The role of galaxies and AGN in reionizing the IGM – III. IGM–galaxy cross-correlations at $z \sim 6$ from eight quasar fields with DEIMOS and *MUSE*

Romain A. Meyer [●],[1][★] Koki Kakiichi,[1,2] Sarah E. I. Bosman,[1] Richard S. Ellis,[1] Nicolas Laporte,[1,3,4] Brant E. Robertson,[5] Emma V. Ryan-Weber,[6,7] Ken Mawatari[8,9] and Adi Zitrin [●][10]

[1]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[2]*Department of Physics, University of California, Santa Barbara, CA 93106, USA*
[3]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[4]*Cavendish Laboratory, University of Cambridge, 19 JJ Thomson Avenue, Cambridge CB3 0HE, UK*
[5]*Department of Astronomy and Astrophysics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA*
[6]*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, VIC 3122, Australia*
[7]*ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Canberra, ACT 2601, Australia*
[8]*Department of Environmental Science and Technology, Faculty of Design Technology, Osaka Sangyo University, 3-1-1, Nagaito, Daito, Osaka 574-8530, Japan*
[9]*Institute for Cosmic Ray Research, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8583, Japan*
[10]*Physics Department, Ben-Gurion University of the Negev, PO Box 653, Be'er-Sheva 8410501, Israel*

ABSTRACT

We present improved results of the measurement of the correlation between galaxies and the intergalactic medium transmission at the end of reionization. We have gathered a sample of 13 spectroscopically confirmed Lyman-break galaxies (LBGs) and 21 Lyman-α emitters (LAEs) at angular separations 20 arcsec $\lesssim \theta \lesssim$ 10 arcmin ($\sim$0.1–4 pMpc at $z \sim 6$) from the sightlines to eight background $z \gtrsim 6$ quasars. We report for the first time the detection of an excess of Lyman-α transmission spikes at $\sim$10–60 cMpc from LAEs (3.2σ) and LBGs (1.9σ). We interpret the data with an improved model of the galaxy–Lyman-α transmission and two-point cross-correlations, which includes the enhanced photoionization due to clustered faint sources, enhanced gas densities around the central bright objects and spatial variations of the mean free path. The observed LAE(LBG)–Lyman-α transmission spike two-point cross-correlation function (2PCCF) constrains the luminosity-averaged escape fraction of all galaxies contributing to reionization to $\langle f_{\rm esc} \rangle_{M_{\rm UV} < -12} = 0.14^{+0.28}_{-0.05} (0.23^{+0.46}_{-0.12})$. We investigate if the 2PCCF measurement can determine whether bright or faint galaxies are the dominant contributors to reionization. Our results show that a contribution from faint galaxies ($M_{\rm UV} > -20 (2\sigma)$) is necessary to reproduce the observed 2PCCF and that reionization might be driven by different sub-populations around LBGs and LAEs at $z \sim 6$.

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

1 INTRODUCTION

Understanding cosmic reionization is of prime importance for cosmology and galaxy evolution. The key open questions are the timing of cosmic hydrogen reionization and the nature of the sources of ionizing photons. The timing of reionization is now constrained

to the redshift range $6 \lesssim z \lesssim 15$ (Planck Collaboration VI 2018). A large set of observational probes such as the Lyman-α forest opacity (e.g. Becker et al. 2001; Fan et al. 2002, 2006; Bosman et al. 2018; Eilers, Davies & Hennawi 2018), the decline of the fraction of Lyman-α emitters (LAEs) (e.g. Stark et al. 2010; Stark, Ellis & Ouchi 2011; Ono et al. 2012; Schenker et al. 2012; De Barros et al. 2017; Mason et al. 2018a; Pentericci et al. 2018), the number of 'dark pixels' in the Lyman-α forests at $z \sim 6$ (Mesinger 2010; McGreer, Mesinger & D'Odorico 2015), and the damping wing of

[★] E-mail: r.meyer.17@ucl.ac.uk

quasars (e.g. Bañados et al. 2018) seem to favour a late, rapid, and patchy reionization process down to $z \sim 5.5$–6 (Greig & Mesinger 2017). In comparison, the nature of the sources has remained more elusive. Galaxies are thought to provide the bulk of the ionizing photon budget necessary to drive reionization (e.g. Robertson et al. 2015, for a review) and the contribution of active galactic nuclei (AGNs), although still debated, thought to be relatively minor (e.g. Giallongo et al. 2015; Parsa, Dunlop & McLure 2018; Kulkarni, Worseck & Hennawi 2019).

