}}}^{z_{\mathrm{LBG}}} c\mathrm{d}z/[(H(z)(1+z)]$ where z_{LBG} and z_{pixel} are the redshifts of a LBG[4] and Lyα forest pixel. We did not divide the Lyα transmitted flux F_i in each pixel by the mean

[4]We take a Lyα redshift as a galaxy redshift, $z_{\mathrm{Ly}\alpha} = z_{\mathrm{LBG}}$. The velocity offsets of Lyα redshifts relative to the systemic galaxy redshifts vary by ~0–500 km s^{-1} (e.g. Mainali et al. 2017, and references therein). At a typical velocity offset $\simeq$200 km s^{-1} the systematic error in distance is $\simeq$300 pkpc at $z = 5.8$. While for small-scale applications this involves a correction (Steidel et al. 2010; Turner et al. 2014), this has a negligible effect on the large-scale cross-correlation presented in this paper.

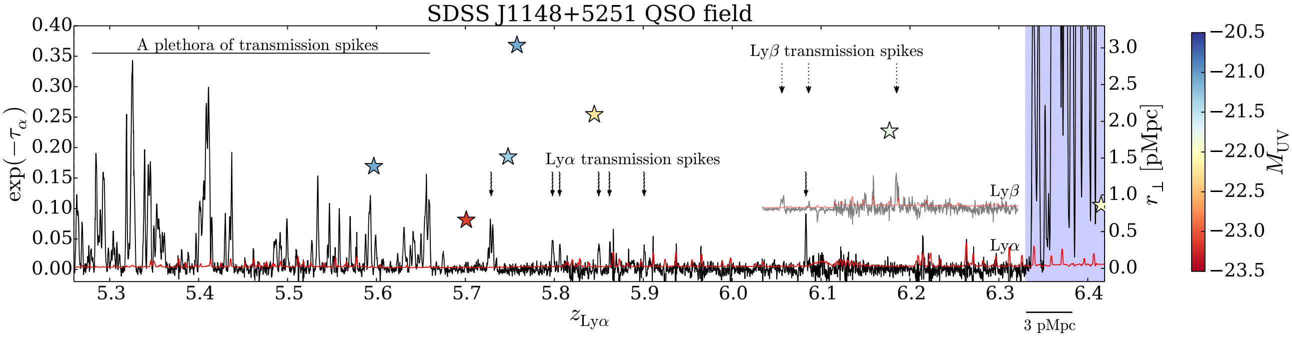

Figure 7. Continuum normalized ESI spectrum (black) of the QSO J1148+5251 plotted in terms of the Lyα opacity, $\exp(-\tau_\alpha)$, alongside the spectroscopically confirmed r, i-dropout galaxies (star symbols). The noise level in the ESI spectrum is shown in red and incorporates OH and O_2 sky line residuals. The region of the absorption spectrum covering the Lyβ forest (between the Lyα and Lyγ forest regions) is shown in grey and offset vertically for convenience. UV luminosities of the spectroscopic sample are indicated by the colour bar and the y-axis refers to the angular distance $r_\perp$ of the galaxies from QSO sightline in proper units. The line-of-sight distance corresponding to $\Delta z = 0.05$ ($\approx$3 pMpc) is indicated by the ruler at the bottom right corner. The proximity zone of J1148+5251 is marked by the blue shaded region. Noticeable Lyα transmission spikes at $z > 5.7$ are marked with arrows, followed by a plethora of transmission spikes at lower redshifts.

Lyα transmission $e^{-\bar{\tau}_{\mathrm{eff}}(z)}$ (to subtract the mean redshift evolution of the IGM, $\bar{\tau}_{\mathrm{eff}}(z)$ (e.g. Fan et al. 2006; Becker et al. 2013) because, at $z \gtrsim 5.75$, the observed Lyα transmitted flux is below the noise level. While equation (5) gives more weight to the Lyα transmission around lower redshift LBGs, it provides the most direct statistical measurement independent of external constraints. A further advantage of this statistical measure is that we need not apply uncertain completeness corrections to our spectroscopic samples. Our procedure provides a measure of the mean H I gas density around *detected* galaxies. This galaxy-centric view contrasts with Lyα forest-centric statistical measures, e.g. the number of galaxies around Lyα transmission spikes, for which completeness corrections in the galaxy sample would be critical.

In Fig. 8, we show the observed mean Lyα transmitted flux around spectroscopically confirmed LBGs with $5.3 < z < 6.3$ as a function of proper distance in the Q1148 field. We consider $\langle z\rangle \simeq 5.8 \pm 0.2$ as the representative redshift based on the mean redshift of the LBG sample. The maximum distance (6 pMpc) is governed by the typical mean free path of ionizing photons at $z \approx 6$ (Worseck et al. 2014). The error is estimated using the Jackknife resampling based on five sub-samples removing one galaxy at a time. As the two innermost bins at $r < 1$ pMpc are based on only one source, we exclude them from the statistical analysis. Although a modest sample, the data presents tentative, intriguing evidence for an increasing Lyα forest transmission closer to the LBGs. This indicates the presence of statistical H I proximity effect at $z \simeq 5.8$. The Spearman rank correlation coefficient is $r_s = -0.47$ which corresponds to a 'moderate' correlation at a $\approx$80$-$90 per cent confidence level (Wall & Jenkins 2012). The correlation is somewhat weaker if the AGN sample is included, degrading the coefficient to $r_s = -0.30$.

Our sample probes only one sightline, thus any interpretation of the positive signal is affected by both potential systematic errors and small number statistics. The apparent hump at $r \approx 4$ pMpc is caused by repeatedly selecting the same prominent Lyα transmission spike at $z \approx 5.73$. We have tested this by artificially masking between $z = 5.64$ and 5.74, where Lyα forest is likely affected by the proximate $z = 5.701$ AGN, and find that the hump is removed. The Jackknife method likely underestimates the error discussed above as the removal of one source near $z \approx 5.7$ contributes little to the variance. At this stage we consider the positive correlation between LBGs and Lyα transmission spikes tentative, but sufficient to demonstrate the potential of our method. Although an increased sample size is clearly required, Fig. 8 demonstrates it is possible to probe the gaseous environment of galaxies at the end of reionization by a spectroscopic survey in $z > 6$ QSO fields.

4 INTERPRETING THE GALAXY–LYα FOREST CROSS-CORRELATIONS

The H I proximity effect is normally thought to arise due to the enhanced UV background around ionizing sources. In this section, we discuss the physical interpretation of the *statistical* H I *proximity effect* seen in the mean Lyα transmitted flux around LBGs in J1148 QSO field. The basis of our method will be to assume that this statistical H I proximity effect arises not only from the detected LBGs but also from undetected faint galaxies which cluster around them. By balancing the ionizing output of this combined population of luminous and fainter galaxies and the UV background via the statistical H I proximity effect, we can constrain the population-averaged LyC escape fraction at $z \simeq 6$. Although the fainter sources cannot be detected in our observing campaign, we will use our spectroscopically detected luminous LBGs effectively as signposts, indicating their likely presence as predicted both by deeper imaging observations and expectations of hierarchical clustering in ΛCDM cosmology.

4.1 Methodology

In order to interpret our data, we have developed a simple radiative transfer model to examine the influence of galaxies on the IGM. Later we use the model to fit the observed mean Lyα transmitted flux around LBGs to derive a constraint on LyC escape fraction. Although more approximate than one based on numerical radiative transfer or radiation hydrodynamic simulations, it has the benefit of illustrating explicitly how various physical processes influence the interaction between galaxies and Lyα forest transmission features.

4.1.1 Model: the mean Lyα transmitted flux around galaxies

The Lyα optical depth around galaxies depends on the density, ionization, and thermal state of the IGM. Using the fluctuating

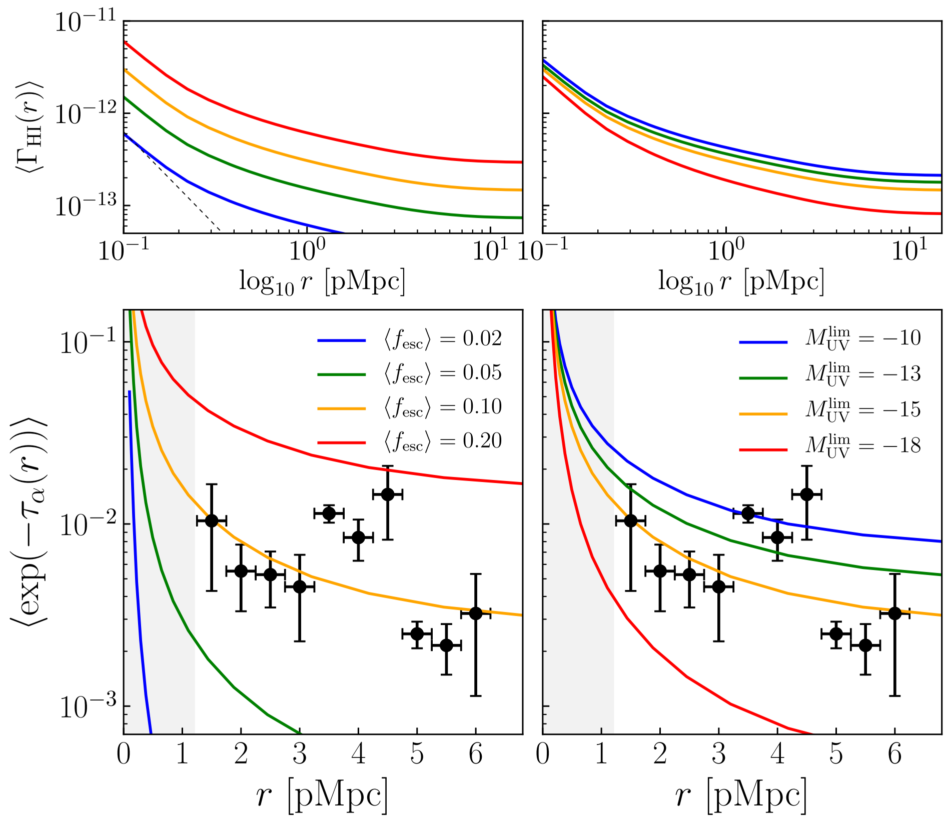

Figure 8. Comparison of the observed mean Lyα transmitted flux around LBGs (black) with the theoretical model at $z = 5.8$. The model shows the contribution to the photoionization rate from sub-luminous galaxies clustered around the LBGs for different values of (*left-hand panel*) the mean LyC escape fraction and (*right-hand panel*) the minimum UV luminosity of ionizing galaxies. In the left (right) panel the value of $M_{\rm UV}^{\rm lim} = -15$ ($\langle f_{\rm esc}\rangle = 0.10$) is fixed. The local contribution from a bright LBG alone is indicated as the dotted line. The average photoionization rate and mean Lyα transmitted flux around LBGs are shown in the top and bottom panels.

Gunn–Peterson approximation (e.g. Becker, Bolton & Lidz 2015a for review), the Lyα optical depth is given by

$$\tau_\alpha \simeq 11\Delta_b^2 \left(\frac{\Gamma_{\rm H\,I}}{10^{-12}\ {\rm s}^{-1}}\right)^{-1} \left(\frac{T}{10^4\ {\rm K}}\right)^{-0.72} \left(\frac{1+z}{7}\right)^{9/2}, \quad (6)$$

where Δ_b is the baryon overdensity, $\Gamma_{\rm H\,I}$ is the H I photoionization rate, T is the temperature of the IGM. For the H I proximity effect, the primarily quantity of interest is the typical H I photoionization rate around a galaxy, $\langle\Gamma_{\rm H\,I}(r)\rangle$, which is enhanced relative to the mean value in the IGM, $\bar{\Gamma}_{\rm H\,I}$. By averaging over many sightlines (ensemble averaging over density fluctuations), the mean Lyα transmitted flux around galaxies is given by

$$\langle\exp(-\tau_\alpha(r))\rangle = \int {\rm d}\Delta_b P_V(\Delta_b) \exp\left[-\bar{\tau}_\alpha(\bar{\Gamma}_{\rm H\,I}, T)\Delta_b^2 \left(\frac{\langle\Gamma_{\rm H\,I}(r)\rangle}{\bar{\Gamma}_{\rm H\,I}}\right)^{-1}\right], \quad (7)$$

where $\bar{\tau}_\alpha(\bar{\Gamma}_{\rm H\,I}, T) \simeq 11 \left(\frac{\bar{\Gamma}_{\rm H\,I}}{10^{-12}\ {\rm s}^{-1}}\right)^{-1} \left(\frac{T}{10^4\ {\rm K}}\right)^{-0.72} \left(\frac{1+z}{7}\right)^{9/2}$ is the optical depth at mean and $P_V(\Delta_b)$ is the volume-weighted density probability distribution function (Miralda-Escudé, Haehnelt & Rees 2000), for which we use the Pawlik, Schaye & van Scherpenzeel (2009) fitting formula based on the cosmological hydrodynamical simulations. We assume a uniform temperature of $T = 10^4$ K as a fiducial value unless otherwise stated, but examine the impact of the IGM temperature later in the paper.

The model embraces a number of physical factors – density fluctuations, UV background, and thermal state of the IGM – important for the mean Lyα transmitted flux around LBGs. We discuss each physical process in the following section.

4.1.2 *Balancing the galaxy abundance with the photoionization rate required by statistical* H I *proximity effect*

To derive a constraint on the LyC escape fraction from the statistical H I proximity effect, we balance the observed galaxy number density with the photoionization rate required from the Lyα transmitted flux. We formulate this cosmological radiative transfer problem using a statistical argument; the full treatment is presented in AppendixA for an interested reader. Here, we focus on the physics essential for understanding the workflow of the methodology. Each star-forming galaxy emits LyC photons at the ionizing photon production rate (Robertson et al. 2013),

$$\dot{N}_{\rm ion} = f_{\rm esc}\xi_{\rm ion}L_{\rm UV}, \quad (8)$$

where $f_{\rm esc}$ is the LyC escape fraction, the LyC photon production efficiency $\xi_{\rm ion}$ is the ratio of ionizing and non-ionizing UV photons, and $L_{\rm UV}$ is the non-ionizing UV (1500 Å) luminosity (in units of ergs^{-1} Hz^{-1}). The total ionizing photon production rate density (in units of photons s^{-1} cm^{-3}) is supplied by all star-forming galaxies above a certain minimum UV luminosity $L_{\rm UV}^{\rm min}$,

$$\dot{\bar{n}}_{\rm ion}(> L_{\rm UV}^{\rm min}) = \langle f_{\rm esc}\xi_{\rm ion}\rangle \int_{L_{\rm UV}^{\rm min}}^{\infty} L_{\rm UV}\Phi(L_{\rm UV}){\rm d}L_{\rm UV}, \qquad (9)$$

where $\langle f_{\rm esc}\xi_{\rm ion}\rangle$ is the population average of the product of the LyC escape fraction and LyC photon production efficiency and $\Phi(L_{\rm UV})$ is the UV luminosity function. $\langle \cdot \rangle$ means the ensemble-averaged quantity.

