Mon. Not. R. Astron. Soc. **386,** 359–369 (2008) doi:10.1111/j.1365-2966.2008.13029.x

Glimpsing through the high-redshift neutral hydrogen fog

S. Gallerani,[1⋆] A. Ferrara,[1⋆] X. Fan[2⋆] and T. Roy Choudhury[3⋆]

[1]*SISSA/International School for Advanced Studies, via Beirut 2-4, 34014 Trieste, Italy*
[2]*Steward Observatory, The University of Arizona, Tucson, AZ 85721, USA*
[3]*Institute of Astronomy, Madingley Road, Cambridge CB3 0HA*

Accepted 2008 January 28. Received 2008 January 28; in original form 2007 June 7

ABSTRACT

We analyse the transmitted flux in a sample of 17 QSOs spectra at $5.74 \leqslant z_{\rm em} \leqslant 6.42$ to obtain tighter constraints on the volume-averaged neutral hydrogen fraction, $x_{\rm H\,I}$, at $z \approx 6$. We study separately the narrow transmission windows (peaks) and the wide dark portions (gaps) in the observed absorption spectra. By comparing the statistics of these spectral features with a semi-analytical model of the Lyα forest, we conclude that $x_{\rm H\,I}$ evolves smoothly from $10^{-4.4}$ at $z = 5.3$ to $10^{-4.2}$ at $z = 5.6$, with a robust upper limit $x_{\rm H\,I} < 0.36$ at $z = 6.3$. The frequency and physical sizes of the peaks imply an origin in cosmic underdense regions and/or in H II regions around faint quasars or galaxies. In one case (the intervening H II region of the faint quasar RD J1148+5253 at $z = 5.70$ along the line of sight of SDSS J1148+5251 at $z = 6.42$) the increase of the peak spectral density is explained by the first-ever detected transverse proximity effect in the H I Lyα forest; this indicates that at least some peaks result from a locally enhanced radiation field. We then obtain a strong lower limit on the foreground QSO lifetime of $t_{\rm Q} > 11$ Myr. The observed widths of the peaks are found to be systematically larger than the simulated ones. Reasons for such discrepancy might reside either in the photoionization equilibrium assumption or in radiative transfer effects.

Key words: intergalactic medium – quasars: absorption lines – large-scale structure of Universe.

1 INTRODUCTION

Although observations of cosmic epochs closer to the present have indisputably shown that the intergalactic medium (IGM) is in an ionized state, it is yet unclear when the phase transition from the neutral state to the ionized one started. Thus, the redshift of reionization, $z_{\rm rei}$, is still very uncertain.

After the first-year *Wilkinson Microwave Anisotropy Probe* (*WMAP*) data a possible tension was identified between cosmic microwave background and Sloan Digital Sky Survey (SDSS) results. The high electron-scattering optical depth inferred from the TE–EE power spectra $\tau_{\rm e} \approx 0.17$ (Kogut et al. 2003; Spergel et al. 2003) seemed difficult to be reconciled with the strong evolution in the Gunn–Peterson (GP) optical depth $\tau_{\rm GP}$ at $z = 6$ (Fan et al. 2001, 2003), accompanied by the appearance of large dark portions in QSO absorption spectra (Becker et al. 2001; Djorgovski et al. 2001; Fan et al. 2006, hereafter F06). The three-year *WMAP* results have released the above tension by providing a smaller value for $\tau_{\rm e} \approx 0.1$, which implies $z_{\rm rei} \approx 11$ for a model with instantaneous reionization (Page et al. 2007; Spergel et al. 2007). However, an increasing number of Lyman α emitters (LAEs) are routinely found at $z > 6$ (e.g. Stern et al. 2005; Iye et al. 2006; Stark et al. 2007), possibly requiring a substantial free electron fraction, resulting in an IGM relatively transparent to Lyα photons.

Constraints on the IGM ionization state derived by using Lyα forest spectroscopy must take into account the extremely high sensitivity of $\tau_{\rm GP}$ to tiny neutral hydrogen amounts. Indeed, a volume averaged neutral hydrogen fraction as low as $x_{\rm H\,I} \sim 10^{-3}$ (Fan et al. 2002) is sufficient to completely depress the transmitted flux in QSO absorption spectra; thus, the detection of a GP trough only translates into a lower limit for $x_{\rm H\,I}$. For this reason, recently, many studies have tried to clarify if the SDSS data effectively require that the IGM was reionized as late as $z \approx 6$ (Gallerani, Choudhury & Ferrara 2006, hereafter GCF06; Becker, Rauch & Sargent 2007): in particular, GCF06 have shown that QSO observational data currently available are compatible with a highly ionized universe at that redshift.

Clearly the determination of the reionization epoch is strictly related to the measurement of the neutral hydrogen fraction at $z \approx 6$. To investigate this issue many different approaches can be used. To start with, it is worth mentioning that many authors have tried to constrain $x_{\rm H\,I}$ at high redshift by analysing statistically the optical depth inside H II regions around high-redshift QSO (Mesinger & Haiman 2004, 2007), or by measuring the QSO bubble sizes (Wyithe & Loeb 2004; Wyithe, Loeb & Carilli 2005a; Maselli et al. 2007;

⋆E-mail: galleran@sissa.it (SG); ferrara@sissa.it (AF); fan@as.arizona.edu (XF); chou@ast.cam.ac.uk (TRC)

Bolton & Haehnelt 2007a; Lidz et al. 2007). However, sufficient ground for controversy remains due to intrinsic uncertainties of the various techniques.

By deriving sizes of H II bubbles surrounding observed $z = 6.5$ LAEs, Malhotra & Rhoads (2006) have provided an upper limit for the neutral hydrogen fraction $x_{\rm H\,I} \lesssim 0.2$–$0.5$. This result is in quite good agreement with the upper limit $x_{\rm H\,I} \lesssim 0.45$ found by Kashikawa et al. (2006), by interpreting the deficit measured at the bright end of the LAE luminosity function at $z > 6$ as a sudden change in the intergalactic neutral hydrogen content. Nevertheless, the increasing attenuation with redshift of the Lyα line transmission could be partially explained as a consequence of the evolution in the mass function of dark matter haloes, thus implying a much lower upper limit, $x_{\rm H\,I} < 0.05$–0.2 (Dijkstra, Wyithe & Haiman 2007).

Gamma-ray burst (GRB) spectroscopy has also tentatively used to constrain $x_{\rm H\,I}$; Totani et al. (2006) have observed a damping wing at wavelengths larger than the Lyα emission line, finding that this feature can be explained at best by assuming an intervening damped Lyα system immersed in a fully ionized IGM, and quoting an upper limit of $x_{\rm H\,I} < 0.17$ and 0.60 (68 and 95 per cent confidence levels, respectively).

Finally, the width distribution of dark portions (gaps) seen in QSO absorption spectra has been recently introduced in order to constrain the IGM ionization state (Paschos & Norman 2005; GCF06; F06). F06 has used the dark gap distribution, as observed in 19 high-z QSO spectra, to put a preliminary upper limit on the IGM neutral fraction $x_{\rm H\,I} < 0.1$–0.5. GCF06, by analysing the statistically properties of the transmitted flux in simulated absorption spectra, have shown that the gap and peak (i.e. transmission windows) width statistics are very promising tools for discriminating between an early ($z_{\rm rei} > 6$) and a late ($z_{\rm rei} \approx 6$) reionization scenario. Here we combine the previous two results: by comparing the observed transmitted flux in high-z QSO spectra with theoretical predictions we obtain tighter constraints on the neutral hydrogen fraction around $z = 6$, a crucial epoch in the reionization history.

The plan of the paper is the following. In Section 2 we describe the semi-analytical modelling adopted; in Section 3 we compare observational data with simulations. The implications of this comparison are given in Section 4; in Section 5 we evaluate the robustness of our method against the specific line of sight (LOS) to the highest redshift QSO. The conclusions are summarized in Section 6.

2 SIMULATIONS

The radiation emitted by QSOs could be absorbed through Lyα transition by the neutral hydrogen intersecting the LOS, the so-called GP effect. The Lyα forest arises from absorption by low amplitude fluctuations in the underlying baryonic density field (Bi, Börner & Chu 1992), and is a natural consequence of the hierarchical structure formation expected in the context of cold dark matter (CDM) cosmologies.[1]

To simulate the GP optical depth ($\tau_{\rm GP}$) distribution we use the method described by GCF06, whose main features are recalled in the following. The spatial distribution of the baryonic density field and its correlation with the peculiar velocity field are taken into account adopting the formalism introduced by Bi & Davidsen (1997). To enter the mildly non-linear regime which characterizes the Lyα forest absorbers we use a lognormal (LN) model introduced by Coles & Jones (1991), widely adopted later on (Bi 1993; Bi & Davidsen 1997; Choudhury, Padmanabhan & Srianand 2001a; Choudhury, Srianand & Padmanabhan 2001b; Viel et al. 2002; GCF06). In particular, GCF06 have compared various Lyα statistics, namely the probability distribution function (PDF) and the gap width (GW) distribution, computed using the LN distribution with those obtained from HYDROPM simulations, finding a good agreement between the results. For a given IGM temperature, the H I fraction, $x_{\rm H\,I}$, can be computed from the photoionization equilibrium as a function of the baryonic density field and photoionization rate due to the ultraviolet background radiation field. For all these quantities we follow the approach of Choudhury & Ferrara (2006), hereafter CF06. By assuming as ionizing sources QSOs, Population II and Population III stars (the latter neglected here, see below), their model provides excellent fits to a large number of observational data, namely the redshift evolution of Lyman-limit systems, Lyα and Lyβ optical depths, electron scattering optical depth, cosmic star formation history, and the number counts of high-redshift sources. In the CF06 model, a reionization scenario is defined by the product of two free parameters: (i) the star formation efficiency f_* and (ii) the escape fraction $f_{\rm esc}$ of ionizing photons of Population II and Population III stars; it is worth noting that these parameters are degenerate, since different parameter values could provide equally good fits to observations. In this work, by fitting all the above observational constraints, we select two sets of free parameters values yielding two different reionization histories: (i) an early reionization model (ERM) for ($f_{*,\rm PopulationII} = 0.1$; $f_{\rm esc,PopulationII} = 0.07$) and (ii) a late reionization model (LRM) for ($f_{*,\rm PopulationII} = 0.08$; $f_{\rm esc,PopulationII} = 0.04$). We do not consider contributions from Population III stars, as Population II stars alone yield $\tau_{\rm e} = 0.07$ (0.06) for ERM (LRM), marginally consistent with WMAP3 results.[2]