None the less, the properties of the galaxies driving reionization and their relative contributions are less clear. Although the existence of a large population of galaxies down to $M_{\rm UV} \lesssim -16$ is now established up to redshift of $z \sim 10$ (e.g. Bouwens et al. 2015; Livermore, Finkelstein & Lotz 2017; Oesch et al. 2018), a fundamental issue is the challenge of measuring the escape fraction $f_{\rm esc}$ and production efficiencies of ionizing photons $\xi_{\rm ion}$,[1] which determine their contribution to the total photoionization budget. At high redshift, matching the total ionization budget to the neutral fraction evolution requires a reionization process driven by faint galaxies ($M_{\rm UV} \lesssim -10$) with a moderate escape fraction $f_{\rm esc} \gtrsim 10$–20 per cent and standard Lyman continuum (LyC) photon production efficiencies $\log \xi_{\rm ion}/[{\rm erg}^{-1}{\rm Hz}] \simeq 25.2$–$25.5$ (Ouchi et al. 2009; Robertson et al. 2013, 2015; Matthee et al. 2017; Nakajima et al. 2018). However, the proposed picture is dependent on whether low-mass faint galaxies do contribute significantly to the reionization of the surrounding intergalactic medium (IGM). Alternatives are possible if massive and efficiently LyC-leaking galaxies and AGNs dominate late reionization (e.g. Naidu et al. 2020). Direct measurements of the ionizing parameters of galaxies have proven challenging and offer a fractured picture. Spectroscopy of the afterglow of gamma-ray bursts indicates extremely low escape fractions $\lesssim 2$ per cent for their hosts (Tanvir et al. 2019), whereas the peak separation of Lyman α in a bright LAE (COLA1) results in an indirect measurement of $f_{\rm esc} \sim 15$–30 per cent (Matthee et al. 2018). Individual galaxies at high redshift have been found to have hard radiation spectra (e.g. Stark et al. 2017; Mainali et al. 2018) comparable to strong LAEs or high [O III]/[O II] emitters at $z \sim 3$ (Nakajima et al. 2016, 2018; Tang et al. 2019). At $z \lesssim 4$ where the escape fraction is directly measurable, it is found to be highly varying in individual objects but on average is around the ~ 5–10 per cent required for reionization (e.g. Shapley et al. 2006; Izotov et al. 2016, 2018; Vanzella et al. 2016, 2017; Steidel et al. 2018; Fletcher et al. 2019).

In this series, we have sought to address the issue of the ensemble contribution of sub-luminous galaxies to reionization. We have presented in Kakiichi et al. (2018, henceforth Paper I), a new approach to uncover the contribution of faint sources by the detection and modelling of the statistical H I proximity effect due to faint galaxies clustered around bright LBGs at $z \sim 6$ in cosmic volumes probed in absorption with quasar spectra. In Meyer et al. (2019, henceforth Paper II), we extended this framework to enable us to correlate metal absorbers, considered to be hosted by sub-luminous LBGs, with the IGM transmission measured in the Lyman-α forest of quasars. Paper I was a pilot study that analysed only one quasar sightline, which raised the question of the statistical significance of the tantalizing proximity effect detected. Indeed, it was shown in Paper II that cosmic variance between sightlines is an important factor also noted independently in simulations (Garaldi, Compostella & Porciani 2019). Though

Paper II overcame this issue by sampling C IV absorbers at $4.5 < z < 6$ in 25 quasar sightlines and detected an excess of transmission around C IV absorbers, they were a proxy for observed starlight from galaxies and raised the questions of the nature of the C IV hosts. None the less, both studies suggested that the faint population of galaxies at $z \sim 6$ has a high ensemble-averaged escape fraction and/or ionizing efficiencies required to sustain reionization (e.g. Robertson et al. 2015). In this third paper of the series, we present an improved study of the correlation between galaxies and the IGM at the end of reionization and the resulting inference on the ionizing capabilities of the sub-luminous population. We have gathered an extensive data set of galaxies in the fields of eight quasars at $z > 6$ through an additional observational campaign with DEep Imaging Multi-Object Spectrograph (DEIMOS)/Keck as well as archival Multi Unit Spectroscopic Explorer (*MUSE*)/VLT data. Moreover, we have extended the analytical framework to include the effect of gas overdensities in addition to the enhanced ultraviolet background (UVB) caused by clustered faint galaxies, spatially varying mean free path and forward modelling of peculiar velocities and observed flux uncertainties. Finally, we present a new probe of the galaxy–IGM connection by measuring and modelling the two-point cross-correlation function (2PCCF) between galaxies and transmission spikes in the Lyman-α forest of background quasars.

The plan of this paper is as follows. Section 2 introduces our new data set, starting with the eight high-redshift quasar spectra used in this study. Section 2.2 presents DEIMOS/Keck spectroscopic data of Lyman-break galaxies (LBGs) in three quasar fields with multislit spectroscopy. Section 2.3 details our data set drawn from *MUSE* archival observations, and our search for LAEs in the Integral Field Unit (IFU) datacubes. In Section 3, we present the galaxies detected in the field of background quasars with redshifts overlapping with the IGM probed by the Lyman-α forest, and the cross-correlations of galaxies with the surrounding IGM are detailed in Section 4. Section 5 presents an extension to the analytical framework of Paper I necessary to interpret our new measurements. The final results and constraints on the ionizing capabilities of galaxies at the end of reionization are presented in Section 6. We discuss the use of our measurement to differentiate between the relative contributions of faint and bright galaxies to reionization and the difference between the cross-correlation statistics in Section 7 before concluding in Section 8.