The UV luminosity function at $z \sim 6$ is now well constrained by both *Hubble* Ultra Deep Field and Frontier Field data (we adopt the UV luminosity function of Bouwens et al. (2015)). Thus, the primary unknowns are $\langle f_{\rm esc}\rangle$ and $L_{\rm UV}^{\rm min}$. Although the unknown parameter always comes in the product, $\langle f_{\rm esc}\xi_{\rm ion}\rangle$, $\xi_{\rm ion}$ can be derived from SED fitting (Bouwens et al. 2016) or UV metal line ratios (Stark et al. 2015, 2017; Matthee et al. 2017; Harikane et al. 2018a).

The independent measure of the ionizing photon production rate density comes from the mean transmitted flux in the Lyα forest, which provides a measure of the H I-photoionization rate of the IGM, $\bar{\Gamma}_{\rm H\,I}$ (e.g. Faucher-Giguère et al. 2008; Becker & Bolton 2013), whence:

$$\bar{\Gamma}_{\rm H\,I} = \int_{\nu_{\rm H\,I}}^{\infty} \sigma_{\rm H\,I}(\nu)\frac{4\pi \bar{J}_\nu}{h\nu}{\rm d}\nu \simeq \frac{\alpha_g}{\alpha_g+3}\sigma_{912}\lambda_{\rm mfp}\dot{\bar{n}}_{\rm ion}(> L_{\rm UV}^{\rm min}), \qquad (10)$$

where $\sigma_{912} = 6.35 \times 10^{-18}$ cm^2 is the H I photonionization cross-section at the Lyman limit and α_g is EUV (>13.6 eV) spectral slope of galaxies. Both the EUV spectral slope α_g and the LyC photon production efficiency $\xi_{\rm ion}$ characterize the hardness of the galaxy spectra; for a given population synthesis model (e.g. Bruzual & Charlot 2003; Eldridge et al. 2017) the best-fitting SED fixes both α_g and $\xi_{\rm ion}$. We use the mean free path of ionizing photons provided by Worseck et al. (2014), $\lambda_{\rm mfp} \simeq 6.0[(1+z)/7]^{-5.4}$ pMpc.

In previous work, Becker & Bolton (2013) (see also Inoue, Iwata & Deharveng 2006; Kuhlen & Faucher-Giguère 2012) have used the *global mean* of the photoionization rate from Lyα forest at $2 < z < 5$ and the observed UV luminosity function of galaxies to derive $\langle f_{\rm esc}\rangle$ at a given $L_{\rm UV}^{\rm min}$. Applying this *global mean* method, however, becomes difficult at $z > 5$ because of the large spatial fluctuations in the intergalactic opacity of the IGM (Becker et al. 2015b; Bosman et al. 2018). The local UV background may differ from the global mean, therefore hindering any balance between the mean galaxy number density and the global mean of the UV background.

The statistical H I proximity effect provides a natural way forward by providing a measure of the local photoionization rate $\langle\Gamma_{\rm H\,I}(r)\rangle$ in the same cosmic volume. The average H I-photoionization rate around a LBG depends on the LyC photons both from a central luminous (detected) LBG and fainter (undetected) galaxies around the central system:

$$\langle\Gamma_{\rm H\,I}(r)\rangle = \langle\Gamma_{\rm H\,I}^{\rm LBG}(r)\rangle + \langle\Gamma_{\rm H\,I}^{\rm CL}(r)\rangle. \qquad (11)$$

The local ionizing effect caused by a spectroscopically detected luminous LBG is

$$\langle\Gamma_{\rm H\,I}^{\rm LBG}(r)\rangle = \frac{\alpha_g\sigma_{912}}{\alpha_g+3}\frac{\langle \dot{N}_{\rm ion}^{\rm LBG}\rangle}{4\pi r^2}{\rm e}^{-r/\lambda_{\rm mfp}}, \qquad (12)$$

where $\langle \dot{N}_{\rm ion}^{\rm LBG}\rangle = \langle f_{\rm esc}\xi_{\rm ion}\rangle\langle L_{\rm UV}\rangle$ is the mean ionizing production rate for which the average UV luminosity is given directly from the observed UV magnitudes. Furthermore, the collective LyC photon flux from the fainter undetected galaxies depends on the *luminosity-weighted* galaxy correlation function $\langle\xi_g(r)\rangle_L$ (or power spectrum $\langle P_g(k)\rangle_L$) between the luminous LBGs and fainter galaxies above a certain minimum UV luminosity $L_{\rm UV}^{\rm min}$ (see Appendix A),

$$\langle\Gamma_{\rm H\,I}^{\rm CL}(r)\rangle = \frac{\bar{\Gamma}_{\rm H\,I}}{\lambda_{\rm mfp}}\int \frac{{\rm e}^{-|\boldsymbol{r}-\boldsymbol{r}'|/\lambda_{\rm mfp}}}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|^2}\left[1+\langle\xi_g(|\boldsymbol{r}'|)\rangle_L\right]{\rm d}^3r',$$
$$= \bar{\Gamma}_{\rm H\,I}\left[1+\int_0^{\infty}\frac{k^2{\rm d}k}{2\pi^2}R(k\lambda_{\rm mfp})\langle P_g(k)\rangle_L\frac{\sin kr}{kr}\right], \qquad (13)$$

where $R(k\lambda_{\rm mfp}) = \arctan(k\lambda_{\rm mfp})/(k\lambda_{\rm mfp})$ is the Fourier transform of the radiative transfer kernel ${\rm e}^{-r/\lambda_{\rm mfp}}/(4\pi r^2\lambda_{\rm mfp})$. The second equality is succinctly expressed in Fourier space. The luminosity-weighted galaxy power spectrum is

$$\langle P_g(k)\rangle_L = \frac{\int_{L_{\rm UV}^{\rm min}}^{\infty} L_{\rm UV}\Phi(L_{\rm UV})P_g(k, L_{\rm UV}){\rm d}L_{\rm UV}}{\int_{L_{\rm UV}^{\rm min}}^{\infty} L_{\rm UV}\Phi(L_{\rm UV}){\rm d}L_{\rm UV}}, \qquad (14)$$

where $P_g(k, L_{\rm UV})$ is the Fourier transform of the galaxy correlation function of LBGs with galaxies of luminosity $L_{\rm UV}$. This captures the contribution of galaxies clustered around the LBGs to the ionizing background. To estimate the galaxy power spectrum, we use the conditional luminosity function (CLF) approach to populate dark matter haloes with galaxies (Yang, Mo & van den Bosch 2003; van den Bosch et al. 2013) described fully in Appendix A. The CLF model is constrained by simultaneously fitting the UV luminosity function of $z \sim 6$ LBGs from *Hubble* Legacy Fields (Bouwens et al. 2015) and the LBG angular correlation function from the *HST*+Subaru/Hyper Suprime-Cam samples (Harikane et al. 2016).

Note that the LyC escape fraction enters as $\langle\Gamma_{\rm H\,I}(r)\rangle \propto \langle f_{\rm esc}\xi_{\rm ion}\rangle$. To see the parameter dependence, it is informative to schematically write:

$$\langle\Gamma_{\rm H\,I}(r)\rangle \propto \langle f_{\rm esc}\rangle \times \frac{\alpha_g\langle\xi_{\rm ion}\rangle}{\alpha_g+3} \times \left[\begin{array}{c}\textit{Galaxy abundance:}\\ \text{LBG} + \textit{galaxy clustering } P_g(k)\end{array}\right], \qquad (15)$$

where we assumed $f_{\rm esc}$ and $\xi_{\rm ion}$ are statistically independent. This highlights how a measure of $\langle\Gamma_{\rm H\,I}(r)\rangle$ from the statistical H I proximity effect is balanced with the galaxy abundance estimate from the luminosity function and angular clustering measurements, leading to a constraint on the product of LyC escape fraction and ionizing photon production efficiency.

Noting the spectral hardness of ionizing sources enters as a combination of the EUV slope and ionizing production efficiency, we define an *effective spectral hardness parameter* $\langle\xi_{\rm ion}^{\rm eff}\rangle$ and assume a fiducial value,

$$\log\langle\xi_{\rm ion}^{\rm eff}\rangle/({\rm erg}^{-1}{\rm Hz}) = \log\left(\frac{\alpha_g\langle\xi_{\rm ion}\rangle}{\alpha_g+3}\right) = 24.8 \text{ (fiducial)}. \qquad (16)$$

We have adopted a canonical value for the ionizing photon production efficiency, $\log\xi_{\rm ion}/({\rm erg}^{-1}{\rm Hz}) = 25.2$ (Robertson et al. 2013) consistent with LBG observations at intermediate redshift (Bouwens et al. 2016; Shivaei et al. 2018). The EUV slope varies from $\alpha_g = 1$ to 3 (Kuhlen & Faucher-Giguère 2012; Becker & Bolton 2013) depending on metallicity and age (Eldridge et al. 2017). For simplicity, we adopt a fiducial value of $\alpha_g = 2$. However, adopting $\alpha_g = 1 - 3$ only changes the value of $\langle\xi_{\rm ion}^{\rm eff}\rangle$ by 0.2 dex, comparable to the typical uncertainty.

In this radiative transfer model, the nominal free parameters of interest are the product of the LyC escape fraction and LyC photon production efficiency, $\langle f_{\rm esc}\xi_{\rm ion}\rangle$, and the minimum UV luminosity of galaxies that contribute to reionization, $L_{\rm UV}^{\rm min}$. We vary both parameters when fitting the model to the observed mean Lyα transmitted flux around LBGs, thereby deriving a constraint on the LyC escape fraction. Before presenting the derived constraint on the LyC escape fraction from the statistical H I proximity effect, we first discuss the impacts of individual physical processes on the mean Lyα transmitted flux around galaxies.

4.2 Physical processes governing the mean Lyα transmitted flux around galaxies

The spatial relationship between galaxies and Lyα forest features carries a wealth of information about the physics of early galaxy formation and reionization.

4.2.1 UV background

Although the UV background includes a contribution from those luminous LBGs detected in our DEIMOS survey, such central LBGs have little impact on the large-scale (>1pMpc) mean Lyα transmitted flux around the LBGs. Their average UV luminosity is $\langle L_{\rm UV}^{\rm LBG}\rangle = 1.9 \pm 0.86 \times 10^{29}$ erg s^{-1} Hz^{-1} where the error indicates the 1σ scatter of luminosities. The local ionizing effect is then

$$\langle\Gamma_{\rm H\,I}^{\rm LBG}(r)\rangle \approx 6.4\times10^{-15} r_{\rm pMpc}^{-2}\left(\frac{\langle f_{\rm esc}\rangle\times\langle\xi_{\rm ion}^{\rm eff}\rangle}{0.1\times10^{24.8}\ {\rm erg^{-1}Hz}}\right)\ {\rm s^{-1}}, \tag{17}$$

for $r \ll \lambda_{\rm mfp}$ and $r_{\rm pMpc} = r/(1\ {\rm pMpc})$ is a distance from the central LBG in proper Mpc. This is more than one order of magnitude lower than the $z\sim$ 6 mean photoionization rate measurement from the mean Lyα transmitted flux of the IGM $\bar{\Gamma}_{\rm H\,I} = 1.8^{+1.8}_{-0.9}\times10^{-13}$ s^{-1} (Wyithe & Bolton 2011). The same would be true even if the ionizing radiation were harder $\log_{10}\xi_{\rm ion}/({\rm erg^{-1}Hz}) = 25.6$ (e.g. Stark et al. 2017) or if we assume a LyC escape fraction of unity. This demonstrates that fainter galaxies, undetected in our survey, are needed to explain the large-scale statistical H I proximity effect. In Fig. 8 the contribution of these fainter galaxies is shown for different values of the mean LyC escape fraction $\langle f_{\rm esc}\rangle$ and the minimum UV luminosity $L_{\rm UV}^{\rm lim}$ (or $M_{\rm UV}^{\rm lim}$) assuming the observed $z\sim6$ UV luminosity function (Bouwens et al. 2015) and angular clustering (Harikane et al. 2016) brighter than $M_{\rm UV}^{\rm lim}$ (see Appendix A). A higher escape fraction increases the average photoionization rate, enhancing the strength of the statistical H I proximity effect. Integrating to a fainter $M_{\rm UV}^{\rm lim}$ clearly has a similar effect.

The radial dependence of the Lyα transmitted flux, however, provides additional information on the clustering bias of ionizing sources, which, in principle, offers a means to break the degeneracy between $\langle f_{\rm esc}\rangle$ and $M_{\rm UV}^{\rm lim}$. Fig. 8 (*right*) shows that if only bright galaxies reionize the IGM, they will be clustered more strongly, producing a somewhat steeper slope of the average photoionization rate and mean Lyα transmitted flux. However, if faint galaxies dominate reionization (extending below the current *Hubble* UV magnitude limit ≈ -15, e.g. Bouwens et al. 2017), their weaker clustering will produce a flatter slope. The luminosity-weighted bias can easily be modelled: on the large scale $\langle P_g(k)\rangle_L \approx b_{\rm LBG}\langle b_g\rangle_L P_m(k)$ we have

$$\langle\Gamma_{\rm H\,I}^{\rm CL}(r)\rangle \approx$$

$$\bar{\Gamma}_{\rm H\,I}\left[1 + b_{\rm LBG}\langle b_g\rangle_L\int_0^\infty\frac{k^2{\rm d}k}{2\pi^2}R(k\lambda_{\rm mfp})P_m(k)\frac{\sin kr}{kr}\right], \tag{18}$$

where $\langle b_g\rangle_L$ is the luminosity-weighted bias factor[5] of ionizing galaxies above $L_{\rm UV}^{\rm min}$:

$$\langle b_g\rangle_L = \frac{\int_{L_{\rm UV}^{\rm min}}^\infty L_{\rm UV}b_g(L_{\rm UV})\Phi(L_{\rm UV}){\rm d}L_{\rm UV}}{\int_{L_{\rm UV}^{\rm min}}^\infty L_{\rm UV}\Phi(L_{\rm UV}){\rm d}L_{\rm UV}}, \tag{19}$$

and $b_{\rm LBG}$ is the bias factor of LBGs ($M_{\rm UV} < -21$) and $b_g(L_{\rm UV})$ is the bias factor of galaxies with luminosity $L_{\rm UV}$. The constraint on $\langle b_g\rangle_L$ from the observed mean Lyα transmitted flux around LBGs can thus be translated to a measure of the minimum UV luminosity once combined with the galaxy luminosity function $\Phi(L_{\rm UV})$ and angular correlation function measurements [i.e. $b_g(L_{\rm UV})$].

The mean free path $\lambda_{\rm mfp}$ of ionizing photons also impacts the radial dependence of the Lyα transmitted flux by setting the maximum distance for influencing the IGM. It is controlled by the number density of H I absorbers, primarily Lyman-limit systems. Our assumed value at $z\sim6$ value is based on an extrapolation of the trend within $2.3 < z < 5.5$ (Worseck et al. 2014). However, hydrodynamical simulations predict $\lambda_{\rm mfp}$ falls markedly at the end of reionization (Gnedin & Fan 2006; Rahmati & Schaye 2018). A further uncertainty may arise if Lyman-limit systems are clustered around galaxies; Rudie et al. (2013) find that inclusion of the CGM of galaxies reduces $\lambda_{\rm mfp}$ by 20 per cent. Ultimately, the galaxy-Lyα forest cross-correlation analysis of many QSO sightlines should be interpreted with detailed hydrodynamical simulations. In this analysis, we quantify this modelling uncertainty by lowering $\lambda_{\rm mfp}$ by 20 per cent (i.e. $\lambda_{\rm mfp} = 4.8$ pMpc) for a comparison.