Fig. 1 shows the global properties of the two reionization models considered. In the ERM the volume filling factor of ionized regions, $Q_{\rm H\,II} = V_{\rm H\,II}/V_{\rm tot} = 1$ at $z \leqslant 7$; in the LRM it evolves from 0.65 to unity in the redshift range 7.0–6.0, implying that the universe is still in the pre-overlap stage at $z \geqslant 6$, i.e. the reionization process is not completed up to this epoch. In the middle panel of the same figure we compare the volume-averaged photoionization rate $\Gamma_{\rm H\,I}$ for the two models with the recent estimate by F06, and the ones by McDonald & Miralda-Escudé (2001); Bolton et al. (2005); Bolton & Haehnelt (2007b), hereafter MM01, B05 and B07, respectively. Finally, the evolution of the volume-averaged neutral hydrogen fraction for the ERM and LRM is presented in the rightmost panel.

The photoionization rate predicted by both models is in agreement with the results by B05 and B07 at in the range $4.0 < z < 6$, whereas at $z = 5.5$ (6) the ERM is characterized by a photoionization rate which is $\approx$2 (6) times larger than the estimates by F06. In spite of these differences, our predictions for $x_{\rm H\,I}$ are consistent with F06 measurements. This apparent contradiction does not come as a surprise. In fact, the derivation of $\Gamma_{\rm H\,I}$ requires an assumption

[1] Throughout this paper we will assume a flat universe with total matter, vacuum and baryonic densities in units of the critical density of $\Omega_{\rm m} = 0.24$, $\Omega_\Lambda = 0.76$ and $\Omega_{\rm b}\,h^2 = 0.022$, respectively, and a Hubble constant of $H_0 = 100\,h\,{\rm km\,s^{-1}\,Mpc^{-1}}$, with $h = 0.73$. The parameters defining the linear dark matter power spectrum are $n = 0.95$, ${\rm d}n/{\rm dln}\,k = 0$, $\sigma_8 = 0.82$. Note that we have chosen a σ_8 value higher than the *WMAP*3 one (0.74). Indeed Viel, Haehnelt & Lewis (2006), by combining *WMAP*3 data with SDSS ones, found $\sigma_8 \approx 0.78$ (0.86) analysing high-resolution (low-resolution) Lyα forest data. Mpc are physical unless differently stated.

[2] Small contributions from Population III stars, i.e. $f_{*,\rm PopulationIII} = 0.013$ ($f_{*,\rm PopulationIII} = 0.08$), in the ERM (LRM), would yield $\tau_{\rm e} = 0.09$ ($\tau_{\rm e} = 0.08$), without affecting sensitively the results below.

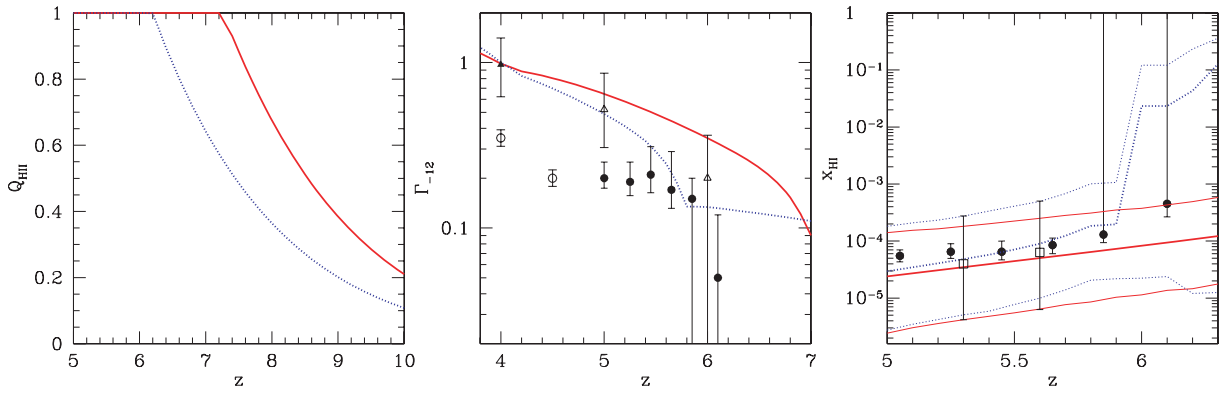

Figure 1. Left-hand panel: Evolution of the volume filling factor of ionized regions for the early (red solid lines) and late (blue dotted lines) reionization models. Middle panel: Volume-averaged photoionization rate $\Gamma_{-12} = \Gamma_{\rm H\,I}/10^{-12}\,{\rm s}^{-1}$. The filled circles, empty circles, filled triangles and empty triangles show results obtained by F06, MM01, B05 and B07, respectively. Right-hand panel: Evolution of the neutral hydrogen fraction. Thick lines represent average results over 100 lines of sight, while the thin lines denote the upper and lower neutral hydrogen fraction extremes in each redshift interval. Solid circles represent neutral hydrogen fraction estimates by F06; empty squares denote the results obtained in this work.

concerning the IGM density distribution. When measuring $\Gamma_{\rm H\,I}$ at $5 < z < 6$, F06 assume the density PDF given by Miralda-Escudé, Haehnelt & Rees (2000), hereafter MHR.[3] We instead adopt an LN model which predicts a higher probability to find overdensities $\Delta = \rho/\bar{\rho} \gtrsim 1$ than MHR. For example, at $z = 6$ and for $\Delta \approx 1.5$, $P_{\rm LN}(\Delta) \approx 2 \times P_{\rm MHR}(\Delta)$. For this reason, once $\tau_{\rm GP}$ is fixed to the observed value, the LN model requires a higher $\Gamma_{\rm H\,I}$. As $x_{\rm H\,I} \propto \Delta$, these two effects combine to give a values of $x_{\rm H\,I}$ consistent with the data.

3 COMPARISON WITH OBSERVATIONS

3.1 Control statistics

We first test the predictions of our model by applying various statistical analysis to the simulated spectra and comparing our results with observations. Specifically, we use the following control statistics: (i) Mean transmitted flux evolution in the redshift range 2–6; (ii) PDF of the transmitted flux at the mean redshifts $z = 5.5$, 5.7, 6.0; (iii) GW distribution in $3.5 \leqslant z \leqslant 5.5$. For what concerns the GW statistics we define gaps as contiguous regions of the spectrum having a $\tau_{\rm GP} > 2.5$ over rest-frame wavelength ($\lambda_{\rm RF}$) intervals >1 Å. This method was first suggested by (Croft 1998) and then adopted by various authors (Songaila & Cowie 2002; Paschos & Norman 2005; F06; GCF06). The comparison of model and observational results in terms of the above three statistics is plotted in Fig. 2. By checking our models we follow the same approach of GCF06, to which we refer for a complete description of the technical details.

The outcome of the test is encouraging, as both ERM and LRM successfully match the observational data at $z \leqslant 6$ for the control statistics considered. This allows us to confidently proceed with the comparison of more advanced statistical tools.

3.2 Advanced statistics

Since at $z \approx 6$ regions with high transmission in the Lyα forest become rare, an appropriate method to analyse the statistical properties of the transmitted flux is the distribution of gaps. In particular GCF06 suggested that the largest gap width (LGW) and the largest peak width (LPW) statistics are suitable tools to study the ionization state of the IGM at high redshift.[4] The LGW (LPW) distribution quantifies the fraction of lines of sight which is characterized by the largest gap (peak) of a given width. As far as this work is concerned, we apply the LGW and the LPW statistics both to simulated and observed spectra with the aim of measuring the evolution of $x_{\rm H\,I}$ with redshift.

We use observational data including 17 QSOs obtained by F06. We divide the observed spectra into two redshift-selected subsamples: the 'low-redshift' (LR) sample (eight emission redshifts $5.7 < z_{\rm em} < 6$), and the 'high-redshift' (HR) one (nine emission redshifts $6 < z_{\rm em} < 6.4$). Simulated spectra have the same $z_{\rm em}$ distribution of the observed samples. For most QSOs we consider the ($\lambda_{\rm RF}$) interval 1026–1200 Å and we normalize each width to the corresponding redshift path. Note that the lines of sight do not extend up to $z_{\rm em}$; the upper (lower) limit of the interval chosen ensures that we exclude from the analysis the portions of the spectra penetrating inside the QSO H II (Lyβ) region. For the QSOs SDSS J1044−0125 and SDSS J1048+4637 we choose different intervals, namely 1050–1183 and 1050–1140, respectively. These two objects have been classified as BAL QSO (Goodrich et al. 2001; Fan et al. 2003; Maiolino et al. 2004), since their spectra present broad absorption lines associated with highly ionized atomic species (e.g. Si IV, C IV). By selecting the above intervals we exclude those portions of the spectra characterized by C IV absorption features which extend to $z \approx 5.56$ ($z \approx 5.75$) in SDSS J1044−0125 (SDSS J1048+4637). Observed data were taken with a spectral resolution $R \approx$ 3000–6000; simulated spectra have been convolved with a Gaussian of FWHM = 67 km s^{-1}, providing $R \sim 4500$. Moreover each observed/simulated spectrum has been rebinned to a resolution of $R = 2600$. Finally, we add noise to the simulated data such that the flux F in each pixel is replaced by $F + G(1)\sigma_n$, where $G(1)$ is a Gaussian random deviate with zero mean and unit variance, and σ_n is the observed noise rms deviation of the corresponding pixel.