Throughout this paper, the magnitudes are quoted in the AB system (Oke, 1974), we refer to proper (co-moving) kiloparsecs as pkpc (ckpc) and megaparsecs as pMpc (cMpc), assuming a concordance cosmology with $H_0 = 70 \ {\rm km \ s^{-1} \ Mpc^{-1}}$, $\Omega_{\rm M} = 0.3$, and $\Omega_\Lambda = 0.7$.

2 OBSERVATIONS

tThis series focuses on the measurement and modelling of the correlations between galaxies and the surrounding IGM at the end of reionization. In order to achieve that goal, we have gathered different data sets of luminous galaxies acting as signposts for overdensities of less luminous sources. We have continued the approach of Paper I by confirming high-redshift LBGs via their Lyman-α emission with the DEIMOS (DEIMOS, Faber et al. 2003) at Keck. For convenience, we refer to those as LBGs because of their selection technique (although formally they are all also LAEs given they were confirmed with this line). Throughout the paper, we thus reserve the terminology LAE for galaxies detected blindly in archival IFU data of the *MUSE* (Bacon et al. 2010) at the VLT.

---

[1] Defined as the number of Lyman Continuum photons emitted per unit UV luminosity of the galaxy (e.g. Robertson et al. 2013).

Table 1. Observational data summary of the quasar fields.

Quasar	z	LAEs	LBGs	DEIMOS ID	*MUSE* ID	Total exptime	Ref	Spectrum	Refereces
J0305−3150	6.61	3	–	–	094.B–0893(A)	2^h30^m	(2)	XShooter	(3)
J0836+0054	5.81	–	1	U182	–	5^h10^m	(1)	XShooter	(5)
J1030+0524	6.28	7	8	C231, U182	095.A–0714(A)	$10^h50^m / 6^h40^m$	(1)/(4)	XShooter	(5)
J1148+5251	6.419	–	4	C231, U182	–	13^h10^m	(1, 6)	ESI	(5)
J1526−2050^a	6.586	2	–	–	099.A–0682(A)	3^h20^m	(7, 8)	XShooter	(9)
J2032−2114^b	6.24	3	–	–	099.A–0682(A)	5^h	(7, 8)	XShooter	(10)
J2100−1715	6.09	4	–	–	097.A–5054(A)	3^h40^m	(7, 8)	XShooter	(11)
J2329−0301	6.43	2	–	–	060.A–9321(A)	2^h	(7, 8)	ESI	(12)

References: (1) this work; (2) Farina et al. (2017); (3) Venemans et al. (2013); (4) Diaz et al. (2020); (5) McGreer et al. (2015); (6) Paper I; (7) Drake et al. (2019); (8) Farina et al. (2019); (9) Mazzucchelli et al. (2017); (10) Bañados et al. (2016); (11) Willott et al. (2007); and (12) Bosman et al. (2018). Fields: (1) quasar name; (2) quasar redshift; (3) number of suitable LAEs in the Lyman-α forest redshift range of the nearby quasar detected with *MUSE*; (4) number of suitable LBGs detected with DEIMOS; (5) DEIMOS Keck programme ID; (6) *MUSE* ESO programme ID; (7) total exposure time of DEIMOS/*MUSE* in the field; (8) reference of the original published paper on the observational programme; (9) instrument used for the quasar spectrum; (10) reference of the discovery paper of the quasar. ^{a}Also known as P231-20. ^{b}Also known as P308-21.

MUSE complements the early DEIMOS approach since we use a different selection method for galaxies and probe the smaller scales appropriate to the small *MUSE* field of view (FoV) (1 arcmin corresponding to ~360 pkpc at $z \sim 5.8$). These complementary data sets of galaxies were gathered in the field of $z \sim 6$ quasars with existing absorption spectroscopy of the Lyman-α forest, which probes the IGM transmission and ultimately enables us to compute the galaxy–IGM correlations. We now proceed to describe this rich observational data set, starting with the quasar spectra and moving then to the DEIMOS and *MUSE* data.

2.1 Quasar spectroscopic observations

The eight quasar fields studied in this work were chosen to have existing moderate or high signal-to-noise ratio (SNR) spectroscopy of the Lyman-α forests and either be accessible to Keck for DEIMOS follow-up or have archival *MUSE* data with adequate (≥ 2h) exposure time. The quasar spectra used in this study were downloaded from the European Southern Observatory (ESO) XShooter Archive or the Keck Observatory Archive (ESI). We use the same ESI spectrum of J1148, as in Paper I, originally observed by Eilers et al. (2017). The quasars already presented in Bosman et al. (2018, J0836, J1030) were reduced using a custom pipeline based on the standard ESOREX XShooter recipes, as detailed therein. The remaining quasars (J0305, J1526, J2032, J2100, and J2329) were reduced with the open-source reduction package PYPEIT^2 (Prochaska et al. 2019). The quasars have a median SNR of ~16 in the rest-frame ultraviolet (UV). The spectra were finally normalized by a power-law $f(\lambda) = A\lambda^b$ fitted to the portion of the continuum redwards of Lyman-α relatively devoid of broad emission lines (1270–1450 Å), as described in Bosman et al. (2018), to compute the transmission in the Lyman-α forest. Table 1 summarizes the quasar spectroscopic data information alongside the galaxy detections.