4.2.2 Gas density fluctuations

The inhomogeneous gas distribution in the IGM has the effect of rendering individual associations between galaxies and Lyα transmission spikes *stochastic*. The Lyα optical depth at the end of reionization, e.g. at $z = 5.8$, is large:

$$\tau_\alpha \approx 48\Delta_b^2\left(\frac{\Gamma_{\rm H\,I}}{2\times10^{-13}\ {\rm s^{-1}}}\right)^{-1}. \tag{20}$$

The level of photoionization rate required by the statistical H I proximity effect is $\langle\Gamma_{\rm H\,I}(r)\rangle \approx 3.1$–$1.6\times10^{-13}$ s^{-1} at radius $r =$ 1–6 pMpc (see Fig. 8), corresponding to the Lyα optical depth value of $\tau_\alpha \approx 32$–61. Thus, observable Lyα transmission spikes only occur within IGM underdensities ($\Delta_b < 1$) even if the UV background is enhanced. The required gas underdensity for producing a Lyα transmission spike larger than $F_\alpha^{\rm th}(= {\rm e}^{-\tau_\alpha^{\rm th}})$ is

$$\Delta_b < \Delta_b^{\rm th} = 0.25\left(\frac{\tau_\alpha^{\rm th}}{3}\right)^{1/2}\left(\frac{\Gamma_{\rm H\,I}}{2\times10^{-13}\ {\rm s^{-1}}}\right)^{1/2}, \tag{21}$$

where $\tau_\alpha^{\rm th}$ is the corresponding pixel optical depth threshold. For a typical identifiable Lyα transmission spike in the Q1148 spectrum (i.e. $\tau_\alpha^{\rm th} = 3$ corresponding to a height $F_\alpha \simeq 0.05$), using the density fluctuations from cosmological simulations (Pawlik et al. 2009), the expected occurrence probability of Lyα transmission spike is found as

$$P(<\Delta_b^{\rm th}) = \int_0^{\Delta_b^{\rm th}} P_V(\Delta_b){\rm d}\Delta_b \simeq 8.7\ \text{per cent} \tag{22}$$

[5] Note that the luminosity-weighted bias factor is typically much larger than the normal bias factor (Croft et al. 2016), contributing to a large spatial cross-correlation.

at $r = 1$ pMpc at an enhanced UV background of $\langle \Gamma_{\mathrm{H\,I}}(r) \rangle \approx 3.1 \times 10^{-13}$ s^{-1} decreasing to $\simeq$1.5 per cent at large distance (for $\langle f_{\mathrm{esc}} \rangle = 0.1$ and $M_{\mathrm{UV}}^{\mathrm{lim}} = -15$). The remaining $\gtrsim$90 per cent of the IGM produces opaque Gunn–Peterson troughs even with an enhanced UV background. Thus, this provides a natural interpretation for the non-exact alignment (see Fig. 7) between a LBG redshift and the nearest Lyα transmission spike. While the enhanced UV background increases the *probability* that the Lyα transmission spikes occur at the IGM around LBGs, but the exact location prefers an underdense IGM.

At smaller radii $\lesssim$1 pMpc approaching the CGM regime, the gaseous overdensity increases. This counteracts with the UV background as $\tau_\alpha \propto \Delta_b^2 \Gamma_{\mathrm{H\,I}}^{-1}$ introducing more absorption and eventually a negative signal in the cross-correlation.[6] In the intermediate redshift range $z \simeq 2$–3, overdensity around LBGs dominates the small-scale mean Lyα transmitted flux (Adelberger et al. 2003, 2005; Crighton et al. 2011; Rakic et al. 2012; Rudie et al. 2012; Tummuangpak et al. 2014; Turner et al. 2014; Bielby et al. 2017), consistent with a wide range of cosmological hydrodynamical simulations (Meiksin, Bolton & Tittley 2015; Rahmati et al. 2015; Meiksin, Bolton & Puchwein 2017; Turner et al. 2017; Sorini et al. 2018).[7] However, the scale where this downturn occurs is $r \lesssim 1.5$ pMpc (Turner et al. 2014; Bielby et al. 2017), i.e. several times the commonly defined CGM scale ($\simeq$300 pkpc). In AppendixB, using the linear theory model we show that the effect of galaxy-gas density correlation is below 10-20 per cent level at $\gtrsim$1 pMpc. Given the range we can measure in the Q1148 field, we therefore expect such small-scale effects to be unimportant.

4.2.3 *Thermal state of the IGM*

Thermal fluctuations of the IGM will introduce further modulation of the Lyα optical depth as $\tau_\alpha \propto \Delta_b^2 \Gamma_{\mathrm{H\,I}}^{-1} T^{-0.72}$, causing the IGM to be more transparent at higher gas temperature. The thermal state of the IGM is primarily controlled by the balance between photoionization heating and the cooling by adiabatic expansion and Compton scattering off CMB photons; it produces a tight asymptotic power-law relation (Hui & Gnedin 1997; McQuinn & Upton Sanderbeck 2016)

$$T = T_0 \Delta^{\gamma - 1}. \tag{23}$$

For $T_0 = 10^4$ K and assuming $\gamma = 1.3$, the Lyα transmitted flux is lower than for the fiducial $\gamma = 1$. This is because the temperature of the underdense IGM which gives rise to Lyα transmission spikes is lower (e.g. $\log_{10} T/\mathrm{K} = 3.82$ at $\Delta_b = 0.25$). Cosmological radiative transfer simulations find a large scatter around $\gamma = 1$ in the temperature–density relation just after the IGM is reionized (Tittley & Meiksin 2007; Trac, Cen & Loeb 2008; Kakiichi et al. 2017; Keating, Puchwein & Haehnelt 2018), which is not captured by the single power-law relation. Thus, we adopt a uniform temperature for simplicity for a fiducial analysis, but also repeat the analysis with $T = T_0 \Delta^{\gamma - 1}$ assuming $T_0 = 10^4$ K and $\gamma = 1.3$. The increased opacity arising from temperature fluctuations requires more ionizing photons to match the statistical H I proximity effect and hence a higher LyC escape fraction.

Large-scale thermal fluctuations may also be caused by environmental effects in the reionization process. In 'inside-out' reionization, highly biased regions around luminous galaxies are thought to have ionized earlier, allowing more time for the gas to cool by adiabatic expansion and CMB Compton cooling. This causes the low-density IGM near luminous galaxies to be preferentially cooler (D'Aloisio, McQuinn & Trac 2015), reducing the mean Lyα transmitted flux around LBGs at inner radii (Davies, Becker & Furlanetto 2017). The extent of this effect is debated (e.g. Keating et al. 2018). For the low-density IGM close to luminous galaxies, the temperature asymptotically relaxes to the value set by the balance between the adiabatic expansion and instantaneous photoionization rate. On the other hand, the IGM away from the galaxies that has been engulfed by a H II I-front raises the temperature to about $\sim 10^4$ K. The large-scale thermal fluctuations vary from $\sim$5000 K to $T \approx 1.0 - 1.5 \times 10^4$ K which contributes to the negative correlation of the mean Lyα transmitted flux around LBGs. As the temperature has a weaker dependence on the optical depth $\tau_\alpha \propto \Delta_b^2 \Gamma_{\mathrm{H\,I}}^{-1} T^{-0.72}$, this can easily be compensated by only moderate enhancement of the UV background. Although both UV background and thermal fluctuations co-exist, because of the steeper dependence on the photoionization rate it is likely that the UV background variation dominates creating a positive correlation, with secondary modulation by thermal fluctuations weakening it (Davies et al. 2017, private communication).

5 CONSTRAINING THE MEAN ESCAPE FRACTION

We now utilize the foregoing to analyse the balance between inferred galaxy abundance in the Q1148 field with the observed mean Lyα transmitted flux in terms of a statistically averaged LyC escape fraction $\langle f_{\mathrm{esc}} \rangle$. To accomplish this we fit the model to the observed mean Lyα transmitted flux data using a Markov Chain Monte Carlo (MCMC) method (Foreman-Mackey et al. 2013) varying $\langle f_{\mathrm{esc}} \rangle$ and $M_{\mathrm{UV}}^{\mathrm{lim}}$. We assume a Gaussian likelihood and place a flat prior in the range of $-2 < \log_{10}\langle f_{\mathrm{esc}} \rangle < 0$ and $-18 < M_{\mathrm{UV}}^{\mathrm{lim}} < -10$. We have tested the result against an enlarged prior range ($-20 < M_{\mathrm{UV}}^{\mathrm{lim}} < -8$) and find a consistent result. For the covariance matrix we only use diagonal elements from the Jackknife error estimate.

In Fig. 9 we show the derived constraint on the $\langle f_{\mathrm{esc}} \rangle - M_{\mathrm{UV}}^{\mathrm{lim}}$ plane. The inferred mean LyC escape fraction at $z \simeq 6$ is found to be

$$\langle f_{\mathrm{esc}} \rangle = 0.083^{+0.037}_{-0.016} \left(\frac{\langle \xi_{\mathrm{ion}}^{\mathrm{eff}} \rangle}{10^{24.8}\ \mathrm{erg^{-1} Hz}} \right)^{-1}, \tag{24}$$

for $M_{\mathrm{UV}}^{\mathrm{lim}} = -14.53^{+2.71}_{-2.53}$ for the fiducial analysis.[8] This constraint is dependent upon the assumed mean free path and IGM temperature. None the less, as discussed in the previous section, a lower mean free path and thermal fluctuations would mean a larger (>10 per cent) mean LyC escape fraction to compensate the increased opacity. These uncertainties on radiative transfer can be included in the MCMC analysis once a larger data set becomes available.

Although our sample is modest, our result suggests that $\langle f_{\mathrm{esc}} \rangle = 0.06$–$0.16$ for star-forming galaxies above $M_{\mathrm{UV}}^{\mathrm{lim}} = -14.53^{+3.16}_{-2.47}$ including modelling systematic error. In Fig. 10 we

[6] As the probability distribution function $P_V(\Delta_b)$ adopted here is measured from the entire simulation box (Pawlik et al. 2009), the effect of a gaseous overdensity around galaxies is ignored in the model.

[7] At scales less than $\sim$100 pkpc, galactic feedback and hydrodynamic processes complicate the distribution of cold gas.

[8] Note that for fiducial analysis we have ignored the three radial bins at 3.5–4.5 pMpc as they are likely affected by systematics. Their inclusion would give a 12 per cent larger $\langle f_{\mathrm{esc}} \rangle$ with two possible best-fitting values of $M_{\mathrm{UV}}^{\mathrm{lim}}$ due to the poor constraint on the shape.

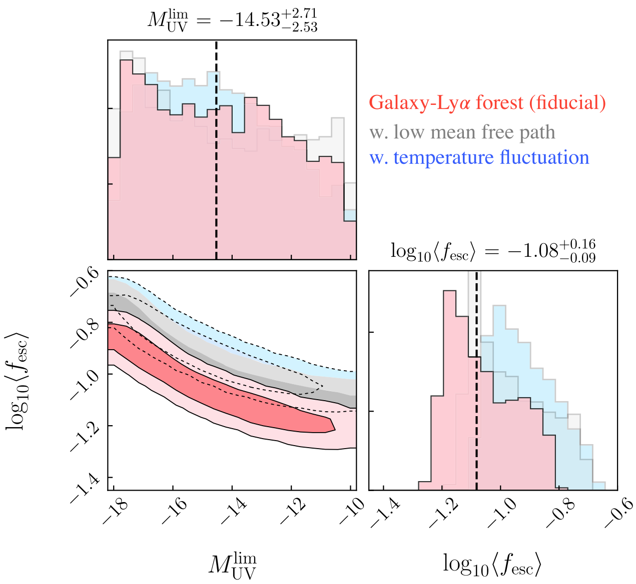

Figure 9. Constraints on the average LyC escape fraction $\langle f_{\rm esc}\rangle$ and the minimum UV luminosity $M_{\rm UV}^{\rm lim}$ with 68 per cent and 95 per cent confidence intervals for the fiducial galaxy–Lyα forest cross-correlation analysis (red: $\lambda_{\rm mfp} = 6$ pMpc, $T = 10^4$ K) and with a lower value of mean free path (grey: $\lambda_{\rm mfp} = 4.8$ pMpc) and with a temperature–density relation (blue: $T_0 = 10^4$ K, $\gamma = 1.3$). The quoted constraint is from the fiducial analysis.

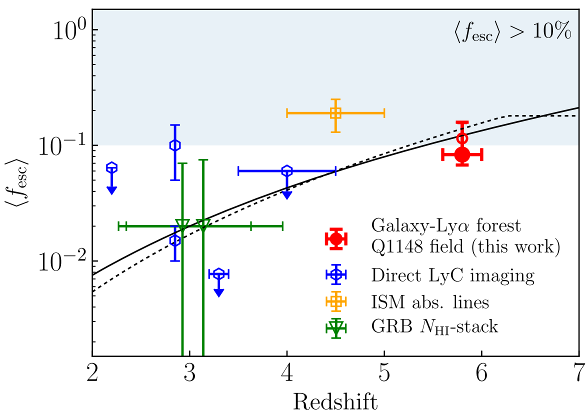

Figure 10. Redshift evolution of the population-averaged LyC escape fraction of galaxies. The $z \simeq 6$ constraint from the galaxy–Lyα forest cross-correlation in Q1148 field is indicated by the filled red circle. A compilation of previous $2 < z < 4$ constraints is indicated by open symbols. These include direct LyC imaging (Vanzella et al. 2010; Mostardi et al. 2013; Grazian et al. 2016; Matthee et al. 2017) and GRB N-stacking (Chen, Prochaska & Gnedin 2007; Fynbo et al. 2009), and ISM absorption line studies (Leethochawalit et al. 2016). The model mean LyC escape fractions adopted by Haardt & Madau (2012) (solid) and Puchwein et al. (2018) (dotted) are overlaid. The shaded region indicate $\langle f_{\rm esc}\rangle > 10$ per cent required for galaxies to drive reionization.

compare our $\langle f_{\rm esc}\rangle$ constraint with earlier estimates from LyC imaging at $z\sim 2-4$ (Vanzella et al. 2010; Mostardi et al. 2013; Grazian et al. 2016; Matthee et al. 2017). Low escape fractions at $z \sim 3$, $\langle f_{\rm esc}\rangle = 0.02 \pm 0.02$ (<0.075 at 95 per cent confidence upper limit), are also indicated from H I covering fractions derived from the spectra of long-duration gamma-ray bursts (GRB) (Chen et al. 2007; Fynbo et al. 2009). Our new estimate suggests a rising mean escape fraction with increasing redshifts consistent with the trend adopted by the recently revised synthesis model of the cosmic UV background (Puchwein et al. 2018) and the minimal reionization model of Haardt & Madau (2012). This means that faint galaxies deposit sufficient ionizing radiation into the IGM for driving the reionization process (see also Faisst 2016). Since the inclusion of temperature fluctuations would require more ionizing photons to match the observed positive correlation of the mean Lyα transmitted flux around LBGs, our fiducial analysis provides a fairly conservative lower limit to the mean LyC escape fraction.