[3] F06 require $\Gamma_{\rm H\,I}$ to match the MM01 measurement at $z = 4.5$. This estimate is based on a mean transmitted flux ($\bar{F} = 0.25$) which is lower than the more recent measurements $\bar{F} \approx 0.32$ by Songaila (2004), which implies $\Gamma_{\rm H\,I} \approx 0.3$.

[4] The definition of 'peak' in the transmitted flux is similar to the 'gap' one. A peak is a contiguous region of the spectrum over $\lambda_{\rm RF}$ intervals greater than the observed pixel size (≈0.5 Å) characterized by a transmission above a given flux threshold ($F_{\rm th} = 0.08$ in this work).

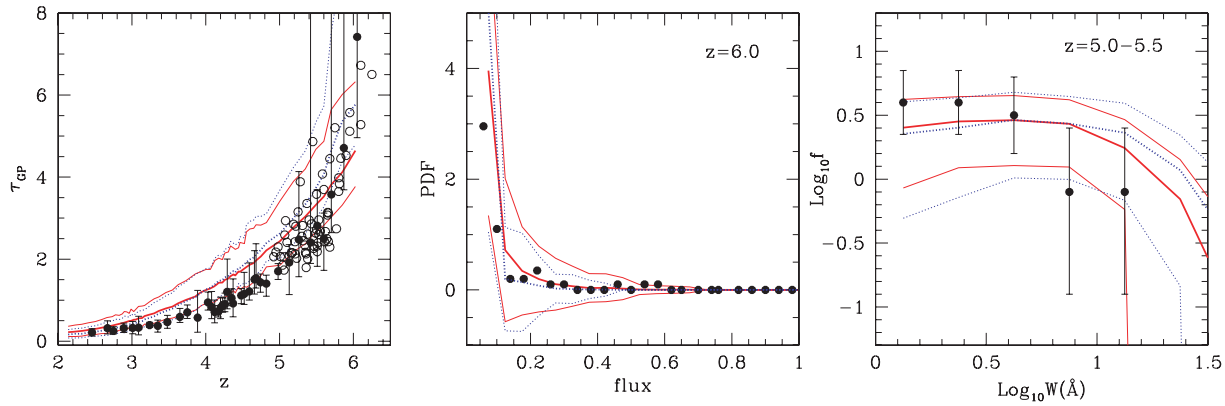

Figure 2. Left-hand panel: Evolution of the GP optical depth for early (ERM, solid red line) and late (LRM, blue dotted). Thick lines represent average results on 100 lines of sight for each emission redshift, while the thin lines denote the upper and lower transmission extremes in each redshift bin, weighted on 100 lines of sight. Filled and empty circles are observational data from Songaila (2004) and F06, respectively. Middle panel: PDF of the flux at $z = 6.0$. Filled circles are obtained by Fan et al. (2002). Thick lines represent simulated results averaged over 500 lines of sight, while the thin lines denote cosmic variance. Right-hand panel: GW distribution in the redshift range 5.0–5.5. Simulated results are compared with observations by Songaila & Cowie (2002) (filled circles). The errors associated to both simulated and observed results denote cosmic variance.

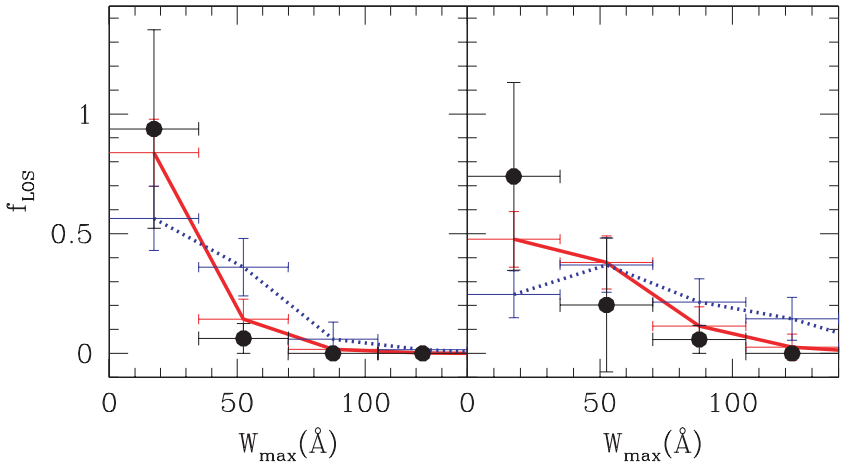

Figure 3. LGW distribution for the LR and the HR cases (left- and right-hand panels, respectively). Filled circles represent the result of the analysis of the 17 QSOs observed spectra. Solid red (dotted blue) lines show the results obtained by the semi-analytical modelling implemented for the ERM (LRM). Vertical error bars measure Poissonian noise, horizontal errors define the bin for the GWs.

The results provided by the statistics adopted in this study are sensitive to the signal-to-noise ratio (S/N), since spurious peaks could arise in spectral regions with noise higher than the F_{th} adopted. Indeed, the shape of the LGW/LPW distributions depends on the F_{th} chosen. Thus, we consider two different values for F_{th}, namely 0.03 and 0.08, respectively, and, for both of them, compute preliminary LGW/LPW distributions. Finally, the LGW/LPW distributions presented are obtained as the mean of the preliminary ones, weighted on the corresponding errors (see Appendix A for a detailed discussion). In our analysis we do not consider the two QSOs presented by F06, namely SDSS J1436+5007 and SDSS J1630+4012, since these spectra have significantly lower S/N to apply LGW/LPW tests (continuum S/N $\lesssim 7$).

3.2.1 LGW distribution

We now discuss the LGW distribution for observed/simulated spectra; the results are shown in Fig. 3. The QSOs emission redshifts used and the λ_{RF} interval chosen for the LR sample are such that the mean redshift of the absorbers is $\langle z \rangle = 5.26$, with a minimum (maximum) redshift $z_{min} = 4.69$ ($z_{max} = 5.86$), and an rms deviation $\sigma = 0.06$. For the HR sample it is $\langle z \rangle = 5.55$, $z_{min} = 4.90$, $z_{max} = 6.32$, $\sigma = 0.14$. The observed LGW distribution evolves rapidly with redshift: in the LR sample most of the lines of sight are characterized by a largest gap <40 Å, whereas gaps as large as 100 Å appear in the HR sample. This means that lines of sight to QSOs emitting at $z_{em} \lesssim 6$ encounter 'optically thick' regions whose size is $\leqslant$20 Mpc, while for $z_{em} \gtrsim 6$ blank regions of size up to 46 Mpc are present. Superposed to the data in Fig. 3 are the predicted LGW distributions corresponding to ERM and LRM, obtained by simulating 800/900 lines of sight in the LR/HR case, corresponding to 100 lines of sight for each emission redshift in each sample. In our ERM simulated spectra, at $z \approx 6$ gaps are produced by regions characterized by a mean overdensity $\bar{\Delta} \approx 1 (\Delta_{min} = 0.05,\ \Delta_{max} = 18)$ with a $x_{H\,I} \approx 10^{-4}$, averaging on 100 lines of sight ($x_{H\,I,min} = 1.1 \times 10^{-5}$, $x_{H\,I,max} = 3.6 \times 10^{-4}$).

It results that both the predicted LGW distributions provide a good fit to observational data. We exploit the agreement between

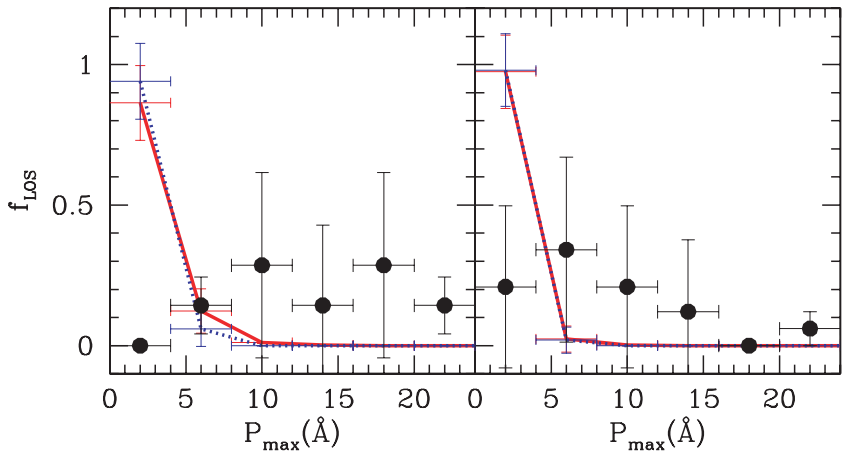

Figure 4. LPW distribution for the LR and the HR cases (left- and right-hand panels, respectively). Filled circles represent observational data obtained by analysing the observed spectra of the 17 QSOs considered. Solid red (dotted blue) lines show the results obtained by the semi-analytical modelling implemented for the ERM (LRM). Vertical error bars measure Poissonian noise, horizontal errors define the bin for the peak widths.

the simulated and observed LGW distributions to derive an estimate of $x_{\rm H\,I}$, shown in Fig. 1. We find $\log_{10}x_{\rm H\,I} = -4.4^{+0.84}_{-0.90}$ at $z \approx 5.3$.[5] By applying the same method to the HR sample we constrain the neutral hydrogen fraction to be within $\log_{10}x_{\rm H\,I} = -4.2^{+0.84}_{-1.0}$ at $z \approx 5.6$.