2.2 DEIMOS spectroscopy of LBGs in three quasar fields

As part of this study, we have re-observed the field of quasar J1148+5251 explored in Paper I to improve our selection of LBGs. As the slitmask design of DEIMOS does not allow small slit separations, only a selected subset of LBGs can be observed in any

Table 2. Summary of the DEIMOS observations.

Quasar	N_{LBG}	#$_{Mask}$	Exptime	Seeing
J0836+0054	4^a	K1	5^h10^m	0.7−0.9 arcsec
J1030+0524	3	K1	4^h00^m	0.9−1.5 arcsec
	3	K2	5^h20^m	0.7−0.9 arcsec
	2	K3	1^h30^m	0.7−0.9 arcsec
J1148+5251	4^b	K1^b	4^h30^m	0.7−1.5 arcsec
	0	K2	8^h40^m	0.7−0.9 arcsec

Notes: The masks J1030-K1, J1030-K2, and J1148-K1 were observed in 2017 March 26–27 (PI: A. Zitrin, ID: C231) and the remainder in 2018 March 7–8 (PI: B. Robertson, ID: U182). The number of confirmed LBGs is only weakly correlated to the total exposure time on the mask. ^{a}Three LBGs in J0836+0054 are in the near-zone of the quasar and hence they do not appear in Table 1 as they are not suited for our purposes. They will be studied in greater in Bosman et al. (2019). ^{b}We remove the faint AGN as well as the least convincing LBG detection (ID = 022) presented in Paper I to harmonize the LBG selection.

given mask. Accordingly, the detection of the proximity signal in Paper I might be affected by the sampling of candidate LBGs in the field. We also include data for two new quasar fields: J1030+0524 ($z = 6.3$) and J0836+0054 ($z = 5.8$) (Table 2).

Deep ground-based photometry of the three fields was used to construct catalogues of r' and i' drop-outs for follow-up. The fields of J1030 and J1148 have been observed in the SDSS r'-, i'-, z'-band filters with the Large Binocular Camera (LBC, Giallongo et al. 2008) at the Large Binocular Telescope (LBT, Hill & Salinari 2000), with the publicly available photometry reduced by Morselli et al. (2014).^{3} For the field of J0836, we used r'-, i'-, z'-band observations (Ajiki et al. 2006) taken with SuprimeCam on the 8.2-m Subaru Telescope (Kaifu et al. 2000; Miyazaki et al. 2002). We chose the following colour cuts to select potential $z \sim 5$–6 LBG candidates (see Fig. 1)

$$[r' - i' > 1.0] \wedge [i' - z' < 1.0] \wedge [z' < z'(3\sigma)] \quad (1)$$

for r'-drop-outs and

$$[i' - z' > 1.0] \wedge [r' > r'(2\sigma) \vee r' - z' > 1.75] \wedge [z' < z'(3\sigma)] \quad (2)$$

for i-drop-outs, where r', i', and z' indicates the magnitude in the corresponding SDSS band, and $r'(2\sigma)$, $z'(3\sigma)$ the limiting 2σ

^{2}https://github.com/pypeit/PypeIt

^{3}http://www.oabo.inaf.it/ LBTz6/1030/lbtz6.html
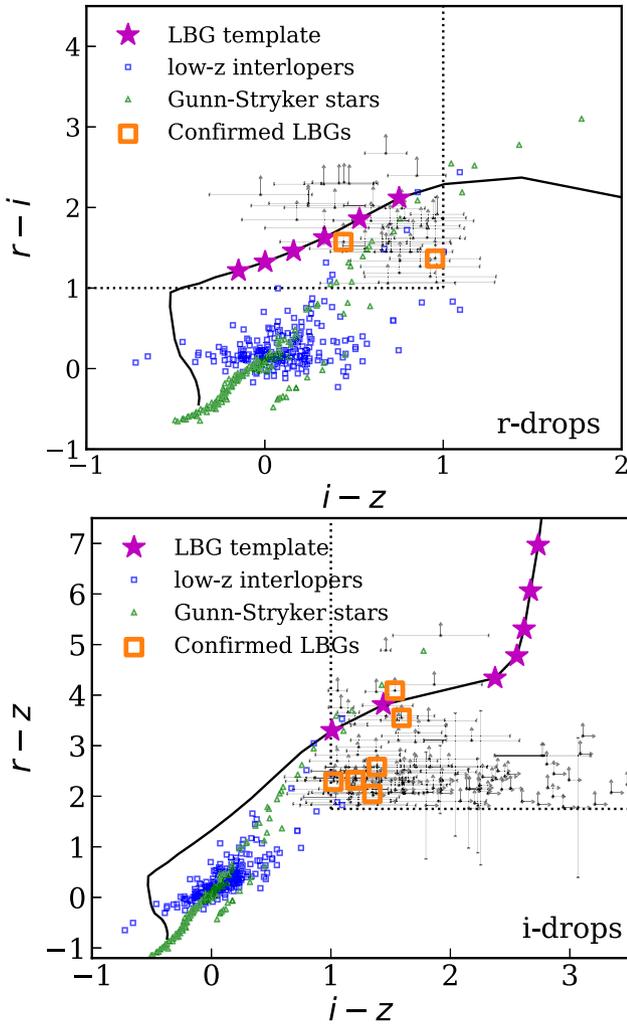
Figure 1. *r*-drop-out (top panel) and *i*-drop-out (bottom panel) selection of candidate LBGs in the fields of J0836, J1030, and J1148. An LBG template (magenta stars, black line) falls into the colour–colour cuts (dotted lines) at the redshift of interest $z \sim 5$–6. Galactic stars (green triangles, Gunn & Stryker 1983) and low-redshift interlopers (small blue squares, VUDS-DR1 samples from the COSMOS field, Le Fèvre et al. 2015; Tasca et al. 2017) are, however, mostly rejected. The target candidates are in shown in black dots, and confirmed LBGs are highlighted in the top-right corner with large orange squares. A colour version is available online.