Table 3. Summary of the IGM environment of the $z = 6.177$ luminous Lyα emitting LBG. The associated Lyβ transmission spike is the evidence of highly ionized intergalactic gas around the LBG, which is maintained likely by the faint galaxy overdensity (indicated by the excess O I absorbers).

LBG's Lyα redshift	$z = 6.177$
Lyβ transmission spike	$z = 6.185$
Lower limit to the H II bubble size[a]	> 1.9 pMpc ($9.4\,h^{-1}$ cMpc)
Photoionization rate of the LBG[b]	$\Gamma_{\rm HI}^{\rm LBG} \simeq 2.1 \times 10^{-15}$ s^{-1}
Photoionization rate at the Lyβ spike[c]	$\Gamma_{\rm HI}^{\rm spike} \simeq 5.7 \times 10^{-13}$ s^{-1}
O I absorbers' redshift (distance to the Lyβ spike)	$z = 6.1293, 6.1968, 6.2555$ (3.2, 0.7, 4.0 pMpc)

[a] From the distance between the LBG and Lyβ spike.
[b] At the Lyβ spike (i.e. 1.9 pMpc distance from the LBG) and for SFR = 28 M$_\odot$ yr^{-1}, $f_{\rm esc} = 0.1$, and $\xi_{\rm ion} = 10^{25.2}$ erg^{-1} Hz.
[c] The expected median value of the photoionization rate at the location of the Lyβ spike (see Fig. 12).

6 THE IMPACT OF LUMINOUS SYSTEMS

Finally, we turn our attention to two individual cases of a LBG and AGN for which we can identify associated transmission spikes in the Q1148 spectrum. We investigate both as examples of spatial fluctuations in the IGM environment induced by luminous sources. We discuss how they might contribute to spatial fluctuations of the ionization and thermal states of the IGM and the possible role of rare, luminous sources on the reionization process.

6.1 $z = 6.177$ LBG J1148+5250 and Lyβ transmission spike

LBG J1148+5250 is a newly discovered Lyα emitting galaxy in our DEIMOS sample. It is a luminous ($M_{\rm UV} = -21.8$) galaxy with a secure asymmetric Lyα line at $z_{\rm Ly\alpha} = 6.177$. Interestingly, the LBG redshift coincides with that of a Lyβ transmission spike at $z = 6.185$. This is the first case of a possible individual transverse proximity effect around a $z > 6$ LBG (Table 3). The Lyβ transmission spike is separated by $d_{\rm spike} = 1.9$ pMpc ($9.4\ h^{-1}$cMpc) from the LBG.

The detection of a Lyβ transmission spike and the high optical depth in the Lyα forest region (see Fig. 11) places a bound on the Lyα transmission of the IGM. The peak transmitted flux is $e^{-\tau_{\alpha+\beta}} = 0.0686 \pm 0.0066$ ($\tau_{\alpha+\beta} = 2.68$). Because the high-redshift ($z > 6$) Lyβ forest overlaps with its lower redshift ($z < 5.26$) Lyα equivalent, this translates into an upper limit on the $z = 6.185$ Lyβ optical depth $\tau_\beta < \tau_{\alpha+\beta}$ and, using the ratio between the Lyβ and Lyα optical depths $\tau_\beta/\tau_\alpha = f_{13}\lambda_\beta/(f_{12}\lambda_\alpha) = 0.16$ predicted by atomic physics, a range of

$$4.2\ (3\sigma) < \tau_\alpha < 16.7 \pm 0.6, \tag{25}$$

consistent with the absence of a clear Lyα transmission spike above the 3σ noise in the QSO spectrum.

Compared to the Gunn–Peterson optical depth at $z = 6.185$, $\tau_{\rm GP} \simeq 1.8 \times 10^5 x\Delta_b$, this upper limit on τ_α is quite low, suggesting that

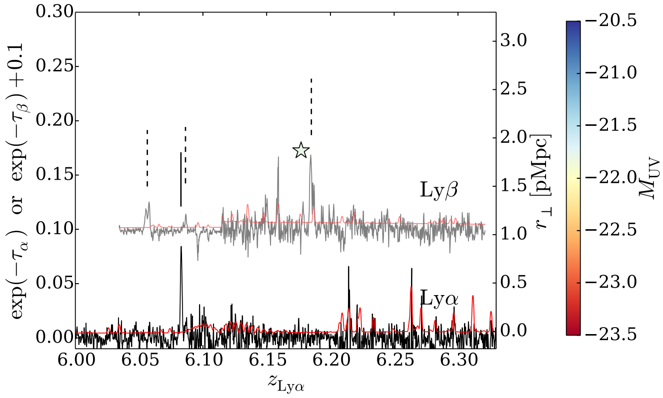

Figure 11. A zoom-in of Fig.7 around the luminous LBG J1148+5250 at $z_{\mathrm{Ly}\alpha}=6.177$ adopting the same colour bar for the galaxy luminosity. Solid and dashed vertical lines indicate the location of wavelet-identified Lyα and Lyβ transmission spikes. The Lyβ forest region is offset by 0.1 in *y*-axis.

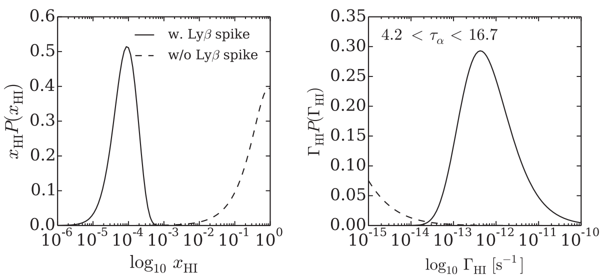

Figure 12. The probability distribution of (*left-hand panel*) the neutral hydrogen fraction and (*right-hand panel*) the H I photoionization rate at the location of the Lyβ transmission spike $z=6.185$ (solid). For comparison, the dashed line shows a hypothetical case without a Lyβ transmission spike (assuming the Lyα optical depth can reach the Gunn–Peterson optical depth of a fully neutral medium, $\tau_\alpha=1.8\times10^5$).

the IGM is highly ionized to $x\lesssim10^{-4}$. As discussed in Section 4.2, the association of individual galaxies and transmission spikes is probabilistic owing to the gas density fluctuations. Thus, we should assess the probability distribution of the neutral hydrogen fraction x at the location of the Lyβ transmission spike given an observed Lyα optical depth. Using the simulated probability distribution function of gas density fluctuations and $\tau_\alpha=\bar{\tau}_{\mathrm{GP}}x_{\mathrm{H\,I}}\Delta_b$ where $\bar{\tau}_{\mathrm{GP}}=1.8\times10^5$ is the Gunn–Peterson optical depth of a fully neutral medium at mean density, we find that

$$P(x_{\mathrm{H\,I}}|\tau_\alpha)=\int\delta_D\left(x_{\mathrm{H\,I}}-\frac{\tau_\alpha}{\bar{\tau}_{\mathrm{GP}}}\Delta_b^{-1}\right)P_V(\Delta_b)\mathrm{d}\Delta_b. \qquad (26)$$

Fig. 12 (left) shows the resulting probability distribution of the neutral fraction x after marginalizing over the observed bound of the Lyα optical depth. The presence of a Lyβ transmission spike indeed indicates that the $z=6.185$ IGM is highly ionized to the expected value of $x\simeq10^{-4}$. Note that this analysis does not assume the medium is photoionized a priori. Thus, a UV luminous galaxy at the reionization epoch ($z>6$) is clearly located in a highly ionized environment.

The distance to the Lyβ transmission spike from LBG J1148+5250 provides a lower limit to the size of the cosmological H II region,

$$R_{\mathrm{H\,II}}>d_{\mathrm{spike}}=1.9\ \mathrm{pMpc}\ (9.4\ h^{-1}\mathrm{cMpc})\quad\text{at}\quad z=6.18. \qquad (27)$$

Can this luminous galaxy alone produce such a large ionized bubble? The UV luminosity $L_{\mathrm{UV}}=2.25\times10^{29}$ erg s^{-1} Hz^{-1} corresponds to a star formation rate SFR $=28.1$ M$_\odot$ yr^{-1} assuming a Salpeter IMF and solar metallicity (Madau, Pozzetti & Dickinson 1998) before any correction for dust extinction. We can estimate the size of the H II region

$$R_{\mathrm{H\,II}}=\left[\frac{3}{4\pi}\frac{\dot{N}_{\mathrm{ion}}^{\mathrm{H\,II}}t_{\mathrm{age}}}{\bar{n}_{\mathrm{H}}(z)}\right]^{1/3},$$

$$\approx1.0\left[\left(\frac{f_{\mathrm{esc}}}{0.1}\right)\left(\frac{\xi_{\mathrm{ion}}}{10^{25.2}\ \mathrm{erg}^{-1}\mathrm{Hz}}\right)\left(\frac{t_{\mathrm{age}}}{300\ \mathrm{Myr}}\right)\right]^{1/3}\mathrm{pMpc}. \qquad (28)$$

assuming a constant star formation history over the median age of UV luminous galaxies ($L>L^*$) at $z\simeq6$ of $\simeq$200–300 Myr (Curtis-Lake et al. 2013). Even for a hard $\log_{10}\xi_{\mathrm{ion}}=25.6$, the radius becomes $R_{\mathrm{H\,II}}\approx1.4$ pMpc below the observed lower limit. It therefore seems necessary to invoke a contribution from fainter galaxies clustered around the luminous LBG.

The probability distribution of the photoionization rate inside the H II region can be estimated as in equation (26) by integrating the Dirac delta function at $\Gamma_{\mathrm{H\,I}}\propto\tau_\alpha^{-1}\Delta_b^2$ with $P_V(\Delta_b)$. Fig. 12 shows that the expected photoionization rate at the Lyβ transmission spike may be as high as $\Gamma_{\mathrm{H\,I}}\simeq10^{-12}$–$10^{-13}$ s^{-1}, close to the value indicated by the statistical analysis in Section 4.2. Such a high photoionization rate cannot be maintained by the luminous LBG alone, which contributes up to $\Gamma_{\mathrm{H\,I}}^{\mathrm{LBG}}(r)\approx7.6$–$19.0\times10^{-15}(r/1\ \mathrm{pMpc})^{-2}$ s^{-1} for $f_{\mathrm{esc}}=0.1$ and $\log_{10}\xi_{\mathrm{ion}}=25.2$–$25.6$.

Becker et al. (2006) report the discovery of four O I absorbers at $z=6.0097$, 6.1293, 6.1968, 6.2555, which indicates the location of low-luminosity galaxies (Finlator et al. 2013) below the LBT detection limit ($M_{\mathrm{UV}}\simeq-21$). The closest $z=6.1968$ O I absorber is separated by $\simeq$0.7 pMpc from the Lyβ transmission spike. Such a surprising excess of O I absorbers near the $z\simeq6.18$ luminous LBG - Lyβ transmission spike association supports the presence of clustered faint galaxies around the LBG, and their collective ionizing contribution.

In summary, the discovery of a Lyβ transmission spike near the $z\simeq6.18$ LBG further supports the conclusion of our statistical analysis. Accelerated reionization is likely driven by the collective ionizing contribution from fainter galaxies clustered around luminous LBGs, possibly enhanced with a harder ionizing spectrum.

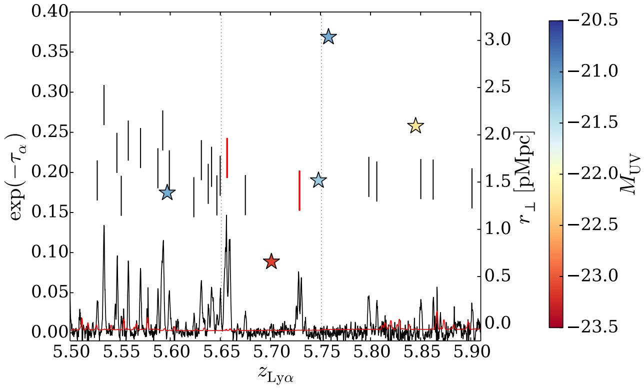

Figure 13. A zoom-in of Fig. 7 around the faint AGN RD J1148+5253 at $z_{\rm Ly\alpha} = 5.701$. Solid vertical lines indicates the wavelet-identified Lyα transmission spikes. The red lines mark the two broad Lyα transmission spikes. The $\Delta z = 0.1$ (≈6.8 pMpc) region around the AGN is shown by the dotted vertical lines.

Table 4. Summary of the IGM environment of the $z = 5.701$ faint AGN. The associated two broad Lyα transmission spikes may be due to the impact of the AGN on the ionization and thermal state of the IGM.

AGN's Lyα redshift	$z = 5.701$
Broad Lyα transmission spikes	$z = 5.657, 5.729$
(distance to the AGN)	(3.07, 2.02 pMpc)
Photoionization rate of the AGN[a]	$\Gamma^{\rm LBG}_{\rm H\,I} \simeq 1.0, 2.2 \times 10^{-13}$ s^{-1}
Photoionization rate at the Lyα spikes[b]	$\Gamma^{\rm spike}_{\rm H\,I} \simeq 8.2, 7.0 \times 10^{-13}$ s^{-1}
An estimated He III bubble size	3.2 pMpc$(t_{\rm Q}/10^8\ {\rm yr}^{-1})$

[a] At the broad Lyβ spikes, 3.07, 2.02 pMpc, respectively. We assumed the 100 per cent escape fraction of the AGN.
[b] The expected median value of the photoionization rate at the location of the broad Lyα spikes at 3.07, 2.02 pMpc, respectively.

6.2 A faint AGN and broad Lyα transmission spikes at $z \sim 5.7$

RD J1148+5253 is a low-luminosity ($M_{\rm UV} = -23.1$) AGN with a redshift of $z_{\rm Ly\alpha} = 5.701$ originally discovered by Mahabal et al. (2005). We confirm the faint AGN with our deep 4.2 h DEIMOS spectroscopy via detection of a broad Lyα, N V λ1240, Lyβ emission lines, and the associated continuum (Fig. 4). Assuming the Eddington luminosity for RD J1148+5253, the super massive black hole (SMBH) mass is estimated to be $M_{\rm BH} \approx 5 \times 10^7$ M$_\odot$ after applying the bolometric correction of 4.4 (Willott et al. 2010) to the observed UV luminosity. Non-thermal emission is evident from the high-ionization metal line N V as well as from the tentative 2.6–2.9σ detection of X-ray emission from 78 ks *Chandra* observation (Gallerani et al. 2017).

In the Mpc-scale environment around the AGN (see Fig. 13 and Table 4), the spectrum of QSO J1148+5251 exhibits two prominent broad Lyα transmission spikes at $z_{\rm Ly\alpha} = 5.729$ and 5.657 located at 2.02 pMpc and 3.07 pMpc away from the AGN, respectively. Fig. 14 shows that the widths of both transmission spikes are broader (≃220 km s^{-1}) than others ($\lesssim$110 km s^{-1}) in the redshift range $5.5 < z < 5.9$. Could the faint AGN impact the physical origin of the broad Lyα transmission spikes?