Although the predicted LGW distributions are quite similar for the two models considered, yet some differences can be pointed out. Both for the LR and HR cases the early reionization LGW distribution provides a very good match to the observed points, thus suggesting $z_{\rm rei} \gtrsim 7$. The agreement is satisfactory also for the LRM, but it is important to note that LRMs predict too many largest gaps $\approx$60 Å in the LR case and too few gaps $\approx$20 Å in the HR one. Given the limited quasar sample available, the statistical relevance of the LRM discrepancies is not sufficient to firmly rule out this scenario. However, since in the HR case 40 per cent of the lines of sight extend at $z \gtrsim 6$, we can use the LRM results to put an upper limit on $x_{\rm H\,I}$ at this epoch. Indeed in the HR case we find that a neutral hydrogen fraction at $z \approx 6$ higher than that one predicted by the LRM would imply an even worst agreement with observations, since a more abundant H I would produce a lower (higher) fraction of lines of sight characterized by the largest gap smaller (higher) than 40 Å with respect to observations. Thus, this study suggests $x_{\rm H\,I} < 0.36$ at $z = 6.32$ (obtained from the maximum value for $x_{\rm H\,I}$ found in the LRM at this epoch).

3.2.2 LPW distribution

Next, we apply the LPW statistics (Fig. 4) to both observed and simulated spectra. From the observed LPW distribution we find that, in the LR (HR) sample, about 50 per cent of the lines of sight exhibit peaks of width <12(8) Å. In more details, the size $P_{\rm max}$ of the largest transmission regions in the observed sample are $3 \lesssim P_{\rm max} \lesssim 10$ ($1 \lesssim P_{\rm max} \lesssim 6$) Mpc at $\langle z \rangle = 5.3$ (5.6). The frequency and the amplitude of the transmission regions rapidly decrease towards high redshift. This could be due both to the enhancement of the neutral hydrogen abundance at epochs approaching reionization or to evolutionary effects of the density field (Songaila 2004). In fact the growth factor D_+ of density fluctuations decreases with redshift [$D_+(z = 6) = (3/5)\, D_+(z = 3)$ for Λ CDM], thus implying a low density contrast at $z = 6$ with respect to later epochs. Stated differently, underdense regions that are transparent at $z = 3$, were less underdense at $z = 6$, thus blocking transmission. As a consequence of the density field evolution towards higher z, only few/small peaks survive and wide GP troughs appear.

Superposed to the data in Fig. 4 are the predicted LPW distributions corresponding to ERM and LRM, obtained by simulating 800/900 lines of sight in the LR/HR case. In our ERM simulated spectra, at $z \approx 6$, gaps are interrupted by narrow transparent windows (i.e. peaks) originating from underdense regions with $\bar{\Delta} \approx 0.1$, averaging on 100 lines of sight ($\Delta_{\rm min} = 0.03$, $\Delta_{\rm max} = 0.26$) and $x_{\rm H\,I} \approx 2 \times 10^{-5}$, ($x_{\rm H\,I,min} = 7.8 \times 10^{-6}$, $x_{\rm H\,I,max} = 3.6 \times 10^{-5}$). Regions characterized by $\Delta \in [0.05;0.26]$ and $x_{\rm H\,I} \in [1.1 \times 10^{-5}\,;3.6 \times 10^{-5}]$ could correspond to both gaps or peaks depending on redshift and peculiar motions of the absorbers producing them.

By comparing the simulated LPW distributions with the observed one, it is evident that simulations predict peak widths that are much smaller than the observed ones both for LR and HR cases. In particular, in no LOS of our simulated samples we find peaks larger than 8 Å. The disagreement between the observed and simulated LPW distributions does not affect the estimate of $x_{\rm H\,I}$ through the LGW distributions, since at high redshift the peaks are narrow ($\lesssim$10 Å). We discuss the possible reasons for this discrepancy in the final section.

4 PHYSICAL INTERPRETATION OF THE PEAKS

The most natural interpretation for the peaks is that they correspond to underdense regions, where the low H I density of the gas allows a high transmissivity. However, in principle they could also arise if individual ionized bubbles produced by QSOs and/or galaxies are crossed by the LOS. In the latter case the typical physical size and frequency of such semitransparent regions must be related to the emission properties and masses of such objects. Stated differently, the fraction of lines of sight, $f_{\rm LOS}$, having the LPW equal to $P_{\rm max}$ can be interpreted as the probability $\wp$ to intersect an H II region

[5] The $x_{\rm H\,I}$ value quoted is the mean between the estimates predicted by the ERM and the LRM. Moreover, we consider the most conservative case in which the errors for the measurement of the neutral hydrogen fraction are provided by the minimum $x_{\rm H\,I}$ value found in the ERM and the maximum one in the LRM.

of radius $R_{\rm H\,II}$ around a dark matter halo hosting either a QSO or a galaxy along the redshift path ($z_{\rm i} - z_{\rm f}$) spanned by the LOS. The comoving number density $n_{\rm h}$ of dark matter haloes of mass $M_{\rm h}$ is related to $\wp$ through the following equation:

$$n_{\rm h}(M_{\rm h}) = \frac{3}{2}\frac{H_0\Omega_{\rm m}^{1/2}}{c}\left(\pi R_{\rm H\,II}^2\right)^{-1}\left[(1+z)^{3/2}\left|_{z_{\rm i}}^{z_{\rm f}}\right.\right]^{-1}\wp, \quad (1)$$

We take $R_{\rm H\,II} = 1$ (10) Mpc, consistent with the smaller (larger) size $P_{\rm max}$ of the observed largest peaks in the HR (LR) sample. As it is likely that statistically the LOS crosses the bubble with non-zero impact parameter, adopting $R_{\rm H\,II} = P_{\rm max}$ seems a reasonable assumption. By further imposing $\wp = f_{\rm LOS}$ we find that $n_{\rm h} = 3.7 \times 10^{-6}$ (2.2×10^{-8}) Mpc^{-3} for $P_{\rm max} = 1$ (10) Mpc in the redshift range $z_{\rm i} = 5$ to $z_{\rm f} = 6$. Given our cosmology, such halo number density can be transformed at $z = 5.5$ into a typical halo mass of $M_{\rm h} \gtrsim 10^{12}$ (10^{13}) M$_\odot$ (Mo & White 2002). Thus, the haloes hosting the putative luminous sources producing the peaks must be massive. Note that this result holds even if the QSO is shining only for a fraction of the Hubble time $t_{\rm Q}/t_{\rm H} \approx 10^{-2}$ at $z = 5.5$.

In addition to the peak frequency, additional constraints on the properties of the ionizing sources come from bubble physical sizes.

4.1 QSO H II regions

First, we consider the case in which the largest peaks are produced by H II regions around QSOs. The bubble size $R_{\rm H\,II}$ is related to the ionizing photons emission rate $\dot{N}_\gamma$ and QSO lifetime $t_{\rm Q}$ as

$$R_{\rm H\,II} = \left(\frac{3\dot{N}_\gamma t_{\rm Q}}{4\pi n_{\rm H\,I}}\right)^{1/3}, \quad (2)$$

where $n_{\rm H\,I}$ is the neutral hydrogen number density. Equation (2) applies for a homogeneous IGM and does not take into account both recombinations and relativistic effects.

The recombination time-scale $t_{\rm rec}$ is given by

$$t_{\rm rec} = [C\alpha_{\rm B} n_{\rm H}(1 - x_{\rm H\,I})]^{-1}, \quad (3)$$

where $C \simeq 26.2917\exp[-0.1822z + 0.003\,505z^2]$ is the clumping factor (Iliev et al. 2007), $\alpha_{\rm B} = 2.6 \times 10^{-13}$ cm^3 s^{-1} is the case B hydrogen recombination coefficient evaluated at $T = 10^4$ K, and $n_{\rm H} = 7 \times 10^{-5}\,[(1+z)/7]^3$ cm^{-3} is the mean hydrogen number density. Thus, at the redshifts of interest $z \approx 6$, even in the limiting case $x_{\rm H\,I} = 0$, $t_{\rm rec} \approx 2 \times 10^8$ yr, thus being larger than typical QSOs lifetime $t_{\rm Q} \approx 10^7$. This shows that equation (2) provides a plausible value for the H II region extent. For instance, Maselli et al. (2007) have shown that equation (2) matches quite well the mean value of the H II region size determined through radiative transfer calculations.

In this section, we neglect relativistic effects which could squash the ionization front along the sightline (White et al. 2003; Wyithe, Loeb & Barnes 2005b; Yu 2005; Shapiro et al. 2006), possibly reducing the length of the lines of sight interested by the proximity effect. We will discuss this issue in detail in Section 5, when addressing the first observed case of transverse proximity effect.

At $z = 5.5$, assuming $x_{\rm H\,I} = 5.6 \times 10^{-5}$ (see Fig. 1), $R_{\rm H\,II} = 1$ (10) Mpc could be produced by a QSO emitting a number of ionizing photons $N_\gamma = \dot{N}_\gamma t_{\rm Q} = 7 \times 10^{65}$ (7×10^{68}). Thus, assuming a QSO lifetime $\approx 10^7$ yr, the observed peaks in the LR (HR) sample require $\dot{N}_\gamma = 2.2 \times 10^{51}$($2.2 \times 10^{54}$) s^{-1}, which would correspond to sources $\approx$6 (3) orders of magnitude fainter than QSOs observed at $z \approx 6$, typically having $\dot{N}_\gamma \approx 10^{57}$ s^{-1} (Haiman & Cen 2002) and black hole masses $M_{\rm BH} \approx 10^9$ M$_\odot$.