and 3σ magnitude in the r'- and z'-band image, respectively. All candidates were visually inspected to produce a final catalogue. In designing the DEIMOS masks, we prioritized drop-outs based on the strength of their colour drop $[i' - z'(r' - i') > 1.0, 1.3, 1.5]$ or r'-band non-detection $[r' > r'(2\sigma, 3\sigma, 5\sigma)]$ to optimize the chance of confirmation. Priority was always given to better candidates first, and then to i'-drop-outs over r'-drop-outs of the same quality. The masks contained ~ 25–40 slits for science targets and five or six square holes for alignment stars.

The candidates were observed with the DEIMOS instrument (Faber et al. 2003) at the Keck II 10-m telescope during two observing runs in 2017 March 26–27 (PI: A. Zitrin) and 2018 March 7–8 (PI: B. Robertson). We confirmed 13 LBGs in the three fields over the course of four nights in good conditions with a seeing of 0.7–0.9 arcsec except for one night at 0.9–1.5 arcsec, as summarized in Table 2. The masks and the LBG detections are shown in the three

fields in Fig. 2. Surprisingly we could not confirm any new LBG in the field of J1148+5251, although Paper I confirmed 5(+ 1 AGN) in a smaller exposure time. The other masks in J0836 and J1030, exposed for $1^\mathrm{h}30^\mathrm{m}$ to $5^\mathrm{h}10^\mathrm{m}$, yielded two to four LBG confirmations each. The completeness of our search for LBGs in the relevant redshift range is not straightforward to estimate but fortunately not a major concern for our analysis. Whilst, in principle, the total number of galaxies in these fields can be estimated using the UV luminosity function (LF) and the depth of the photometric data, we find no proportionality with the observed numbers of LBGs. The scatter from field to field detailed above is therefore mainly driven by the random selection of objects on each mask ($\lesssim 20$ 'prime' candidates) and the number of mask observed for each given field. Indeed, the number of objects confirmed per mask is roughly constant, regardless of the redshift and field. However, this does not affect our results since we are aiming to measure the average Ly-α transmission around the detected bright galaxies only. As we cross-correlate them with the Ly-α forest and do not measure their number density, we do not need to correct for completeness (see further Section 4).

The data were reduced using the DEEP2 pipeline (Cooper et al. 2012; Newman et al. 2013), and the 1D spectra were extracted using optimal extraction with a 1.2-arcsec boxcar aperture (Horne 1986). The 2D spectra were inspected visually by five authors (RAM, KK, SEIB, RSE, and NL) for line emitters. Candidate LBGs were retained if they were found by three authors or more in the 2D spectra. We show examples of a clear LBG detection and a less convincing one in Fig. 3. The remaining LBG detections are presented individually in Appendix A. We also present serendipitous line emitters (without optical counterpart) that are not used in this study.

2.3 Archival *MUSE* quasar fields

We exploit six $z \sim 6$ quasar fields with deep ($\gtrsim 2$h) archival *MUSE* data to search for galaxies close to the sightline. The *MUSE* quasar fields are listed in Table 1.