The hard ionizing spectra of an AGN will enhance the local UV background. Assuming a broken power-law spectrum $L_\nu \propto \nu^{-0.5}$ for 1050 Å $<\lambda <$ 1450 Å and $L_\nu \propto \nu^{-1.5}$ for $\lambda <$ 1050 Å (e.g. Telfer et al. 2002), the H I ionizing photon production rate of RD J1148+5253 is $\dot{N}^{\rm H\,II}_{\rm ion} = \int_{\nu_{\rm H\,I}}^\infty L_\nu/(h\nu){\rm d}\nu \approx 5.3 \times 10^{55}$ s^{-1}, providing

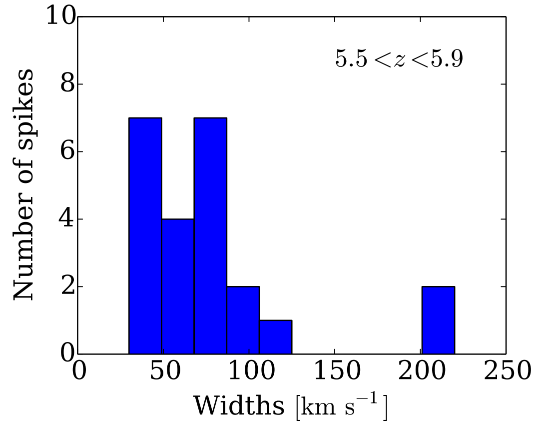

Figure 14. Histogram of the widths of wavelet-identified Lyα transmission spikes in the redshift range of $5.5 < z < 5.9$. The two values with ≃220 km s^{-1} width are those indicated by the red lines in Fig. 13.

the photoionization rate of

$$\Gamma^{\rm AGN}_{\rm H\,I} = \frac{\alpha_Q}{3+\alpha_Q}\frac{\sigma_{912}\dot{N}^{\rm H\,II}_{\rm ion}}{4\pi r^2} \approx 1.0 \times 10^{-13}\left(\frac{r}{3\ {\rm pMpc}}\right)^{-2}\ {\rm s}^{-1}, \quad (29)$$

for $\alpha_Q = 1.5$. At the location of the broad Lyα transmission spikes ($r = 2$–3 pMpc), the faint AGN alone gives an optical depth $\tau_\alpha \approx 48$–$110\Delta_b^2(T/10^4\ {\rm K})^{-0.72}$. While the ionizing contribution of the faint AGN is somewhat larger than a luminous LBG, once again in order to match the observed spikes ($\tau_\alpha \simeq 2$–3), either a gaseous underdensity ($\Delta_b \lesssim 0.25$) or associated fainter ionizing sources are required.

This faint AGN may drive thermal fluctuations of the IGM through He II photoheating (Bolton et al. 2012). Although outside the DEIMOS wavelength coverage, N V emission indicates that there should be photons above 54.4 eV to ionize He II → He III. Using the EUV spectral slope of $\alpha_Q = 1.5$, the He II ionizing photon production is $\dot{N}^{\rm He\,II}_{\rm ion} \approx 6.6 \times 10^{54}$ photons s^{-1}. The size of the He III region so produced is

$$R_{\rm He\,III} = \left[\frac{3}{4\pi}\frac{\dot{N}^{\rm He\,II}_{\rm ion} t_Q}{(Y/X)\bar{n}_{\rm H}(z)}\right]^{1/3},$$

$$\approx 3.2\left(\frac{\dot{N}^{\rm He\,II}_{\rm ion}}{6.6\times 10^{54}\ {\rm s}^{-1}}\right)^{1/3}\left(\frac{t_Q}{10^8\ {\rm yr}}\right)^{1/3}\ {\rm pMpc}. \quad (30)$$

A fiducial AGN lifetime of order 10^8 yr can be estimated from the time-scale required to grow the relevant SMBH. For Eddington-limit accretion, even a massive 100 M$_\odot$ black hole seed requires $t = t_{\rm BH}\ln(M_{\rm BH}/M_{\rm seed}) \approx 5.8 \times 10^8$ yr where $t_{\rm BH} \approx 4.4 \times 10^7(\epsilon_r/0.1)$ yr. Therefore, based on the BH growth time-scale and outflow time-scale[9] arguments, during the plausible AGN lifetime, the two broad Lyα transmission spikes lie within the region of influence of the He III I-front of the AGN.

[9]The upper age limit can be estimated from the non-detection of metal line absorbers at the redshift of RD J1148+5253 AGN. There is broad absorption blueward of N V λ1240, indicating an outflow of $v_{\rm outflow} \approx 1000$–10 000 km s^{-1}. This constrains the AGN lifetime to $t \lesssim d_{\rm spike}/v_{\rm outflow} \approx 3 \times 10^{8-9}$ yr where $d_{\rm spike} \approx 3.1$ pMpc.

The He II photoheating across the He III I-front raises the temperature approximately by (e.g. Kakiichi et al. 2017)

$$\Delta T_{\mathrm{He\,III}} = \frac{2}{3k_{\mathrm{B}}} \frac{G_{\mathrm{He\,II}}/\Gamma_{\mathrm{He\,II}}}{2(X/Y)+3} \approx 6400 \left(\frac{2+\alpha_Q^{\mathrm{eff}}}{2.5}\right)^{-1}, \tag{31}$$

where we adopt a EUV spectral index of $\alpha_Q^{\mathrm{eff}} = 0.5$ to include the effect of spectral hardening $\alpha_Q^{\mathrm{eff}} < \alpha_Q$, resulting in the IGM temperature of $\simeq$16 000 K immediately after the He III I-front (Meiksin, Tittley & Brown 2010; Bolton et al. 2012; Ciardi et al. 2012; Khrykin, Hennawi & McQuinn 2017). Such He II photoheating reduces the optical depth by a factor of $(1.6 \times 10^4\,\mathrm{K}/10^4\,\mathrm{K})^{-0.72} = 0.71$. While a small decrease, now a slightly less underdense gas below $\Delta_b \lesssim 0.30$ can give rise to a transmission spike. This He II heating by AGN doubles the occurrence probability of a transmission spike from $P(<\Delta_b^{\mathrm{th}} = 0.25) = 3.6$ per cent at $T = 10^4$ K to $P(<\Delta_b^{\mathrm{th}} = 0.30) = 8.0$ per cent at $T = 1.6 \times 10^4$ K. The AGN He II photoheating is also a convenient hypothesis as the spatially coherent increase in the Lyα transmission in the He II photoheated region could produce broader transmission spikes (e.g. 220 km s^{-1} corresponds to $\simeq$330 pkpc patch of the IGM) whereas the other spikes widths (100 km s^{-1} are of order the Jeans length of the H I-photoionized IGM.

Hence, the association between the $z \simeq 5.7$ AGN and the proximate broad Lyα transmission spikes suggests that while faint AGN are unlikely a main driver of H I reionization, the hard ionizing spectra of AGN may be important to drive the spatial fluctuations of the ionization and thermal state of the IGM, via possibly an early onset of He II reionization.

7 DISCUSSION AND SUMMARY

We have initiated a spectroscopic programme involving the 3D mapping of $5 < z < 7$ galaxies around the Lyα forest region illuminated by background QSOs which enables us to examine the ionizing capabilities of galaxies and AGN at high redshift. In this paper we describe a science verification of this method using DEIMOS spectroscopy of $5.3 < z < 6.4$ LBGs in the SDSS J1148+5251 field.

Although our sample of confirmed sources is modest, cross-correlation of the spectroscopically confirmed LBGs with the Lyα forest, reveals tentative, but promising, evidence for a 'statistical H I proximity effect' indicating that the Lyα transmission of the IGM is preferentially higher in the vicinity of the galaxies. We have interpreted this signal as evidence for an enhanced UV background around luminous LBGs caused by their ionizing radiation together with that arising from fainter undetected sources clustered around them. We demonstrate that the required ionizing radiation from the luminous LBGs alone is insufficient. This conclusion is supported by independence evidence from deeper imaging observations as well as the expectations of hierarchical clustering in ΛCDM cosmology. This explanation for the statistical H I proximity effect is preferred over alternative hypotheses based solely on gas density or thermal fluctuations of the IGM. Such explanations would produce an anticorrelation yielding an excess Lyα absorption around galaxies. Only UV background fluctuations driven by ionizing radiation from galaxies can predict the H I proximity effect. Balancing the UV background required by the statistical H I proximity effect with the abundance of spectroscopically confirmed LBGs and their fainter associates has enabled us to constrain the *average escape fraction of LyC photons* at $\langle f_{\mathrm{esc}}\rangle \simeq 0.08^{+0.08}_{-0.02}$ with $M_{\mathrm{UV}}^{\mathrm{lim}} \simeq -15 \pm 3$ at $z \simeq 5.8$ using the CLF/HOD framework.

The present method for constraining f_{esc} has some advantages over previous approaches. It examines the direct influence of galaxies on the local IGM as well as the bias of ionizing sources estimated from the galaxy-Lyα cross-correlation; this allows us to deduce the relative contributions of luminous and feeble sources as well as that of AGN. The largest uncertainty at present arises from application to a single QSO sightline and small number statistics. Fortunately, this is easy to remedy with further observations. While a number of assumptions have been made in deriving this value of f_{esc}, we have argued that the uncertainties affecting assumed values for the mean free path and thermal fluctuations in the IGM are likely to increase the derived fraction, strengthening the conclusion that the galaxy population is capable of driving cosmic reionization. Fundamental to our method however, is the assumption that our spectroscopically confirmed sample is unbiased and independent of the surrounding gaseous environment. Since the bulk of our redshifts are based on detecting Lyα emission, if such photons are attenuated by nearby gas this may lower the spectroscopic success rate and may bias the cross-correlation. Such a problem may however be mitigated by examining Lyα haloes as the postulated reduced visibility of Lyα line from galaxies would still produce a bright halo detectable with integral field spectroscopy (Kakiichi & Dijkstra 2017).

As discussed above, the widely held view that the abundant population of intrinsically faint galaxies drives cosmic reionization is supported by this work. This is also consistent with the belief that the typical escape fraction rises at higher redshift as younger, lower mass, galaxies are more susceptible to feedback from intense star-forming activity creating a porous interstellar medium (e.g. Kimm & Cen 2014; Wise et al. 2014). Although there is evidence that reionization may be accelerated around luminous star-forming galaxies (Stark et al. 2017), the statistical H I proximity effect can only be understood if there are intrinsically fainter galaxies clustered around the luminous systems. We note this need not conflict with suggestions that the some of the most luminous systems have harder ionizing radiation (Laporte et al. 2017).

Finally, we explored the specific role of one luminous LBG and a faint AGN where proximate transmission spikes can be directly (as opposed to statistically) associated. A discovery of individual transverse proximity effect via a Lyβ transmission spike in the vicinity of a luminous LBG at $z = 6.177$ suggests that luminous star-forming systems preferentially reside in highly ionized environments. This supports a deduction from the high fraction of Lyα emission in luminous LBGs at $z > 6$ (Curtis-Lake et al. 2012; Stark et al. 2017), for which the visibility of Lyα is boosted by large ionized bubbles (e.g. Dijkstra 2016, for a review). Accelerated reionization around the luminous system likely requires clustered fainter galaxies, whose presence may be indicated by excess O I absorbers (Becker et al. 2006). This scenario may gain further support from an observed galaxy overdensity around a pair of bright Lyα emitting galaxies at $z \sim 7$ (Vanzella et al. 2011; Castellano et al. 2016). The broad Lyα transmission spikes in the vicinity of a $z = 5.701$ faint AGN suggests that the hard ionizing spectra may have an important contribution to the large-scale spatial fluctuations of the UV background and thermal state of the IGM. An interesting possibility is that a patchy early ($z > 5.7$) onset of He II reionization by AGN (Bolton et al. 2012) heats the IGM through He II photoionization heating. This late-time He II heating induces thermal fluctuations so that the intergalactic Lyα opacity is preferentially reduced around luminous systems, without conflicting with the observed statistical H I proximity effect. This may explain the large scatter of intergalactic Lyα opacity at the tail end of reionization (Becker et al. 2015b) without need for a large ($\gtrsim$50 per cent) contribution of AGN to the UV background (Chardin et al. 2015, 2017, 2018; D'Aloisio et al. 2017, see also Finlator et al. 2016) or extreme thermal injection via H II photoheating at the early time of H I reionization (D'Aloisio et al. 2015).

Putting all together, a hypothesis emerging from the initial DEIMOS spectroscopy in the QSO field J1148+5251 is that while the faint galaxies with high escape fraction primarily drive reionization, luminous galaxies and AGN may play an increasingly important role towards the end of the reionization process by sourcing the large-scale spatial fluctuations of the UV background and thermal state of the IGM. This demonstrates the potential of spectroscopic survey of $5 < z < 7$ galaxies towards QSO fields for making a progress with existing facilities before the *JWST* and Extremely Large Telescopes, allowing us to tackle the most challenging aspect of cosmic reionization: *'What reionized the Universe?'*.

ACKNOWLEDGEMENTS

We acknowledge financial support from European Research Council Advanced Grant FP7/669253 (KK, NL, RSE, RAM, SEIB). DPS acknowledges support from the National Science Foundation through grant AST-1410155. ERW acknowledges support from the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. It is a pleasure to thank the following for useful discussion: Will Hartley, Andreu Font-Ribera, Harley Katz, Tom Fletcher, Jamie Bolton, Masami Ouchi, and Akio Inoue. We acknowledge useful email correspondence with Fred Davies and Ali Rahmati regarding their simulations. We thank the referee for useful comments and, in particular, on suggestions for improving the analysis with larger data sets. The data presented herein were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California, and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this beautiful mountain. This work was undertaken using the UCL Grace High Performance Computing Facility (Grace@UCL) and we thank the associated support services. BR acknowledges partial support through NASA contract NNG16PJ25C, and grants 17-ATP17-0034 and *HST*-GO-14747.