So far we have assumed that the gas inside the H II region is fully ionized or, stated differently, that along the redshift path encompassed by the ionized bubble the flux is *completely* transmitted. However, this is unlikely since a sufficiently high opacity due to resonant (damping wing) optical depth associated with the neutral hydrogen inside (outside) the H II region can produce dark gaps. Thus, the relation between $P_{\rm max}$ and $R_{\rm H\,II}$ is

$$P_{\rm max} = \frac{H(\bar{z})\lambda_{\rm Ly\alpha}}{c}\frac{(1+\bar{z})}{(1+z_{\rm em})} f_{\rm t} R_{\rm H\,II} = A(z) f_{\rm t} R_{\rm H\,II}, \quad (4)$$

where $f_{\rm t}$ is the mean transmitted flux computed inside the proximity region. We will derive $f_{\rm t}$ in Section 5 from an observed case of transverse proximity effect, note that values of $f_{\rm t} < 1$ would result in a larger luminosity of the QSO producing the transmissivity window. Finally, powerful QSOs, as those observed at $z \approx 6$, could produce transmission windows consistent with observational data if they are embedded in overdense regions where the high density sustains an initial neutral fraction, $x_{\rm H\,I} \gtrsim 0.1$, *before the QSO turns on*. The expansion of the H II region in such environment would result in considerably smaller sizes (Maselli et al. 2007). In this case, both the host dark matter halo mass found above ($M_{\rm h} \approx 10^{12}$ M$_\odot$), and the size of the H II region would combine to give the correct frequency and spectral width of the observed peaks.

4.2 Galaxy H II regions

In addition to QSOs, transmissivity windows could be produced by H II regions around high-z galaxies. Adopting the canonical relations

$$M_* = f_* \frac{\Omega_{\rm b}}{\Omega_{\rm m}} M_{\rm h}; \qquad N_\gamma = \bar{n}_\gamma \frac{M_*}{m_{\rm p}}; \qquad f_{\rm esc} N_\gamma = \frac{4\pi}{3} n_{\rm H\,I} R_{\rm H\,II}^3, \quad (5)$$

where M_* is the stellar mass, $\bar{n}_\gamma$ is the number of ionizing photons per baryon into stars, and $m_{\rm p}$ is the proton mass, the relation between $M_{\rm h}$ and $R_{\rm H\,II}$ is given by

$$M_{\rm h} = 3 \times 10^8\,{\rm M}_\odot \left(\frac{1+z}{6.5}\right)^3 y_{-1} R_{\rm H\,II}^3, \quad (6)$$

where $y_{-1} = (x_{\rm H\,I} f_*^{-1} f_{\rm esc}^{-1})/0.1$ and we are assuming $\bar{n}_\gamma = 4000$, appropriate for a Population II stellar population with a standard Salpeter initial mass function; we assume the fiducial values $x_{\rm H\,I} = 5.6 \times 10^{-5}$, $f_* = 0.1$, $f_{\rm esc} = 0.01$. The mass of an halo hosting a star-forming region able to produce $R_{\rm H\,II} \approx 1$ (10) Mpc is 2×10^8 (2×10^{11}) M$_\odot$. At $z \approx 5.5$ objects of these masses corresponds to fluctuations of the density field $\gtrsim 1\sigma$ (2σ) (Barkana & Loeb 2001).

As for QSOs, the bubble size–peak frequency tension could be alleviated if the galaxies live in overdense environments where the photoionization rate only supports a $x_{\rm H\,I} \approx 0.1$ (resulting in a larger value of y_{-1}, and hence of $M_{\rm h}$ in equation (6) prior to the onset of star formation in the galaxy. Obviously, the previous arguments neglect that because of clustering (Yu & Lu 2005; Kramer, Hiaman & Oh 2006), as multiple sources could power a single H II region; in order to get firmer results radiative transfer cosmological simulations are required.

5 PEAKS FROM THE PROXIMITY EFFECT

In Section 4, we have discussed the possibility that the observed peaks are produced by ionizing sources whose bubbles intersect the lines of sight to the target QSO. In this case one could ask if the source responsible for the H II region would be detected in the observed field. If the origin of transmissivity regions resides in bubbles around high-z galaxies, these sources are too faint to be

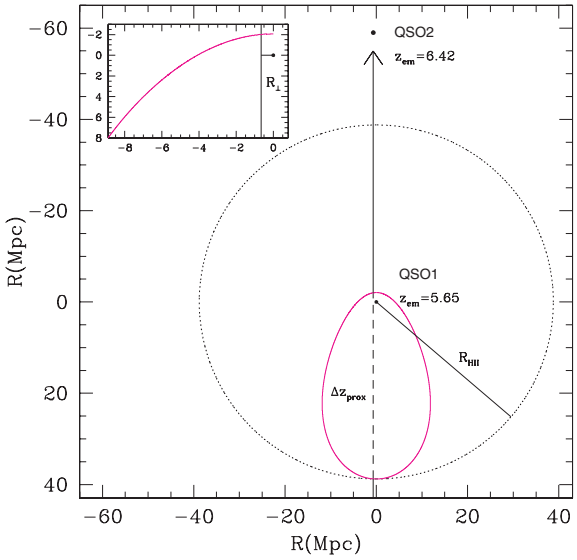

Figure 5. Schematic positions of quasars SDSS J1148+5251 (QSO2, $z_{\rm em} = 6.42$, redshift position not in scale) and RD J1148+5252 (QSO1, $z_{\rm em} = 5.65$). The projected separation is denoted by $R_\perp$, the size of the H II region, $R_{\rm H\,II}$, in the QSO1 rest frame is represented by the dotted circle; the magenta solid line shows the apparent shape of the ionization front; the dashed black line shows the redshift path ($\Delta z_{\rm prox}$) in which the bubble produced by QSO1 intersects the LOS to QSO2.

seen in the SDSS; however, deep *Hubble Space Telescope* imaging (Stiavelli et al. 2005) could detect such objects. On the contrary, if the H II region of a quasar intervenes along the LOS to an higher redshift quasar, the first could be observed in the SDSS field.

Mahabal et al. (2005) have discovered a faint quasar (RD J1148+5253, hereafter QSO1) at $z = 5.70$ in the field of the highest redshift quasar currently known (SDSS J1148+5251, hereafter QSO2) at $z = 6.42$. In this section we study the QSO2 transmitted flux, in order to analyse the proximity effect of QSO1 on the QSO2 spectrum. For clarity, Fig. 5 presents a schematic picture of the considered geometry. As the redshift $z = 5.70$ quoted by Mahabal et al. (2005) is based on the peak of the Lyα emission line, and the estimated error from such procedure is $\Delta z \approx 0.05$ (Goodrich et al. 2001), we assume $z_{\rm em}^{\rm QSO} = 5.65$ and we discuss this issue in further details in Appendix B.

The two QSOs have a projected separation of 109 arcsec, which corresponds to $R_\perp = 0.66$ Mpc. The LOS to QSO2 intersects the bubble produced by QSO1 for a redshift path ($\Delta z_{\rm prox}$) whose length depends on the radius of the H II region ($R_{\rm H\,II}$) itself. We find $R_{\rm H\,II} = 39$ Mpc, by plugging in equation (2) the following values: $t_{\rm Q} = 1.34 \times 10^7$ yr, $x_{\rm H\,I} = 8.4 \times 10^{-5}$, $\dot{N}_\gamma = 8.6 \times 10^{55}\,{\rm s}^{-1}$, where $x_{\rm H\,I}$ is provided by the mean value between those predicted by our models at $z = 5.7$ (see rightmost panel of Fig. 1), while $\dot{N}_\gamma$ is compatible with the luminosity of a QSO 3.5 mag fainter than QSO2 (Mahabal et al. 2005). In this section, we also take into account relativistic effects which could squash the ionization front along the LOS (White et al. 2003; Wyithe et al. 2005b; Yu 2005; Shapiro et al. 2006). The apparent size of the H II region, computed following the method outlined in Yu (2005), is shown in Fig. 5. By zooming the region close to QSO1 (small box in Fig. 5) it is clear that the apparent size of the H II region extends up to 2 Mpc in the direction towards QSO2. Given $R_{\rm H\,II}$, taking into account relativistic effects, the region $\Delta z_{\rm prox}$ extends from $z = 5.16$ up to 5.68. We recompute $x_{\rm H\,I}$ along the LOS to QSO2, adding to the uniform UVB photoionization rate $\Gamma_{\rm H\,I}$ the photoionization rate $\Gamma_{\rm H\,I}^{\rm QSO1}$ provided by QSO1, obtained starting from the following equations:

$$\Gamma_{\rm H\,I}^{\rm QSO1} = 4\pi \int_{\nu_{\rm H\,I}}^{\infty} \frac{J_\nu}{h\nu} \sigma_0 \left(\frac{\nu}{\nu_{\rm H\,I}}\right)^{-3} {\rm d}\nu, \tag{7}$$

$$J_\nu = \frac{\dot{N}_\nu h\nu}{16\pi^2 R^2}, \tag{8}$$

$$\dot{N}_\gamma = \int_{\nu_{\rm H\,I}}^{\infty} \dot{N}_\nu {\rm d}\nu = \int_{\nu_{\rm H\,I}}^{\infty} \dot{N}_{\nu_{\rm H\,I}} \left(\frac{\nu}{\nu_{\rm H\,I}}\right)^{-\alpha} {\rm d}\nu, \tag{9}$$

where $\nu_{\rm H\,I}$ is the hydrogen photoionization frequency threshold, σ_0 is the Thompson scattering cross-section, R is the distance from QSO1 to the LOS, $\dot{N}_{\nu_{\rm H\,I}}$ is the rate of the emitted ionizing photons at the hydrogen photoionization frequency threshold and $\alpha = 1.5$ is the spectral index of the QSO continuum. Integrating equation (9) we obtain

$$\dot{N}_{\nu_{\rm H\,I}} = \frac{(\alpha - 1)\dot{N}_\gamma}{\nu_{\rm H\,I}}. \tag{10}$$

Thus, it results

$$\Gamma_{\rm H\,I}^{\rm QSO1} = \left(\frac{\alpha - 1}{\alpha + 2}\right) \frac{\dot{N}_\gamma \sigma_0}{4\pi R^2}. \tag{11}$$