We reduce the archival *MUSE* data using the *MUSE* v2.6 pipeline (Weilbacher et al. 2012, 2015) with the standard parameters. We further clean the skylines emission using the Zurich Atmospheric Purge code (Soto et al. 2016). After masking the bright sources and the edges of the data cubes, we run *MUSE*LET (Bacon et al. 2016) and LSDCAT (Herenz & Wisotzki 2017) to extract line emitters. We find that both algorithms are complementary due to their different search strategy. *MUSE*LET reduces the IFU cube to a series of narrow-band images (6.25 -Å width) and uses SEXTRACTOR to identify emission lines by subtracting a median continuum constructed from the adjacent wavelength planes. Detections in several narrow bands at similar positions can be grouped together to find a redshift solution. Whilst it is a robust technique, continuum absorptions or rapid continuum variations often produce spurious detections (when the adjacent narrow bands are subtracted). LSDCAT improves the removal of foreground continuum objects by utilizing a median filtering of the cube in the wavelength dimension. The emission lines are then detected with a three-dimensional (3D) matched-filtering approach. LSDCAT also allows one to mask brighter sources with custom masks. It, however, then fails to pick faint sources next to bright objects if the masking and/or the median filtering is too aggressive. Finally, the width of the narrow-bands in *MUSE*LET and the convolution kernel sizes of the matched-filter in LSDCAT can lead to different false positives or negatives. Therefore, we use both algorithms to generate a consolidated list of line emitter candidates
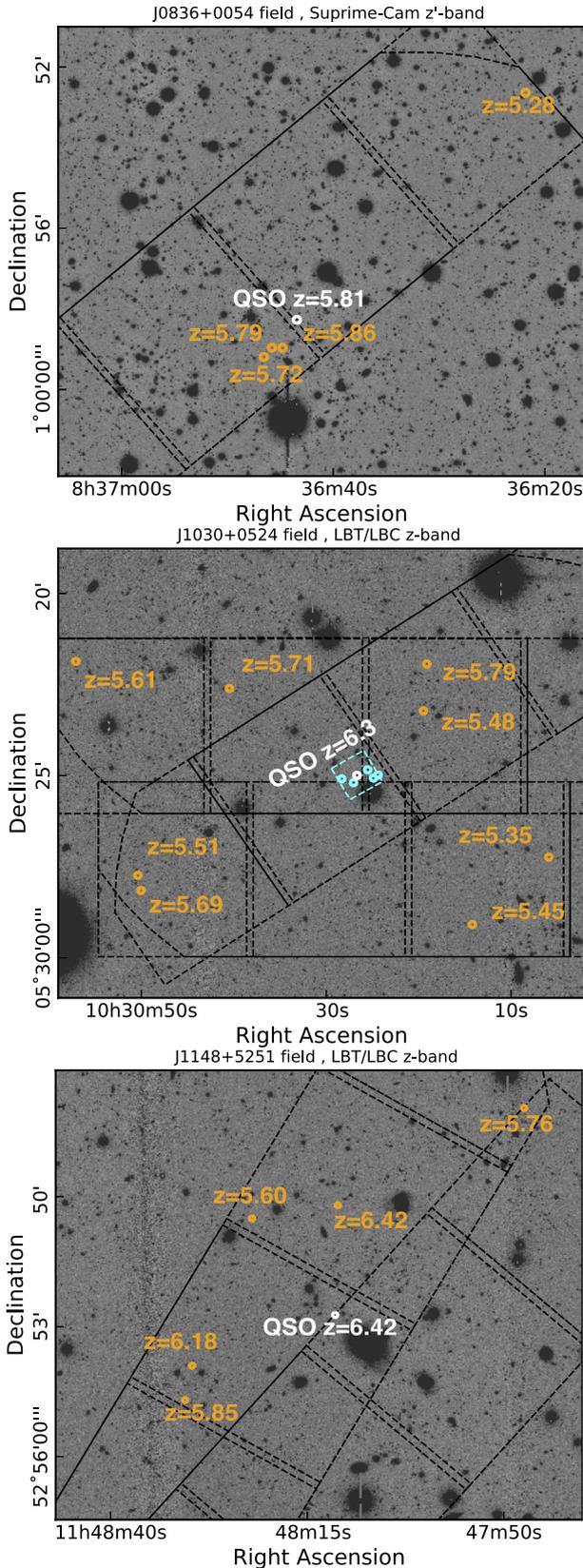
Figure 2. Spectroscopically confirmed high-redshift LBGs (orange squares) and potential LAEs (green circles) identified with DEIMOS in the quasar fields of J0836, J1030, and J1148. We overlay the DEIMOS slitmask FoV in black. For J1030 (middle panel), we also add the *MUSE* FoV (cyan-dashed square) and the LAEs detected in the *MUSE* datacube (cyan circles). A colour version is available online.
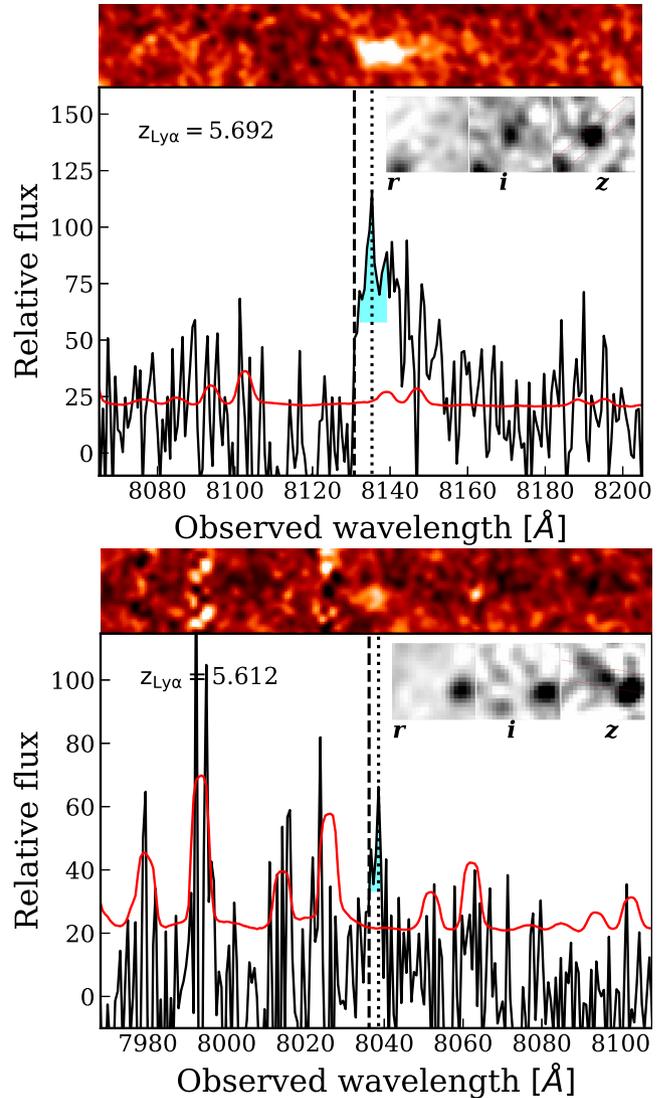
Figure 3. Examples of a clear and a marginal detection of Lyman-α emission for LBG-selected galaxies in the field of J1030. The top panels show the 2D spectra from which the spectrum (black line) and noise (red) are optimally extracted using a boxcar aperture of 1.2 arcsec. The peak of Lyman-α is identified with a dotted line, and the location of the systemic by a dashed line. The systemic redshift is found by applying a correction based on the full width at half-maximum (FWHM) of the Lyman-α emission (cyan, see Section 2.4). In the upper right-hand corner is displayed the *riz* photometry used for the drop-out selection. The remaining detections are presented in Appendix A. A colour version is available online.