REFERENCES

Adelberger K. L., Shapley A. E., Steidel C. C., Pettini M., Erb D. K., Reddy N. A., 2005, ApJ, 629, 636
Adelberger K. L., Steidel C. C., Shapley A. E., Pettini M., 2003, ApJ, 584, 45
Atek H. et al., 2015, ApJ, 814, 69
Barone-Nugent R. L. et al., 2014, ApJ, 793, 17
Becker G. D., Bolton J. S., 2013, MNRAS, 436, 1023
Becker G. D., Bolton J. S., Lidz A., 2015a, PASA, 32, e045
Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015b, MNRAS, 447, 3402
Becker G. D., Hewett P. C., Worseck G., Prochaska J. X., 2013, MNRAS, 430, 2067
Becker G. D., Sargent W. L. W., Rauch M., Simcoe R. A., 2006, ApJ, 640, 69
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bielby R. M. et al., 2017, MNRAS, 471, 2174
Bolton J. S., Becker G. D., Raskutti S., Wyithe J. S. B., Haehnelt M. G., Sargent W. L. W., 2012, MNRAS, 419, 2880
Bolton J. S., Haehnelt M. G., 2007, MNRAS, 382, 325
Bosman S. E. I., Fan X., Jiang L., Reed S. L., Matsuoka Y., Becker G. D., Haehnelt M. G., 2018, MNRAS, preprint (arXiv:1802.08177)
Bouwens R. J., Illingworth G. D., Blakeslee J. P., Franx M., 2006, ApJ, 653, 53
Bouwens R. J., Illingworth G. D., Franx M., Ford H., 2007, ApJ, 670, 928
Bouwens R. J., Oesch P. A., Illingworth G. D., Ellis R. S., Stefanon M., 2017, ApJ, 843, 129
Bouwens R. J., Smit R., Labbé I., Franx M., Caruana J., Oesch P., Stefanon M., Rasappu N., 2016, ApJ, 831, 176
Bouwens R. J. et al., 2015, ApJ, 803, 34
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Cai Z., Fan X., Dave R., Finlator K., Oppenheimer B., 2017, ApJ, 849, L18
Carilli C. L. et al., 2010, ApJ, 714, 834
Castellano M. et al., 2016, ApJ, 818, L3
Chardin J., Haehnelt M. G., Aubert D., Puchwein E., 2015, MNRAS, 453, 2943
Chardin J., Haehnelt M. G., Bosman S. E. I., Puchwein E., 2018, MNRAS, 473, 765
Chardin J., Puchwein E., Haehnelt M. G., 2017, MNRAS, 465, 3429
Chen H.-W., Prochaska J. X., Gnedin N. Y., 2007, ApJ, 667, L125
Ciardi B., Bolton J. S., Maselli A., Graziani L., 2012, MNRAS, 423, 558
Cooper M. C., Newman J. A., Davis M., Finkbeiner D. P., Gerke B. F., 2012, Astrophysics Source Code Library, record ascl:1203.003
Cooray A., 2006, MNRAS, 365, 842
Crighton N. H. M. et al., 2011, MNRAS, 414, 28
Croft R. A. C. et al., 2016, MNRAS, 457, 3541
Curtis-Lake E. et al., 2012, MNRAS, 422, 1425
Curtis-Lake E. et al., 2013, MNRAS, 429, 302
Danforth C. W. et al., 2016, ApJ, 817, 111
Davies F. B., Becker G. D., Furlanetto S. R., 2017, preprint (arXiv:1708.08927)
Dijkstra M., 2016, in Mesinger A., ed., Astrophysics and Space Science Library, Vol. 423, Understanding the Epoch of Cosmic Reionization. Springer, Berlin. p. 145
Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., 2008, MNRAS, 390, L64
du Mas des Bourboux H. et al., 2017, A&A, 608, A130
Díaz C. G., Koyama Y., Ryan-Weber E. V., Cooke J., Ouchi M., Shimasaku K., Nakata F., 2014, MNRAS, 442, 946
Díaz C. G., Ryan-Weber E. V., Cooke J., Koyama Y., Ouchi M., 2015, MNRAS, 448, 1240
Díaz C. G., Ryan-Weber E. V., Cooke J., Pettini M., Madau P., 2011, MNRAS, 418, 820
D'Aloisio A., McQuinn M., Trac H., 2015, ApJ, 813, L38
D'Aloisio A., Upton Sanderbeck P. R., McQuinn M., Trac H., Shapiro P. R., 2017, MNRAS, 468, 4691
Eilers A.-C., Davies F. B., Hennawi J. F., Prochaska J. X., Lukić Z., Mazzucchelli C., 2017, ApJ, 840, 24
Eldridge J. J., Stanway E. R., Xiao L. , McClelland L. A. S. , Taylor G. , Ng M. , Greis S. M. L., Bray J. C. , 2017, PASA, 34, e058
Faber S. M. et al., 2003, in Iye M., Moorwood A. F. M., eds, Proc. SPIE Conf. Ser. Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes. SPIE, Bellingham. p. 1657
Faisst A. L., 2016, ApJ, 829, 99
Fan X. et al., 2003, AJ, 125, 1649
Fan X. et al., 2006, AJ, 132, 117
Faucher-Giguère C.-A., Lidz A., Hernquist L., Zaldarriaga M., 2008, ApJ, 688, 85
Finlator K., Muñoz J. A., Oppenheimer B. D., Oh S. P., Özel F., Davé R., 2013, MNRAS, 436, 1818
Finlator K., Oppenheimer B. D., Davé R., Zackrisson E., Thompson R., Huang S., 2016, MNRAS, 459, 2299
Font-Ribera A. et al., 2013, J. Cosmol. Astropart. Phys., 5, 018

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Fynbo J. P. U. et al., 2009, ApJS, 185, 526
Gallerani S., Ferrara A., Fan X., Choudhury T. R., 2008, MNRAS, 386, 359
Gallerani S. et al., 2017, MNRAS, 467, 3590
García L. A., Tescari E., Ryan-Weber E. V., Wyithe J. S. B., 2017, MNRAS, 469, L53
Giallongo E. et al., 2015, A&A, 578, A83
Gnedin N. Y., Fan X., 2006, ApJ, 648, 1
Gnedin N. Y., Ostriker J. P., 1997, ApJ, 486, 581
Grazian A. et al., 2016, A&A, 585, A48
Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633
Gunn J. E., Stryker L. L., 1983, ApJS, 52, 121
Haardt F., Madau P., 2012, ApJ, 746, 125
Harikane Y. et al., 2016, ApJ, 821, 123
Harikane Y. et al., 2018a, ApJ, 859, 84
Harikane Y. et al., 2018b, PASJ, 70, S11
Harrison C. M., Meiksin A., Stock D., 2011, preprint (arXiv:1105.6208)
Hassan S., Davé R., Mitra S., Finlator K., Ciardi B., Santos M. G., 2018, MNRAS, 473, 227
Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27
Inoue A. K., Iwata I., Deharveng J.-M., 2006, MNRAS, 371, L1
Ishigaki M., Kawamata R., Ouchi M., Oguri M., Shimasaku K., Ono Y., 2018, ApJ, 854, 73
Jiang L. et al., 2006, AJ, 132, 2127
Jones T. A., Ellis R. S., Schenker M. A., Stark D. P., 2013, ApJ, 779, 52
Kakiichi K., Dijkstra M., 2017, preprint (arXiv:1710.10053)
Kakiichi K., Graziani L., Ciardi B., Meiksin A., Compostella M., Eide M. B., Zaroubi S., 2017, MNRAS, 468, 3718
Kakiichi K., Meiksin A., Tittley E., 2012, MNRAS, 426, 2129
Keating L. C., Puchwein E., Haehnelt M. G., 2018, MNRAS, 477, 5501
Khrykin I. S., Hennawi J. F., McQuinn M., 2017, ApJ, 838, 96
Kimm T., Cen R., 2014, ApJ, 788, 121
Kuhlen M., Faucher-Giguère C.-A., 2012, MNRAS, 423, 862
Laporte N., Nakajima K., Ellis R. S., Zitrin A., Stark D. P., Mainali R., Roberts-Borsani G. W., 2017, ApJ, 851, 40
Leethochawalit N., Jones T. A., Ellis R. S., Stark D. P., Zitrin A., 2016, ApJ, 831, 152
Le Fèvre O. et al., 2015, A&A, 576, A79
Livermore R. C., Finkelstein S. L., Lotz J. M., 2017, ApJ, 835, 113
Madau P., Haardt F., 2015, ApJ, 813, L8
Madau P., Pozzetti L., Dickinson M., 1998, ApJ, 498, 106
Mahabal A., Stern D., Bogosavljević M., Djorgovski S. G., Thompson D., 2005, ApJ, 634, L9
Mainali R., Kollmeier J. A., Stark D. P., Simcoe R. A., Walth G., Newman A. B., Miller D. R., 2017, ApJ, 836, L14
Malhotra S. et al., 2005, ApJ, 626, 666
Mas-Ribas L., Dijkstra M., 2016, ApJ, 822, 84
Matthee J., Sobral D., Best P., Khostovan A. A., Oteo I., Bouwens R., Röttgering H., 2017, MNRAS, 465, 3637
McLure R. J., Cirasuolo M., Dunlop J. S., Foucaud S., Almaini O., 2009, MNRAS, 395, 2196
McQuinn M., Upton Sanderbeck P. R., 2016, MNRAS, 456, 47
Meiksin A., 2006a, MNRAS, 365, 807
Meiksin A., 2006b, MNRAS, 365, 833
Meiksin A., Bolton J. S., Puchwein E., 2017, MNRAS, 468, 1893
Meiksin A., Bolton J. S., Tittley E. R., 2015, MNRAS, 453, 899
Meiksin A., Tittley E. R., Brown C. K., 2010, MNRAS, 401, 77
Meiksin A., White M., 2003, MNRAS, 342, 1205
Meiksin A. A., 2009, Rev. Mod. Phys., 81, 1405
Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, ApJ, 530, 1
Mitra S., Choudhury T. R., Ferrara A., 2018, MNRAS, 473, 1416
Morselli L. et al., 2014, A&A, 568, A1
Mostardi R. E., Shapley A. E., Nestor D. B., Steidel C. C., Reddy N. A., Trainor R. F., 2013, ApJ, 779, 65
Mostardi R. E., Shapley A. E., Steidel C. C., Trainor R. F., Reddy N. A., Siana B., 2015, ApJ, 810, 107
Naidu R. P. et al., 2017, ApJ, 847, 12
Newman J. A. et al., 2013, ApJS, 208, 5
Oesch P. A. et al., 2015, ApJ, 804, L30
Oh S. P., Furlanetto S. R., 2005, ApJ, 620, L9
Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
Ono Y. et al., 2018, PASJ, 70, S10
Osterbrock D. E., Fulbright J. P., Martel A. R., Keane M. J., Trager S. C., Basri G., 1996, PASP, 108, 277
Paresce F., McKee C. F., Bowyer S., 1980, ApJ, 240, 387
Parsa S., Dunlop J. S., McLure R. J., 2018, MNRAS, 474, 2904
Pawlik A. H., Schaye J., van Scherpenzeel E., 2009, MNRAS, 394, 1812
Peacock J. A., Dodds S. J., 1996, MNRAS, 280, L19
Planck Collaboration XIII, 2016, A&A, 594, A13
Prochaska J. X. et al., 2013, ApJ, 776, 136
Puchwein E., Haardt F., Haehnelt M. G., Madau P., 2018, preprint (arXiv:1801.04931)
Qin Y. et al., 2017, MNRAS, 472, 2009
Rahmati A., Schaye J., 2018, MNRAS, preprint (arXiv:1708.04238)
Rahmati A., Schaye J., Bower R. G., Crain R. A., Furlong M., Schaller M., Theuns T., 2015, MNRAS, 452, 2034
Rakic O., Schaye J., Steidel C. C., Rudie G. C., 2012, ApJ, 751, 94
Reddy N. A., Steidel C. C., Pettini M., Bogosavljević M., Shapley A. E., 2016, ApJ, 828, 108
Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, ApJ, 802, L19
Robertson B. E. et al., 2013, ApJ, 768, 71
Rudie G. C., Steidel C. C., Shapley A. E., Pettini M., 2013, ApJ, 769, 146
Rudie G. C. et al., 2012, ApJ, 750, 67
Schirber M., Bullock J. S., 2003, ApJ, 584, 110
Schmidt T. M., Worseck G., Hennawi J. F., Prochaska J. X., Crighton N. H. M., 2017, ApJ, 847, 81
Shivaei I. et al., 2018, ApJ, 855, 42
Slosar A. et al., 2011, J. Cosmol. Astropart. Phys., 9, 001
Sorini D., Oñorbe J., Hennawi J. F., Lukić Z., 2018, ApJ, 859, 125
Springel V. et al., 2005, Nature, 435, 629
Stanway E. R., Eldridge J. J., Becker G. D., 2016, MNRAS, 456, 485
Stark D. P., 2016, ARA&A, 54, 761
Stark D. P., Ellis R. S., Chiu K., Ouchi M., Bunker A., 2010, MNRAS, 408, 1628
Stark D. P. et al., 2015, MNRAS, 454, 1393
Stark D. P. et al., 2017, MNRAS, 464, 469
Steidel C. C., Erb D. K., Shapley A. E., Pettini M., Reddy N., Bogosavljević M., Rudie G. C., Rakic O., 2010, ApJ, 717, 289
Tasca L. A. M. et al., 2017, A&A, 600, A110
Telfer R. C., Zheng W., Kriss G. A., Davidsen A. F., 2002, ApJ, 565, 773
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878
Tittley E. R., Meiksin A., 2007, MNRAS, 380, 1369
Trac H., Cen R., Loeb A., 2008, ApJ, 689, L81
Tummuangpak P., Bielby R. M., Shanks T., Theuns T., Crighton N. H. M., Francke H., Infante L., 2014, MNRAS, 442, 2094
Turner M. L., Schaye J., Crain R. A., Rudie G., Steidel C. C., Strom A., Theuns T., 2017, MNRAS, 471, 690
Turner M. L., Schaye J., Steidel C. C., Rudie G. C., Strom A. L., 2014, MNRAS, 445, 794
van den Bosch F. C., More S., Cacciato M., Mo H., Yang X., 2013, MNRAS, 430, 725
van Kampen N.G., 2007, Stochastic Processes in Physics and Chemistry, 3rd edn. Elsevier, Amsterdam
Vanzella E. et al., 2009, ApJ, 695, 1163
Vanzella E. et al., 2010, ApJ, 725, 1011
Vanzella E. et al., 2011, ApJ, 730, L35
Vasei K. et al., 2016, ApJ, 831, 38
Wall J. V., Jenkins C. R., 2012, Practical Statistics for Astronomers. Cambridge Univ. Press, Cambridge
White R. L., Becker R. H., Fan X., Strauss M. A., 2003, AJ, 126, 1
White R. L., Becker R. H., Fan X., Strauss M. A., 2005, AJ, 129, 2102

Willott C. J. et al., 2010, AJ, 140, 546
Wise J. H., Demchenko V. G., Halicek M. T., Norman M. L., Turk M. J., Abel T., Smith B. D., 2014, MNRAS, 442, 2560
Worseck G. et al., 2014, MNRAS, 445, 1745
Wyithe J. S. B., Bolton J. S., 2011, MNRAS, 412, 1926
Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057
Yang X., Mo H. J., van den Bosch F. C., 2008, ApJ, 676, 248
Zackrisson E., Inoue A. K., Jensen H., 2013, ApJ, 777, 39
Zackrisson E. et al., 2017, ApJ, 836, 78
Zitrin A. et al., 2015, ApJ, 810, L12
Zuo L., 1992a, MNRAS, 258, 36
Zuo L., 1992b, MNRAS, 258, 45

APPENDIX A: THEORETICAL FRAMEWORK

A1 Cosmological radiative transfer

Here, we present a more complete treatment of cosmological radiative transfer of ionizing photons. The equation of cosmological radiative transfer follows (e.g. Gnedin & Ostriker 1997; Meiksin 2009)

$$\frac{1}{c}\frac{\partial I_\nu}{\partial t} + \boldsymbol{n}\cdot\nabla I_\nu - \frac{H}{c}\left(\nu\frac{\partial I_\nu}{\partial \nu} - 3HI_\nu\right) = -\alpha_\nu I_\nu + \varepsilon_\nu, \tag{A1}$$

where α_ν is the absorption coefficient and ϵ_ν is the emissivity. The direct solution to this clearly requires expensive numerical radiative transfer simulations. Instead, we seek an approximate statistical solution following the approach of Zuo (1992a,b), Meiksin & White (2003), Kakiichi, Meiksin & Tittley (2012). First, consider a small patch of the universe at position $\boldsymbol{r}$ with volume V so that the cosmological redshifting can be ignored, which should have a minor impact at $z > 5$ (Becker & Bolton 2013). The number of galaxies above a luminosity $L_{\rm min}$ in the patch follows the Poisson distribution $P(N) = \bar{N}^N \mathrm{e}^{-\bar{N}}/N!$ with the mean value $\bar{N}$.