In Fig. 6 we compare the observed transmitted flux in the spectrum of QSO2 with the simulated fluxes along three different lines of sight with (bottom row) or without (top) including the contribution from QSO1 to the total ionizing flux. For brevity, we refer to these cases as 'with bubble' or 'without bubble'. Visual inspection of Fig. 6 shows that the case 'with bubble' is in better agreement with observations. Such statement can be made more quantitative by introducing a

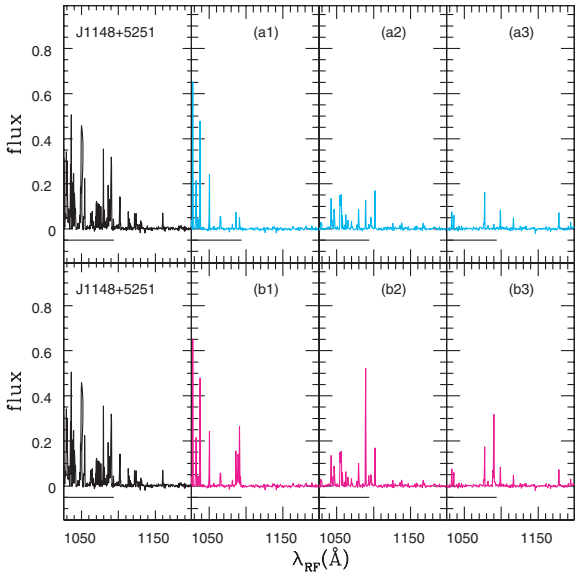

Figure 6. Leftmost panels: Observed transmitted flux (black spectra) in the spectrum of QSO SDSS J1148+5251 (QSO2, $z_{\rm em} = 6.42$). The solid black line shows the redshift path ($\Delta z_{\rm prox}$) in which the bubble produced by QSO RD J1148+5252 (QSO1, $z_{\rm em} = 5.65$) intersects the LOS to QSO2. Top panels (a*i*), with $i = 1, 3$: Simulated fluxes (cyan spectra) along three different random lines of sight (cases 'without bubble'). Bottom panels (b*i*), with $i = 1, 3$: Simulated fluxes (magenta spectra) along the same lines of sight shown in the top panels, taking into account the contribution from QSO1 to the total ionizing flux (cases 'with bubble').

quantity denoted peak spectral density (PSD), i.e. the number of peaks per unit $\lambda_{\rm RF}$ interval.

To compute the PSD for the two cases, we fix two different values for the flux threshold inside ($F_{\rm th}^{\rm IN} = 0.01$) and outside ($F_{\rm th}^{\rm OUT} = 0.08$) the bubble. While $F_{\rm th}^{\rm OUT}$ is the same as the value used in this work so far, $F_{\rm th}^{\rm IN}$ has been chosen accordingly to the maximum observed noise rms deviation (for reasons explained in Appendix A) in the $\lambda_{\rm RF}$ interval $\Delta\lambda = 1087$–1092 Å, where $\Gamma_{\rm H\,I}^{\rm QSO1} \gtrsim \Gamma_{\rm H\,I}$. For both the observed and simulated spectra, we compute the PSD inside and outside the bubble, finding the following results:

$$\left({\rm PSD}_{\rm obs}^{\rm OUT}, {\rm PSD}_{\rm obs}^{\rm IN}\right) = (0.11, 0.40),$$

$$\left({\rm PSD}_{\rm sim}^{\rm OUT}, {\rm PSD}_{\rm sim}^{\rm IN}\right) = \left(0.04^{+0.06}_{-0.04}, 0.24^{+0.35}_{-0.24}\right),$$

where error bars provide the maximum and minimum PSD values found in the simulated LOS. Observationally, the PSD is found to be approximately four times[6] larger inside that bubble than outside it. This boost is quite well reproduced by the simulated PSD, although their absolute values are somewhat lower than the observed ones.

The physical interpretation of the results reported in this section is the following. In the $\lambda_{\rm RF}$ (redshift) interval $\Delta\lambda = 1087$–1092 ($\Delta z_{\rm det} = 5.63$–$5.67$), where $\Gamma_{\rm H\,I}^{\rm QSO1} \gtrsim \Gamma_{\rm H\,I}$, most of the gaps present in the case 'without bubble' disappear, making room for peaks, as a consequence of the decreased opacity in the proximity of QSO1. Note that the $\Gamma_{\rm H\,I}$ value adopted in our calculations (shown in Fig. 1, middle panel, red line) is close to the maximum value suggested by previous studies. Moreover, $\Gamma_{\rm H\,I}^{\rm QSO1} \approx 3.4\Gamma_{\rm H\,I}$ at 0.66 Mpc from the foreground QSO. Thus, an implausible $\Gamma_{\rm H\,I}$ value should be assumed to explain the observed boost in the PSD with a uniform UVB. The enhancement in the transmissivity decreases for $\lambda_{\rm RF}$ smaller (larger) than 1087 (1092) Å, since at the corresponding redshift $\Gamma_{\rm H\,I}^{\rm QSO1} \lesssim \Gamma_{\rm H\,I}$. These results (i) confirm the detection of a proximity effect, (ii) show that the redshift stretch affected by the proximity effect is $\Delta z_{\rm det} < \Delta z_{\rm prox}$.

As a final test for our model, we compute the observed evolution of the optical depth as a function of the distance R from QSO1 and compare it with the predictions of model 'with bubble'; the result is shown in Fig. 7. The agreement between observations and simulations is at 1σ confidence level for 86 per cent of the plotted points. For $R \lesssim 2$ Mpc, the mean optical depth $1.5 \lesssim \bar{\tau} \lesssim 3.5$ is lower than the mean value expected at $\bar{z} = 5.65 (\bar{\tau}_{5.65} \approx 4)$; it approaches $\bar{\tau}_{5.65}$ at distances larger than $R_\tau \sim 2$ Mpc. By taking the difference between R_τ and $R_\perp$, we set a lower limit on the foreground QSO lifetime $t_{\rm Q} > (R_\tau - R_\perp)/c + (t_\tau - t_{\rm QSO1}) \approx 11$ Myr, where t_τ and $t_{\rm QSO1}$ represent the cosmic times corresponding to the redshifts $z_\tau = 5.68$ and $z_{\rm em}^{\rm QSO1} = 5.65$, respectively.

It is worth noting that our model does not take into account either (i) the clustering of the ionizing sources or (ii) the overdense environment expected around the QSO. Both these effects, in principle, could strongly affect the IGM ionization state, albeit in opposite ways. While clustering of sources would enhance the transmissivity in the QSO near-zones, the overdense environment would tend to suppress it. The fact that we found agreement between observations and our modelling could indicate that, at least along this LOS, the two effects compensate. For what concerns (ii), by comparing the optical depth evolution observed in the proximity regions of 45 QSOs at $z_{\rm em} \gtrsim 4$ with theoretical expectations, Guimaraes et al. (2007) find evidence for a density bias correlated with the QSO luminosity. Since QSO1 is much fainter than the QSOs studied by Guimaraes et al. (2007) it seems likely that neglecting such effect does not introduce a significant error. However, the extension of the proposed approach to a larger sample could clarify the relation between the clustering of sources and the overdensities in which massive objects are likely to be embedded.

[6] This factor depends on the flux threshold used. For example, it is reduced to ≈2.7 if $F_{\rm th}^{\rm IN} = F_{\rm th}^{\rm OUT} = 0.05$. Nevertheless, for the purpose of our test what really matters is the boost of this factor moving from outside towards inside the bubble.

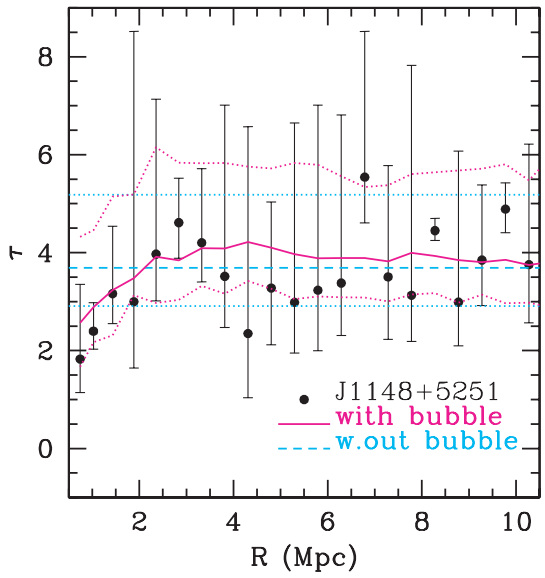

Figure 7. Evolution of the optical depth τ as a function of the distance R from QSO1. Filled circles denote the observed mean value for τ, while error bars represent the maximum and the minimum observed τ at a given distance from the foreground QSO. Solid (dotted) magenta lines are the mean (maximum/minimum) values from 500 simulated lines of sight, computed adopting the case 'with bubble'. The dashed cyan horizontal line shows the mean optical depth predicted by the ERM in correspondence of the emission redshift of the foreground QSO. The dotted cyan horizontal lines denote the maximum/minimum optical depth at the same redshift.

It is important to note that the LOS towards SDSSJ1148+5251 contributes to the LPW distribution with the smallest peak ($P_{\rm max} \approx 2$ Å) in the entire sample. Thus, even if we succeeded in reproducing the features of this LOS with our model 'with bubble', we still have to explain the mysterious origin of transmissivity windows as large as 10–15 Å.

In Section 4.1, we estimate the QSO1 luminosity required to explain the observed $P_{\rm max}$ value, and we comment on the result dependence from $f_{\rm t}$. Plugging in equation (4) the value $f_{\rm t} \approx 0.03$ computed inside the proximity region, we obtain an effective size for $R_{\rm H\,II}$; by further using equation (2), this translates into $\dot{N}_\gamma = 9.2 \times 10^{55} {\rm s}^{-1}$, a value in quite good agreement with the QSO1 ionizing rate quoted by Mahabal et al. (2005).