that are then visually inspected to compile a robust sample of high-redshift LAEs. We check that candidates are present in the two datacubes produced with the two halves of the exposures to remove cosmic rays and other artifacts, and we remove double-peaked emissions, which are likely to be low-redshift [O II] $\lambda\lambda$ 3727 Å doublets, as it would mimic a double-peaked $z \sim 5$–6 Lyman α with a reasonable peak separation $\Delta v \sim 220$ km s^{-1}. Double-peaked Lyman-α profiles at $z > 5$ are exceedingly rare due to absorption by the IGM (Hu et al. 2016; Songaila et al. 2018), so we expect to lose very few high-redshift LAEs in being so conservative. Finally, we produce a white light image of the *MUSE* cube and check that the line detection is not caused by a poor continuum subtraction of a bright foreground object or a
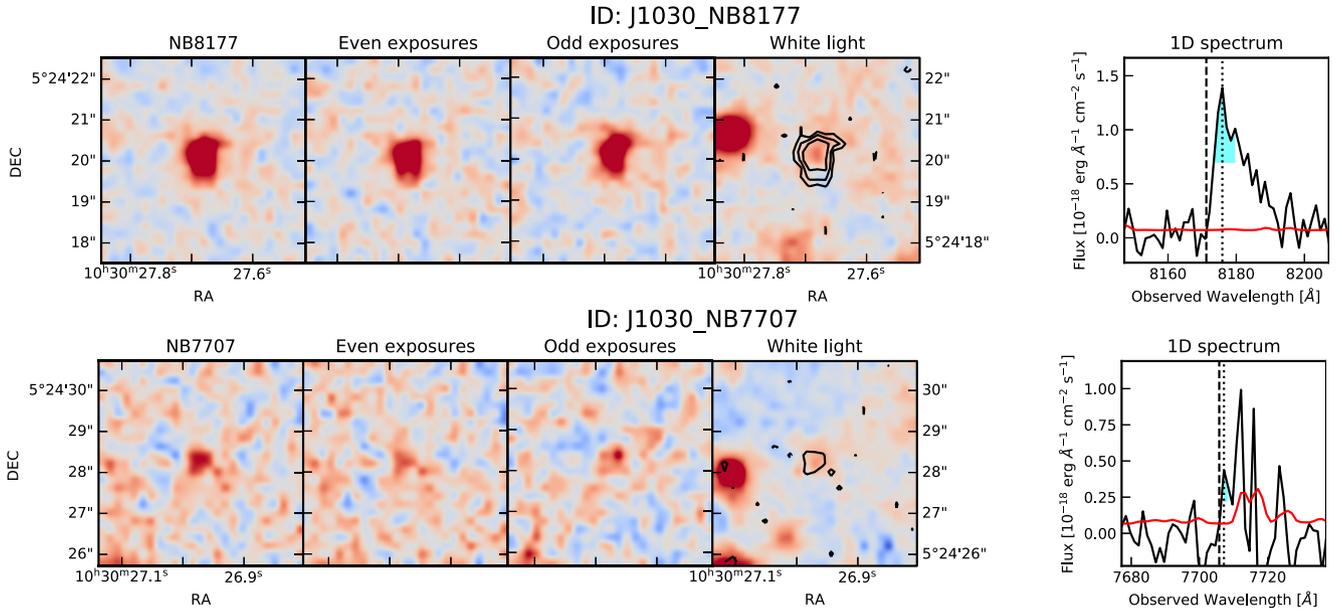
Figure 4. Two representative LAEs detected in *MUSE* archival data in the field of quasar J1030+0524. The first detection is a strong LAE in the field of J1030 already reported by Díaz et al. (2015). The four first panels show, in order, the detection in the narrow band centred on the detection in the full combined cube, followed by the same location in cubes with either half of the exposures. The fourth panel shows the white light image with superimposed black significance levels (-3σ, 3σ, 5σ, and 8σ) of the narrow-band detection. The fifth panel shows the optimally extracted 1D spectra (black line) and the noise level (red line). The cyan shading highlights the FWHM of the line, used to correct the peak redshift (dotted vertical line) to systemic (dashed vertical line). The remainder of the LAE detections are summarized in Table B1 and individual detections are presented in Appendix B. A colour version is available online.

nearby contaminant (see fig. B7, online material, for typical false positives). We show two representative detections in Fig. 4 and the remainder in Appendix B (online material).