As we are interested in the radiation field around a spectroscopically detected LBG, we split the specific intensity into the contribution from the central LBG $J_0(\boldsymbol{r})$ and surrounding galaxies $J_\nu(\boldsymbol{r})$. Then, in a patch with N galaxies around a LBG with specific luminosity $L_{\rm LBG}$, integrating equation (A1) we find that the specific intensity at a distance $\boldsymbol{r}$ from the LBG is given by

$$I_\nu(\boldsymbol{r}) = J_0(\boldsymbol{r}) + J_\nu(\boldsymbol{r}), \tag{A2}$$

where

$$J_0(\boldsymbol{r}) = \frac{L_{\rm LBG}\mathrm{e}^{-\tau_c(|\boldsymbol{r}|)}}{(4\pi)^2|\boldsymbol{r}|^2}, \quad J_\nu(\boldsymbol{r}) = \sum_{k=1}^{N}\frac{L_k \mathrm{e}^{-\tau_c(|\boldsymbol{r}-\boldsymbol{r}_k|)}}{(4\pi)^2|\boldsymbol{r}-\boldsymbol{r}_k|^2}, \tag{A3}$$

and L_k is the specific luminosity and $\boldsymbol{r}_k$ is the proper distance of kth galaxy from the LBG, and τ_c is the optical depth of ionizing photons over a distance $|\boldsymbol{r}-\boldsymbol{r}_k|$.

To derive the statistically averaged specific intensity around LBGs, we take the ensemble-averaging over many realizations of patches with various numbers of galaxies. Using the statistical method of characteristic functions (Meiksin & White 2003; Kakiichi et al. 2012) or otherwise (Zuo 1992a,b), this gives the average specific intensity,

$$\langle I_\nu(\boldsymbol{r})\rangle = \langle J_0(\boldsymbol{r})\rangle + \sum_{N=0}^{\infty} P(N)\int J_\nu(\boldsymbol{r})P[J_\nu(\boldsymbol{r})|N]\mathrm{d}J_\nu(\boldsymbol{r}), \tag{A4}$$

where $P[J_\nu(\boldsymbol{r})|N]$ is the probability distribution function of specific intensity in a patch with N galaxies. When the positions and luminosities of surrounding galaxies are statistically independent to each other (but can be correlated with the LBG) (e.g. van Kampen 2007), we may express $P[J_\nu(\boldsymbol{r})|N]$ as a product of the probabilities of finding each galaxy at a position $\boldsymbol{r}_k$ with a luminosity L_k,

$$P[J_\nu(\boldsymbol{r})|N]\mathrm{d}J_\nu(\boldsymbol{r}) = \prod_{k=1}^{N}\frac{\Phi(L_k)\mathrm{d}L_k}{\bar{n}_g(>L_{\rm min})}\left[1+\xi_g(\boldsymbol{r}_k, L_k)\right]\frac{\mathrm{d}^3 r_k}{V}, \tag{A5}$$

where $\bar{n}_g(>L_{\rm min}) = \int_{L_{\rm min}}^{\infty}\Phi(L)\mathrm{d}L$ and $\xi_g(\boldsymbol{r}, L)$ is the correlation function of LBGs with galaxies of luminosity L. Therefore, by substituting equations (A3) and (A5) into (A4) we obtain, after some algebra,

$$\langle I_\nu(\boldsymbol{r})\rangle = \langle J_0(\boldsymbol{r})\rangle + \bar{\varepsilon}_\nu\int\frac{\langle \mathrm{e}^{-\tau_c(|\boldsymbol{r}-\boldsymbol{r}'|)}\rangle}{(4\pi)^2|\boldsymbol{r}-\boldsymbol{r}'|^2}\left[1+\langle\xi_g(\boldsymbol{r}')\rangle_L\right]\mathrm{d}^3r', \tag{A6}$$

where the local contribution from the LBG is given by $\langle J_0(\boldsymbol{r})\rangle = \langle L_{\rm LBG}\rangle\langle \mathrm{e}^{-\tau_c(|\boldsymbol{r}|)}\rangle/(4\pi|\boldsymbol{r}|)^2$ and $\bar{\varepsilon}_\nu = \int_{L_{\rm min}}^{\infty}L\Phi(L)\mathrm{d}L$ is the mean emissivity of galaxies. Expressions such as equation (A6) are intuitive and quoted elsewhere in literature (e.g. Mas-Ribas & Dijkstra 2016). The above statistical derivation elucidates various assumptions and clarifies that the average specific intensity depends on the *luminosity-weighted correlation function* defined as

$$\langle\xi_g(\boldsymbol{r})\rangle_L = \frac{\int_{L_{\rm min}}^{\infty}L\Phi(L)\xi_g(\boldsymbol{r}, L)\mathrm{d}L}{\int_{L_{\rm min}}^{\infty}L\Phi(L)\mathrm{d}L}. \tag{A7}$$

In the derivation, because we are dealing with the propagation of photons in an ensemble-averaged sense, we have replaced the attenuation along an individual sightline $\mathrm{e}^{-\tau_c(|\boldsymbol{r}-\boldsymbol{r}_k|)}$ with the average value $\langle \mathrm{e}^{-\tau_c(|\boldsymbol{r}-\boldsymbol{r}_k|)}\rangle$ and approximated as

$$\langle \mathrm{e}^{-\tau_c(|\boldsymbol{r}-\boldsymbol{r}'|)}\rangle \approx \mathrm{e}^{-|\boldsymbol{r}-\boldsymbol{r}'|/\lambda_{\rm mfp}}. \tag{A8}$$

The mean free path of ionizing photons is given by

$$\lambda_{\rm mfp}^{-1} = \left|\frac{\mathrm{d}z}{\mathrm{d}l_p}\right|\int \mathrm{d}N_{\rm H\,I}\frac{\partial^2\mathcal{N}}{\partial N_{\rm H\,I}\partial z}\left[1-\mathrm{e}^{-\sigma_{\rm H\,I}(\nu)N_{\rm H\,I}}\right], \tag{A9}$$

for Poisson-distributed absorbers (Paresce, McKee & Bowyer 1980; Haardt & Madau 2012). The mean free path depends on the number density of H I absorbers, which is quantified by the H I column density distribution function (CDDF) $\frac{\partial^2\mathcal{N}}{\partial N_{\rm H\,I}\partial z}$. Parametrizing the CDDF as a power-law $\frac{\partial^2\mathcal{N}}{\partial N_{\rm H\,I}\partial z}\propto N_{\rm H\,I}^{-\beta_{\rm N}}$ (e.g. $\beta_N = 1.33 \pm 0.05$, Becker & Bolton 2013), it may be written as $\lambda_{\rm mfp} = \lambda_{912}(\nu/\nu_{912})^{3(\beta_N-1)}$ (Faucher-Giguère et al. 2008) where λ_{912} is the mean free path of Lyman-limit photons. In this paper, we assume a constant mean free path at Lyman-limit $\lambda_{\rm mfp}\approx\lambda_{912}$. This produces a systematic error, underestimating the mean photoionization rate by a small factor of $[3+\alpha_g-3(\beta_{\rm N}-1)]/(3+\alpha_g)\approx 0.84$ when $\alpha_g = 3$ and $\beta_N = 1.33$ because ignoring the effect that higher frequency photons can reach longer distance before being attenuated. Although adopting a constant spatially uniform mean free path is clearly an oversimplification, it gives a first-order approximation to the mean free path. To encapsulate this model uncertainty (see Section 4.2), we use the Gaussian prior on $\lambda_{\rm mfp}$ with variance of 2 pMpc.

Furthermore, equation (A6) can be written more succinctly in Fourier space. By realizing that equation (A6) is the convolution between the radiative transfer kernel $\frac{\mathrm{e}^{-|\boldsymbol{r}-\boldsymbol{r}'|/\lambda_{\rm mfp}}}{(4\pi\lambda_{\rm mfp})|\boldsymbol{r}-\boldsymbol{r}'|^2}$ and the luminosity-dependent correlation function $\langle\xi_g(\boldsymbol{r})\rangle_L$, we arrive at

$$\langle I_\nu(r)\rangle = \langle J_0(r)\rangle + \frac{\bar{\varepsilon}_\nu\lambda_{\rm mfp}}{4\pi}\times\left[1+\int_0^\infty\frac{k^2\mathrm{d}k}{2\pi^2}R(k\lambda_{\rm mfp})\langle P_g(k)\rangle_L\frac{\sin kr}{kr}\right], \tag{A10}$$

where $R(x) = \arctan(x)/x$ comes from the Fourier transform of the radiative transfer kernel and the luminosity-dependent galaxy power spectrum is

$$\langle P_g(k)\rangle_L = \int_0^\infty \langle \xi_g(r)\rangle_L 4\pi r^2 \frac{\sin kr}{kr} \mathrm{d}r. \quad \text{(A11)}$$

The expression reduces to the local approximation of the Poisson-distributed sources $\langle J_\nu(r)\rangle = \bar{\varepsilon}_\nu \lambda_{\rm mfp}/(4\pi)$ (e.g. Schirber & Bullock 2003) when there is no galaxy clustering around LBGs, $\langle P_g(k)\rangle_L = 0$.

Finally, we suppose that all galaxies have the same spectral energy distribution with the EUV (>13.6 eV) slope α_g to evaluate a typical photoionization rate at a distance r from a LBG, in which the EUV emissivity from star-forming galaxies is

$$\bar{\varepsilon}_\nu = h\alpha_g \left(\frac{\nu}{\nu_{912}}\right)^{-\alpha_g} \dot{n}_{\rm ion}(> L_{\rm min}). \quad \text{(A12)}$$

Hence, using the approximate statistical solution (A10) of the radiation field, we obtain the typical photoionization rate at a distance r from a LBG:

$$\langle \Gamma_{\rm H\,I}(r)\rangle = \int_{\nu_{\rm H\,I}}^\infty \sigma_{\rm H\,I}(\nu) \frac{4\pi \langle I_\nu(r)\rangle}{h\nu} \mathrm{d}\nu,$$
$$= \langle \Gamma^{\rm LBG}_{\rm H\,I}(r)\rangle + \bar{\Gamma}_{\rm H\,I}\left[1 + \int_0^\infty \frac{k^2 \mathrm{d}k}{2\pi^2} R(k\lambda_{\rm mfp}) \langle P_g(k)\rangle_L \frac{\sin kr}{kr}\right], \quad \text{(A13)}$$

where the first term $\langle \Gamma^{\rm LBG}_{\rm H\,I}(r)\rangle = \frac{\alpha_g \sigma_{912}}{\alpha_g + 3} \frac{\langle \dot{N}^{\rm LBG}_{\rm ion}\rangle}{4\pi r^2} \mathrm{e}^{-r/\lambda_{\rm mfp}}$ is the local contribution from the central LBGs and the second term is the clustering contribution from the surrounding galaxies. We use equation (A13) throughout the analysis presented in this paper.

A2 Galaxy abundance from HOD framework

We use the HOD framework to estimate the number of fainter, undetected, galaxies clustered around LBGs. We use the conditional luminosity function (CLF) approach (e.g. Yang et al. 2003) to the halo occupation distribution (HOD) framework. The CLF, $\Phi(L|M_h)$, specifies the average number of galaxies with luminosities in the range of $L \pm \mathrm{d}L/2$ that reside in a halo of mass M_h. Thus, by combining the best-fitting CLF with the theoretical estimate of the clustering of dark matter haloes around the LBG-host haloes from N-body simulations (e.g. Tinker et al. 2008, 2010), we can infer the number of (undetected) galaxies around the (observed) LBGs. To this end, we constrain the CLF model by simultaneously fitting to the observed UV luminosity function (Bouwens et al. 2015) and angular correlation function of $z \sim 6$ LBGs (Harikane et al. 2016). We follow the CLF model of van den Bosch et al. (2013), which is summarized below. We drop the subscript $_{\rm UV}$ of $L_{\rm UV}$ for notational clarity.

We split the CLF from the contribution from central galaxies, $\Phi_{\rm cen}(L|M_h)$, and satellite galaxies, $\Phi_{\rm sat}(L|M_h)$:

$$\Phi(L|M_h) = \Phi_{\rm cen}(L|M_h) + \Phi_{\rm sat}(L|M_h). \quad \text{(A14)}$$

We model the CLF of central galaxies model as a log-normal distribution,

$$\Phi_{\rm cen}(L|M_h)\mathrm{d}L = \frac{\log_{10} e}{\sqrt{2\pi}\sigma_{\rm c}} \exp\left[-\frac{(\log_{10} L - \log_{10} L_{\rm c})^2}{2\sigma_{\rm c}^2}\right] \frac{\mathrm{d}L}{L}, \quad \text{(A15)}$$

where σ_c quantifies the scatter in UV luminosity of central galaxies and halo mass and we adopt a following parametrization for the central UV luminosity–halo mass relation,

$$L_{\rm c}(M_h) = L_0 \frac{(M_h/M_h^*)^{\gamma_1}}{[1 + (M_h/M_h^*)]^{\gamma_1 - \gamma_2}}, \quad \text{(A16)}$$

where L_0 is the normalization, M_h^* is a characteristic halo mass, γ_1 and γ_2 are the power-law slope at low-mass ($M_h \ll M_h^*$) and high-mass ($M_h \gg M_h^*$) ends, respectively. The CLF for satellite galaxies is modelled as a modified Schechter function,

$$\Phi_{\rm sat}(L|M_h)\mathrm{d}L = \phi_s^* \left(\frac{L}{L_s^*}\right)^{\alpha_s + 1} \exp\left[-\left(\frac{L}{L_s^*}\right)^2\right] \frac{\mathrm{d}L}{L}, \quad \text{(A17)}$$

where $L_s^*(M_h) = 0.562 L_{\rm c}(M_h)$ (Yang, Mo & van den Bosch 2008) and

$$\phi_s^*(M_h) = \phi_0 \left(\frac{M_h}{10^{12} h^{-1} \mathrm{M}_\odot}\right)^{\beta_s}. \quad \text{(A18)}$$

Therefore, the CLF model contains eight free parameters, $\boldsymbol{\theta}_{\rm CLF} = (\log L_0, \log M_h^*, \gamma_1, \gamma_2, \sigma_c, \log \phi_0, \alpha_s, \beta_s)$ (strictly speaking, we express $\log L_0$ in terms of the corresponding UV magnitude $M_{\rm UV,\,0}$).

Once the CLF is specified, we can compute the luminosity function and the correlation function (power spectrum) of galaxies. The luminosity function is given by

$$\Phi(L) = \int \Phi(L|M_h) \frac{\mathrm{d}n}{\mathrm{d}M_h} \mathrm{d}M_h, \quad \text{(A19)}$$

where $\mathrm{d}n/\mathrm{d}M_h$ is the halo mass function for which we use Tinker et al. (2008) halo mass function.