6 DISCUSSION

We have studied several statistical properties of the transmitted flux in high-z QSO spectra and compared them with those obtained from simulated Lyα forest spectra to infer constraints on the ionization state of the IGM at $z \approx 6$. We have considered two different reionization models: (i) an ERM, in which the universe reionizes at $z_{\rm rei} = 6$ and (ii) an LRM ($z_{\rm rei} \approx 7$).

By first using standard control statistics (mean transmitted flux evolution, PDF of the transmitted flux, GW distribution) in the

redshift range $3.5 < z < 6$, we show that both ERM and LRM match the observational data. This implies that current observations do not exclude that reionization can have taken place at redshift well beyond 6.

We then apply the LGW and LPW statistics introduced by GCF06 to a sample of 17 QSOs in the redshift range 5.74–6.42. Both ERM and LRM provide good fits to the observed LGW distribution, favouring a scenario in which $x_{\rm H\,I}$ smoothly evolves from $10^{-4.4}$ at $z \approx 5.3$ to $10^{-4.2}$ at $z \approx 5.6$.

Discriminating among the two reionization scenarios would require a sample of QSO at even higher redshifts. In fact, although according to LRM at $z \gtrsim 6$ the reionization process is still in the overlap phase with a mixture of ionized and neutral regions characterizing the IGM, only ≈10 per cent of the simulated LOS pierce the overlap epoch, and for a redshift depth $\Delta z \lesssim 0.2$. This explains why the predicted LGW distributions are quite similar for the two models considered.

None the less, ERM provides a slightly better fit to observational data with respect to LRM, favouring $z_{\rm rei} \gtrsim 7$. Within the statistical relevance of our sample, we have shown that LRM models can be used to put a robust upper limit $x_{\rm H\,I} < 0.36$ at $z = 6.3$.

We have suggested that peaks preferentially arise from underdense regions of the cosmic density field and also from isolated H II regions produced by either faint quasars or galaxies. The frequency of the observed peaks implies that the dark matter haloes hosting such sources is relatively large, $\approx 10^{12}$ $(10^{13})\,{\rm M}_{\odot}$. Bright QSOs are unlikely to contribute significantly in terms of peaks, because given the required size of the H II regions, they should be located close enough to the LOS to the target QSO, so that they should be detectable in the field.

The previous conclusions are substantiated by the specific case of an intervening H II region produced by the faint quasar RD J1148+5253 (QSO1) at $z = 5.70$ along the LOS towards the highest redshift quasar currently known (SDSS J1148+5251, QSO2) at $z = 6.42$. It is worth noting that searches for the transverse proximity effect in the H I Lyα forest at $z \approx 3$ (Schirber, Miralda-Escudé & McDonald 2004) have been so far unsuccessful. Such effect has been isolated only by He II absorption studies (Worseck & Wisotzki 2006; Worseck et al. 2007). Thus, our results represent the first-ever detection in the H I Lyα forest. We have analysed the proximity effect of QSO1 on the QSO2 spectrum. Moreover, we have build up a simple model to estimate the location/extension of the proximity zone. Within the proximity region of QSO1 we have found an increased number of peaks per unit frequency with respect to segments of the LOS located outside the quasar H II bubble. This supports the idea that we are indeed sampling the proximity region of the QSO1 and that at least some peaks originate within ionized regions around (faint) sources. We then obtain a strong lower limit on the foreground QSO lifetime of $t_{\rm Q} > 11$ Myr. Proper inclusion of galaxy clustering, which requires numerical simulations, might affect our conclusions (Faucher-Giguere et al. 2008). Note that even in this clear-cut case, the size of the largest observed peak in the spectrum of QSO2 is only of 2 Å.

Thus we are left with the puzzling discrepancy between observed and simulated transmissivity windows (peaks) size, the former being systematically larger. Very likely, this reflects an unwarranted assumption made by the model. We do not believe that the discrepancy could be imputed to the assumption of an LN model, tested against HYDROPM simulations by GCF06. Nevertheless, we plan to compare the observed LPW distribution with full hydrodynamical simulations in a future work to study the correlation properties of the underdense regions, since Coles, Melott & Shandarin (1993) have shown that the LN model produces a too 'clumpy' distribution of the density field, when compared with *N*-body simulations.

At least two physical effects, neglected here, could affect the calculation of $x_{\rm H\,I}$: (i) non-equilibrium photoionization and (ii) UV background radiation fluctuations.

The first assumption is made by the majority of studies dealing with the Lyα forest. However, if a fraction of the Lyα forest gas has been shock-heated as it condenses into the cosmic web filaments, it might cool faster than it recombines. For example, the recombination time $t_{\rm r}$ becomes longer than the Hubble time when the density contrast is $\Delta < 7.5\,[(1+z)/6.5]^{-3/2}$; hence, large deviations from photoionization equilibrium are expected where $\Delta \ll 1$. Lower values of $x_{\rm H\,I}$ with respect to equilibrium are expected in such regions, as a result of the exceedingly slow recombination rates.

The second possible explanation for the too narrow simulated peaks might reside in radiative transfer effects, also neglected here. At $z \approx 6$ the increase in the mean GP optical depth is accompanied by an evident enhancement of the dispersion of this measurement which has been ascribed to spatial fluctuations of the UVB intensity near the end of reionization. A considerable (up to 10 per cent) scatter in the UVB H I photoionization rate is expected already at $z \approx 3$, as shown by Maselli & Ferrara (2005) through detailed radiative transfer calculations. The amplitude of such illumination fluctuations tend to increase with redshift because of the overall thickening of the forest. Although the observed dispersion in the mean GP optical depth may be compatible with a spatially uniform UVB (Liu et al. 2006; Lidz, Oh & Furlanetto 2006), it is likely that a proper radiative transfer treatment becomes mandatory at earlier times. Basically, the main effect of fluctuations is to break the dependence of the H I neutral fraction on density. This is readily understood by considering two perturbations with the same density contrast Δ. If the first is close to a luminous source it will have its $x_{\rm H\,I}$ depressed well below that of the second one located away from it. Thus, opacity fluctuations naturally arise. If it is so, then peaks of larger width could be produced if the density perturbation associated with it happens to be located in a region where the UVB intensity is higher than the mean.

ACKNOWLEDGMENTS

We thank R. Cen, B. Ciardi, D. Eisenstein, J. P. Ostriker and S. White for stimulating discussions. We are particularly grateful to Z. Haiman, A. Lidz and A. Maselli for enlightening comments on the manuscript. XF acknowledges support from NSF grant AST 03-07384, a Sloan Research Fellowship, a Packard Fellowship for Science and Engineering.

REFERENCES

Barkana R., Loeb A., 2001, Phys. Rep., 349, 125
Becker R. H. et al., 2001, AJ, 122, 2850
Becker G. D., Rauch M., Sargent W. L. W., 2007, ApJ, 662, 72
Bi H., 1993, ApJ, 405, 479
Bi H., Davidsen A. F., 1997, ApJ, 479, 523
Bi H. G., Börner G., Chu Y., 1992, A&A, 266, 1
Bolton J. S., Haehnelt M. G., 2007a, MNRAS, 374, 493
Bolton J. S., Haehnelt M. G., 2007b, MNRAS, 382, 325 (B07)
Bolton J. S., Haehnelt M. G., Viel M., Springel V., 2005, MNRAS, 357, 1178 (B05)
Choudhury T. R., Ferrara A., 2005, MNRAS, 361, 577
Choudhury T. R., Ferrara A., 2006, MNRAS, 371, L55 (CF06)
Choudhury T. R., Padmanabhan T., Srianand R., 2001a, MNRAS, 322, 561
Choudhury T. R., Srianand R., Padmanabhan T., 2001b, ApJ, 559, 29

Coles P., Jones B., 1991, MNRAS, 248, 1
Coles P., Melott A. L., Shandarin S. F., 1993, MNRAS, 260, 765
Croft R. A. C., 1998, in Olinto A. V., Frieman J. A., Schramm, D. N. eds, Eighteenth Texas Symposium on Relativistic Astrophysics. World Scientific, River Edge, NJ, p. 664
Dijkstra M., Wyithe S., Haiman Z., 2007, MNRAS, 379, 253
Djorgovski S. G., Castro S., Stern D., Mahabal A. A., 2001, AJ, 560, L5
Fan X. et al., 2001, AJ, 122, 2833
Fan X., Narayanan V. K., Strauss M. A., White R. L., Becker R. H., Pentericci L., Rix H., 2002, AJ, 123, 1247
Fan X. et al., 2003, AJ, 125, 1649
Fan X. et al., 2006, AJ, 132, 117 (F06)
Fang T. et al., 2005, preprint (astro-ph/0505182)
Faucher-Giguere C. A., Lidz A., Zaldarriaga M., Hernquist L., 2008, ApJ, 673, 39
Gallerani S., Choudhury T. R., Ferrara A., 2006, MNRAS, 370, 1401 (GCF06)
Goodrich R. W. et al., 2001, ApJ, 561, L23
Guimaraes R., Petitjean P., Rollinde E., de Carvalho R. R., Djorgovski S. G., Srianand R., Aghaee A., Castro S., 2007, MNRAS, 377, 657
Haiman Z., Cen R., 2002, ApJ, 578, 702
Iliev I. T., Mellema G., Shapiro P. R., Pen U.-L., 2007, MNRAS, 376, 534
Iye M. et al., 2006, Nat, 443, 186
Kashikawa N. et al., 2006, ApJ, 648, 7
Kogut A. et al., 2003, ApJS, 148, 161
Kramer R. H., Hiaman Z., Oh S. P., 2006, ApJ, 649, 570
Lidz A., Oh S. P., Furlanetto S., 2006, ApJ, 639, L47
Lidz A., McQuinn M., Zaldarriaga M., Hernquist L., Dutta S., 2007, ApJ, 670, 39
Liu J., Bi H., Feng L., Fang L., 2006, ApJ, 645, L1
Mahabal A., Stern D., Bogosavljevic M., Djorgovski S. G., Thompson D., 2005, ApJ, 634, L9
Maiolino R. et al., 2004, Nat, 431, 533
Malhotra S., Rhoads J. E., 2006, ApJ, 647, L95
Maselli A., Ferrara A., 2005, MNRAS, 364, 1429
Maselli A., Gallerani S., Ferrara A., Choudhury T. R., 2007, MNRAS, 376, 34
McDonald P., Miralda-Escudé J., 2001, ApJ, 549, L11 (MM01)
Mesinger A., Haiman Z., 2004, ApJ, 611, L69
Mesinger A., Haiman Z., 2007, ApJ, 660, 923
Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, ApJ, 530, 1 (MHR)
Mo H. J., White S. D. M., 2002, MNRAS, 336, 112
Oh S. P., Furlanetto S. R., 2005, ApJ, 620, L9
Page L. et al., 2007, ApJS, 170, 335
Paschos P., Norman M. L., 2005, ApJ, 631, 59
Schirber M., Miralda-Escudé J., McDonald P., 2004, AJ, 610, 105
Shapiro P. R., Iliev I. T., Alvarez M. A., Scannapieco E., 2006, ApJ, 648, 922
Songaila A., 2004, AJ, 127, 2598
Songaila A., Cowie L. L., 2002, AJ, 123, 2183
Spergel D. N. et al., 2003, ApJS, 148, 175
Spergel D. N. et al., 2007, ApJS, 170, 377
Stark D. P., Ellis R. S., Richard J., Kneib J. P., Smith G. P., Santos M. R., 2007, ApJ, 663, 10
Stern D., Yost S. A., Eckart M. E., Harrison F. A., Helfand D. J., Djorgovski S. G., Malhotra S., Rhoads J. E., 2005, ApJ, 619, 12
Stiavelli M. et al., 2005, A&AS, 37, 1329
Totani, Kawai N., Kosugi G., Aoki K., Yamada T., Iye M., Ohta K., Hattori T., 2006, PASJ, 58, 485
Viel M., Matarrese S., Mo H. J., Haehnelt M. G., Theuns T., 2002, MNRAS, 329, 848
Viel M., Haehnelt M. G., Lewis A., 2006, MNRAS, 370, L51
White R. L., Becker R. H., Fan X., Strauss M. A., 2003, AJ, 126, 1
Wyithe J. S. B., Loeb A., 2004, Nat, 432, 194
Wyithe J. S. B., Loeb A., Carilli C., 2005a, ApJ, 628, 575
Wyithe J. S. B., Loeb A., Barnes D. G., 2005b, ApJ, 634, 715
Worseck G., Wisotzki L., 2006, A&A, 450, 495
Worseck G., Fechner C., Wisotzki L., Dall'Aglio A., 2007, A&A, 473, 805
Yu Q., 2005, ApJ, 623, 683
Yu Q., Lu Y., 2005, ApJ, 620, 31

APPENDIX A: ON THE FLUX THRESHOLD CHOICE

The gap/peak statistics are sensitive to the S/N, since spurious peaks could arise in spectral regions with noise higher than the flux threshold ($F_{\rm th}$) adopted. In what follows we restrict our attention to 'gaps', since the extension of the conclusions on the 'peaks' is direct. In particular, in this appendix, we discuss the LGW distribution shape dependence on the $F_{\rm th}$ chosen. We consider two values for $F_{\rm th}$, namely 0.03 and 0.08, which correspond to $\tau = 3.5$ and 2.5, respectively. It is not obvious what criterion to apply in order to choose a proper value for $F_{\rm th}$, since a too high (low) $F_{\rm th}$ could overestimate (underestimate) the gap length. In Fig. A1 we show two examples of spectra in which the $F_{\rm th}$ choice strongly affects the gap measurement. On the bottom, the spectrum of QSO J1030+0524 is shown, and, in the small box, the region marked by the solid black line is zoomed. It is evident that $F_{\rm th} = 0.03$ would break the gap at $\lambda_{\rm RF} \approx 1190$ Å, instead of at $\lambda_{\rm RF} \approx 1160$ Å, as also noted by F06; thus, in this case $F_{\rm th} = 0.08$ seems to be a better choice.

The opposite is true for the spectrum of QSO J1148+5251, shown in Fig. A1 on the top. Indeed, in this case $F_{\rm th} = 0.08$ would provide a gap as large as ≈100 Å, terminated by transmission at $\lambda_{\rm RF} \approx$ 1100 Å. However, the peak at $\lambda_{\rm RF} \approx 1160$ Å is consistent with pure transmission (White et al. 2003; Oh & Furlanetto 2005; F06); thus, in this case, $F_{\rm th} = 0.03$ would provide the correct gap measurement. For this reason, we compute the LGW distribution, considering both $F_{\rm th} = 0.03$ and 0.08, alternatively.

The final LGW distribution is obtained as the mean of the preliminary ones, weighted on the corresponding errors.

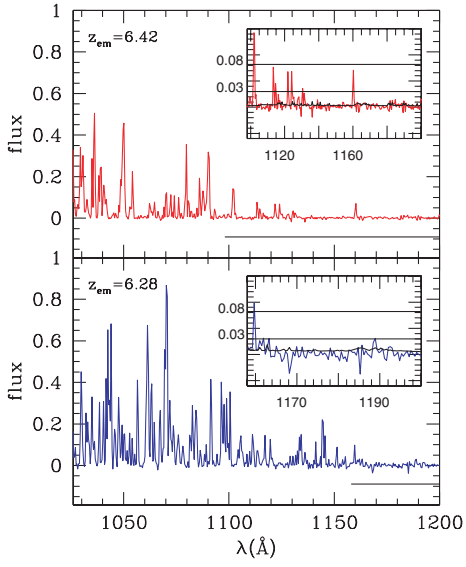

Figure A1. Observed spectra of the QSO SDSS J1148+5251 (top panel) and SDSS J1030+0524 (bottom panel). The black line denote the largest dark gap, measured by assuming $F_{\rm th} = 0.08$. In the small box the region interested by the largest dark gap is zoomed. The two black lines indicate $F_{\rm th} = 0.08$ and 0.03. From the top (bottom) panel is evident that $F_{\rm th} = 0.08$ ($F_{\rm th} = 0.03$) overestimates (underestimates) the size of the largest dark gap.

APPENDIX B: ON THE FOREGROUND QSO REDSHIFT

The redshift quoted by Mahabal et al. (2005) for RD J1148+5253 ($z_{\rm em} = 5.7$) is based on the peak of the Lyα emission line. This standard approach typically overestimates the true redshift by $\approx$0.05 (e.g. Goodrich et al. 2001). For this reason we adopt as fiducial value $z_{\rm em} = 5.65$. As the object RD J1148+5253 (QSO1) is a BAL QSO (Mahabal et al. 2005), its emission redshift cannot be established with accuracy from the broad metal lines, thus remaining uncertain. By comparing the QSO1 absorption spectrum with a BAL composite, also $z_{\rm em} = 5.725$ could be a plausible choice for the QSO1 emission redshift (C. Willot, private communication). In this appendix we repeat the analysis shown in Section 5 considering different possibilities for the QSO1 emission redshift. In Fig. B1 we compare the optical depth evolution as a function of the distance from QSO1 obtained assuming $z_{\rm em}^{\rm QSO1} = 5.65$ (left-hand panel) with the cases in which $z_{\rm em}^{\rm QSO1} = 5.70$ (middle panel) and $z_{\rm em}^{\rm QSO1} = 5.725$ (right-hand panel). In Table B1 the results of the PSD for the three different choices of QS01 emission redshift are shown, together with the wavelength interval $\Delta\lambda$ where $\Gamma_{\rm H\,I}^{\rm QSO1} > \Gamma_{\rm H\,I}$.

Table B1. Results of the PSD for the three different choices of QS01 emission redshift.

$z_{\rm QSO1}$	PSD $_{\rm OUT}$	PSD $_{\rm IN}$	$\Delta\lambda$
5.65	0.11	0.40	1087–1092
5.70	0.12	0.60	1095–1100
5.725	0.11	0.40	1099–1104

Even though Fig. B1 shows that the observed optical depth evolution as a function of the distance from QSO1 is better explained by our model assuming $z_{\rm em}^{\rm QSO1} = 5.65$, this result does not rule out other choices of the QSO1 emission redshift. The results shown in Table 8 confirm the evidence of the transverse proximity effect, since the boost in the PSD moving from outside towards inside the bubble is present in all the three cases considered.

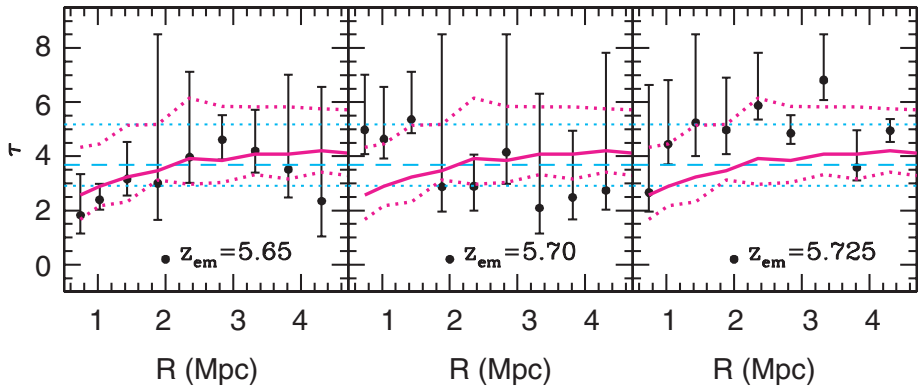

Figure B1. Same as Fig. 7, but for different choices of foreground QSO emission redshift: $z_{\rm em} = 5.65$ (leftmost panel); $z_{\rm em} = 5.70$ (middle panel); $z_{\rm em} = 5.725$ (rightmost panel).

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