The galaxy power spectrum is computed using the standard HOD framework. Using the CLF, the halo occupation number of central galaxies above a limiting luminosity threshold of sample $L_{\rm th}$ is given by,

$$\langle N_{\rm cen}|M_h\rangle = \int_{L_{\rm th}}^\infty \Phi_{\rm cen}(L|M_h)\mathrm{d}L, \quad \text{(A20)}$$

and for satellite galaxies, $\langle N_{\rm sat}|M_h\rangle = \int_{L_{\rm th}}^\infty \Phi_{\rm sat}(L|M_h)\mathrm{d}L$. The number density of galaxies is $\bar{n}_g(> L_{\rm th}) = \int \langle N|M_h\rangle \frac{dn}{dM_h} dM_h$ where $\langle N|M_h\rangle = \langle N_{\rm cen}|M_h\rangle + \langle N_{\rm sat}|M_h\rangle$ is the total halo occupation number of galaxies. In the halo model (e.g. Cooray 2006), the power spectrum of galaxies is expressed in terms of one-halo and two-halo terms containing all possible combinations of central and satellites,

$$P_g(k) = 2P^{1h}_{\rm cs}(k) + P^{1h}_{\rm ss}(k)$$
$$+ P^{2h}_{\rm cc}(k) + 2P^{2h}_{\rm cs}(k) + P^{2h}_{\rm ss}(k). \quad \text{(A21)}$$

Following the notation of van den Bosch et al. (2013), we have defined the necessary one-halo $P^{1h}_{\rm xy}(k)$ and two-halo terms $P^{2h}_{\rm xy}(k)$ as

$$P^{1h}_{\rm xy}(k) = \int \mathrm{d}M_h \mathcal{H}_{\rm x}(k, M_h) \mathcal{H}_{\rm y}(k, M_h) \frac{\mathrm{d}n}{\mathrm{d}M_h}, \quad \text{(A22)}$$

and

$$P^{2h}_{\rm xy}(k) = P_m(k) \int \mathrm{d}M_h \mathcal{H}_{\rm x}(k|M_h) b_h(M_h) \frac{\mathrm{d}n}{\mathrm{d}M_h}$$
$$\times \int \mathrm{d}M_h' \mathcal{H}_{\rm y}(k|M_h') b_h(M_h') \frac{\mathrm{d}n}{\mathrm{d}M_h'} \quad \text{(A23)}$$

where 'x' and 'y' are either 'c' (for central) or 's' (for satellite), and

$$\mathcal{H}_c(k, M_h) = \frac{\langle N_{\rm cen}|M_h\rangle}{\bar{n}_g(> L_{\rm th})}, \quad \mathcal{H}_s(k, M_h) = \frac{\langle N_{\rm sat}|M_h\rangle}{\bar{n}_g(> L_{\rm th})} \tilde{u}(k|M_h). \quad \text{(A24)}$$

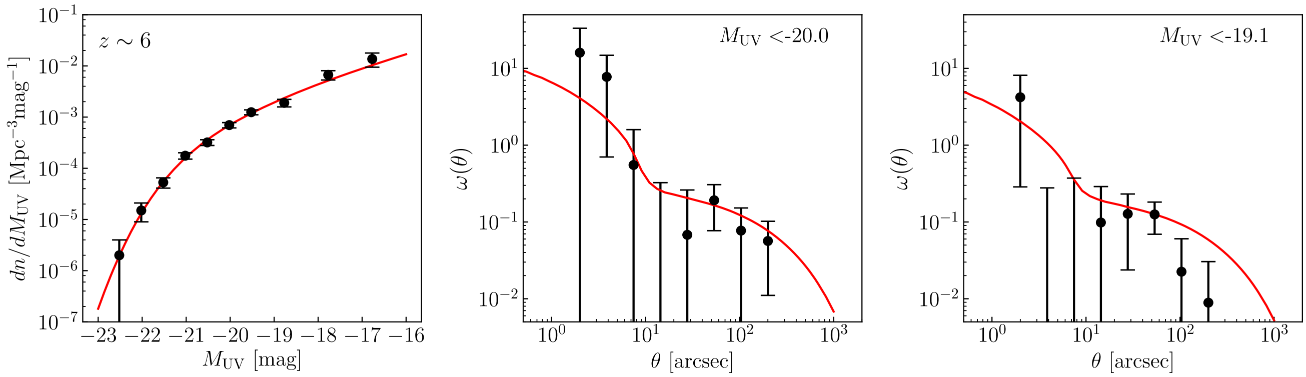

Figure A1. Comparison of the best-fitting CLF model with the $z \sim 6$ UV luminosity function of Bouwens et al (2015) and LBG angular correlation functions of Harikane et al. (2016) (red: model, black: observations).

Table A1. The best-fitting CLF parameters.

Parameter	Best-fitting value
$M_{\rm UV,0}$	$-21.43^{+1.36}_{-0.96}$
$\log M_h^*$	$11.56^{+0.28}_{-0.43}$
γ_1	$2.10^{+0.55}_{-0.25}$
γ_2	$0.25^{+0.36}_{-0.48}$
σ_c	0.2 (fixed)
$\log\phi_0$	$-0.94^{+0.08}_{-0.04}$
α_s	$-1.15^{+0.11}_{-0.18}$
β_s	$1.11^{+0.45}_{-0.53}$

For the dark matter power spectrum $P_m(k)$, we use the non-linear fitting formula of Peacock & Dodds (1996). The result is marginally affected even if we use the linear matter power spectrum. For the halo bias factor $b_h(M_h)$, we adopt the fitting function of Tinker et al. (2010). Here, $\tilde{u}(k|M_h)$ is the Fourier transform of the NFW halo profile and for the halo concentration parameter we use Duffy et al. (2008) fitting function.

Finally, we compute the angular correlation function of galaxies from the galaxy power spectrum. Using the Limber approximation, the angular correlation function at a perpendicular separation $r_\perp$ is given by

$$\omega_{ij}(r_\perp) = \int {\rm d}z N^2(z) \left|\frac{{\rm d}\chi}{{\rm d}z}\right|^{-1} \int \frac{{\rm d}k}{2\pi} k P_{ij}(k) J_0(kr_\perp), \tag{A25}$$

where $N(z)$ is the normalized redshift distribution of galaxies, $|{\rm d}\chi/{\rm d}z| = c/H(z)$, and $J_0(kr_\perp)$ is the zeroth-order Bessel function of the first kind. We use $N(z)$ from the Monte Carlo simulation of i-dropouts by Bouwens et al. (2015).

To specify the CLF parameters, we simultaneously fit the model with the $z \sim 6$ UV luminosity function of Bouwens et al. (2015) and the angular correlation function of LBGs of Harikane et al. (2016). using the MCMC method using EMCEE package (Foreman-Mackey et al. 2013). We assume a Gaussian likelihood and only use the diagonal element of the error covariance matrix. We assume flat priors for all CLF parameters. The best-fitting parameters are computed as the 50 percentiles of the posterior distributions. We use the best-fitting CLF parameters in the analysis throughout this paper. The result of joint fitting procedure is shown in Fig. A1 and the best-fitting parameters are tabulated in Table A1. With a larger data set we can readily improve our analysis by simultaneously fitting the CLF parameters, $\langle f_{\rm esc}\rangle$, and $M_{\rm UV}^{\rm lim}$ in a full MCMC framework. For the purpose of the paper, we keep the CLF parameters to be fixed at the best-fitting values for simplicity.

For the application to cosmological radiative transfer, we need to specify the luminosity-dependent cross-power spectrum between our LBG samples ($M_{\rm UV} < -21$) and galaxies with luminosity L. In halo model, this is given by

$$P_g(k, L) = P^{1h}_{\rm cs}(k, L) + P^{1h}_{\rm sc}(k, L) + P^{1h}_{\rm ss}(k, L) + P^{2h}_{\rm cc}(k, L) + P^{2h}_{\rm cs}(k, L) + P^{2h}_{\rm sc}(k, L) + P^{2h}_{\rm ss}(k, L). \tag{A26}$$

where

$$P^{1h}_{\rm xy}(k, L) = \int {\rm d}M_h \mathcal{H}_{\rm x}(k, M_h)\mathcal{C}_{\rm y}(k, L, M_h)\frac{{\rm d}n}{{\rm d}M_h}, \tag{A27}$$

$$P^{2h}_{\rm xy}(k, L) = P_m(k)\int {\rm d}M_h \mathcal{H}_{\rm x}(k, M_h) b_h(M_h)\frac{{\rm d}n}{{\rm d}M_h} \times \int {\rm d}M_h' \mathcal{C}_{\rm y}(k, L, M_h') b_h(M_h')\frac{{\rm d}n}{{\rm d}M_h'}, \tag{A28}$$

and $\mathcal{H}_{\rm x}(k, M_h)$ is defined in the same way as equation (A24) but using a luminosity threshold $L_{\rm th}$ corresponding to our LBG samples, and

$$\mathcal{C}_c(k, L, M) = \frac{\Phi_{\rm cen}(L|M_h)}{\Phi(L)}, \ \mathcal{C}_s(k, M) = \frac{\Phi_{\rm sat}(L|M_h)}{\Phi(L)}\tilde{u}(k|M_h). \tag{A29}$$

Finally, using the best-fitting CLF parameters we evaluate and substitute equation (A26) into the luminosity-weighted galaxy power spectrum, equation (14), to model the enhanced photoionization rate around LBGs throughout this paper.

APPENDIX B: LINEAR THEORY

Here, we quantify the effect of galaxy-gas density correlation on the mean Lyα transmitted flux around LBGs. While in the main analysis we use a fully non-linear galaxy–galaxy correlation function in the UV background around LBGs $\langle\Gamma_{\rm H\,I}(r)\rangle$ (Appendix A), to examine the relative contribution of galaxy-galaxy and galaxy-gas density correlations, we use the linear theory so that a fair comparison of the two competing effects can be made at the same linear order.

Taylor expanding our model of the mean Lyα transmitted flux around LBGs (equation 7) in terms of the photoionization rate, we

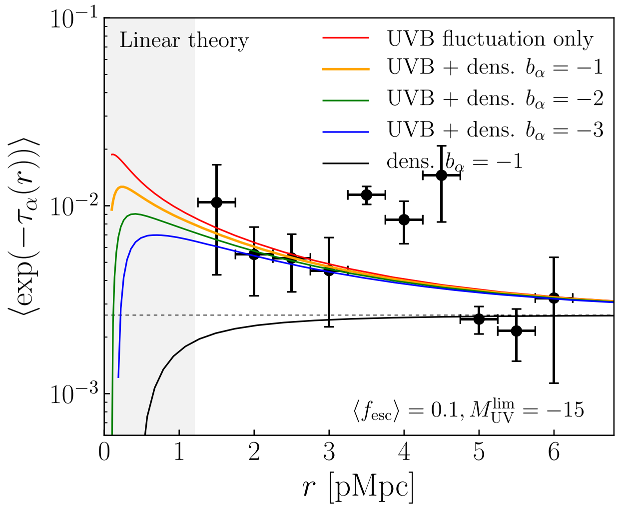

Figure B1. Effect of the galaxy–gas density correlation on the mean Lyα transmitted flux around LBGs in the linear regime. The BOSS-based estimate of Lyα forest bias $b_\alpha \simeq -1$ at z= 5.8 (orange) shows only a modest effect of matter correlation on the galaxy–Lyα forest cross-correlation on the large-scale ($\gtrsim$1 pMpc) presented in the paper. The galaxy–gas density correlation only case is also shown (black). All models assume $\langle f_{\rm esc}\rangle = 0.1$, $M_{\rm UV}^{\rm lim} = -15$, $\lambda_{\rm mfp} = 6$ pMpc, $T = 10^4$ K, and the best-fitting CLF parameters.

find

$$\langle \exp(-\tau_\alpha(r))\rangle \approx \bar{F}_\alpha \left[1 + b_\Gamma \langle \delta_\Gamma(r)\rangle\right], \tag{B1}$$

where $\bar{F}_\alpha = \int {\rm d}\Delta_b P_V(\Delta_b) {\rm e}^{-\bar{\tau}_\alpha(\bar{\Gamma}_{\rm H\,I}, T)\Delta_b^2}$ is the mean Lyα transmitted flux of the IGM. The UV background fluctuation $\langle \delta_\Gamma(r)\rangle = \langle \Gamma_{\rm H\,I}(r)\rangle / \bar{\Gamma}_{\rm H\,I} - 1$ reduces to

$$\langle \delta_\Gamma(r)\rangle = b_{\rm LBG} \langle b_g\rangle_L \int \frac{k^2 {\rm d}k}{2\pi^2} R(k\lambda_{\rm mfp}) P_m^{\rm lin}(k) \frac{\sin kr}{kr}, \tag{B2}$$

in the linear regime, and the bias factor is the response of the Lyα transmitted flux to a small perturbation of UV background,

$$\begin{aligned} b_\Gamma &= \frac{1}{\bar{F}_\alpha} \left.\frac{{\rm d}\langle F_\alpha\rangle}{{\rm d}\langle \delta_\Gamma\rangle}\right|_{\langle\delta_\Gamma\rangle = 0}, \\ &= \frac{1}{\bar{F}_\alpha} \int {\rm d}\Delta_b P_V(\Delta_b) \bar{\tau}_\alpha(\bar{\Gamma}_{\rm H\,I}, T) \Delta_b^2 {\rm e}^{-\bar{\tau}_\alpha(\bar{\Gamma}_{\rm H\,I}, T)\Delta_b^2}. \end{aligned} \tag{B3}$$

This shows that our non-linear model is equivalent to the well-known linear theory (Font-Ribera et al. 2013; du Mas des Bourboux et al. 2017) at the correct limit.

Thus, following the linear theory model, the contribution of galaxy-gas density correlation can be included as (Font-Ribera et al. 2013; du Mas des Bourboux et al. 2017)

$$\langle \exp(-\tau_\alpha(r))\rangle \approx \bar{F}_\alpha \left[1 + b_\Gamma \langle \delta_\Gamma(r)\rangle + b_{\rm LBG} b_\alpha \xi_m^{\rm lin}(r)\right], \tag{B4}$$

where b_α is the Lyα forest bias factor and $\xi_m^{\rm lin}(r)$ is the linear matter correlation function. We estimate the Lyα forest bias using the BOSS Lyα forest result $b_\alpha(z) \simeq -0.134[(1+z)/(1+2.4)]^{2.9}$ (Slosar et al. 2011; du Mas des Bourboux et al. 2017), leading $b_\alpha \simeq -1$ at $z = 5.8$. To complement this large extrapolation, we also examine the cases with $b_\alpha \simeq -2$ and -3.

In Fig. B1 we show the effect of the galaxy-gas density correlation on the mean Lyα transmitted flux around LBGs. The increasing mean gas overdensity around LBGs reduces the Lyα transmission at smaller radii as argued in the main text. The effect would become prominent only at smaller scale ($\lesssim$1 pMpc), which is below the scale presented in the paper. The relative contribution is below 10 per cent for the BOSS-based estimate $b_\alpha \simeq -1$ at the innermost bin (1.5 pMpc), and only modestly increases with Lyα forest bias at the scale of interest. The effect of galaxy-gas density correlation should thus be small. Note that, regardless of the precise value of the effect, the contribution of galaxy-gas density correlation requires *more* ionizing photons to match the observed Lyα transmitted flux in order to compensate the mean gas overdensity, leading to an even higher value of escape fraction. Our main result will therefore remain unchanged.

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