We checked that the number of LAEs is consistent with expectations from the LAE luminosity function (LF) integrated down to the *MUSE* sensitivity limit. We first compute the number of LAEs in a given co-moving volume using the LAE LF from de La Vieuville et al. (2019) and Herenz et al. (2019), which measured the LAE LF in deep *MUSE* datacubes with *MUSE*LET and LSDCAT, respectively, i.e. the same algorithms that we use. By comparing the numbers of LAEs predicted with the LAE LF on a deep 27 h field realized by the *MUSE* GTO team (Bacon et al. 2015) to the numbers of LAEs those authors recovered with LSDCAT, we estimate that LSDCAT has a recovery rate of $\simeq 32$ per cent for LAEs at $z \sim$ 5.5. This is a global recovery rate for all LAEs with peak flux density greater than $f > 4.8 \times 10^{-19}$ erg s^{-1} cm^{-2}, which is below the sensitivity of all *MUSE* observations used in this study. We then predict the number of LAEs, we expect to find in each of our *MUSE* quasar fields depending on the exposure time, effective survey area, and the redshift interval of the central quasar Lyman-α forest. We find good agreement between the predicted number (including LSDCAT efficiency) and the number of retrieved LAEs (Fig. 5), indicating a successful search for LAEs and low levels of contamination.

Table 1 summarizes all the LBGs and LAEs detected in our quasar fields, alongside the reference of the *MUSE* and DEIMOS programmes, and the quasar discovery reference study.

2.4 Correcting the Lyman-α-based redshifts

The red peak of the Lyman-α emission line, commonly observed without its blue counterpart at high-redshift due to the increasingly neutral IGM, is often shifted from the systemic redshift of the
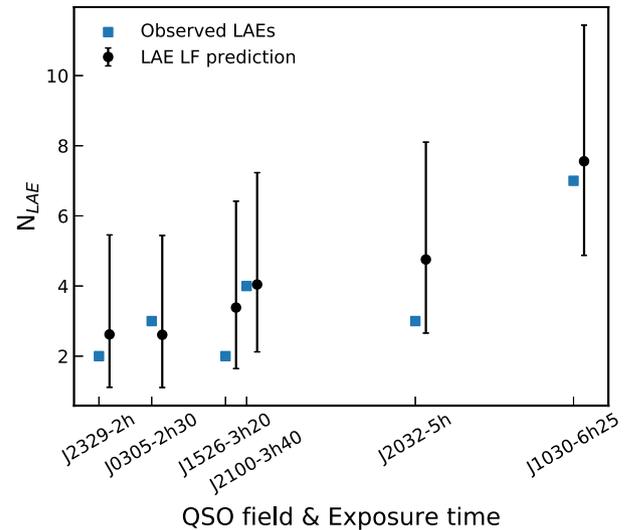
Figure 5. The predicted number of recovered LAEs in each *MUSE* cube (black 1σ Poisson ranges) compared to the number of retrieved LAEs (blue squares), with the quasar fields sorted (x-axis) by increasing the exposure time of the *MUSE* data. The prediction is made from the LAE LF of de La Vieuville et al. (2019) and Herenz et al. (2019) in the redshift range of the Lyman-α forest of each central quasar and effective efficiencies of LSDCAT and *MUSE* reduction. We apply an effective efficiency of the *MUSE* LAE search that is derived by comparison with the 27-h GTO observation of a single field searched for LAEs at high redshift with LSDCAT in Bacon et al. (2015). A colour version is available online.

galaxy. Steidel et al. (2010) give a typical offset of $v_{\rm red}^{\rm peak} \sim$ 200 km s^{-1}, but the range is large and can span $\sim 0 - 500$ km s^{-1} at high redshift (e.g. Vanzella et al. 2016; Stark et al. 2017; Hashimoto et al. 2018). A velocity offset of $\sim$200 km s^{-1} translates at $z \sim 6$ to

~2 cMpc (~ 280 pkpc), which is not negligible given the expected scale of ~10 cMpc for the peak of the cross-correlation signal. As the cross-correlation is computed in 3D space and then radially averaged, any offset would damp the sought-after signal.

In order to get a better estimate of the systemic redshift of the galaxy, we thus apply a correction to the Lyman-α redshift based on the FWHM of the line. We follow the approach of Verhamme et al. (2018), who developed an empirical calibration, using the FWHM directly measured from the data without modelling

$$v_{red}^{peak} = 0.9(\pm 0.14) \times \text{FWHM}(\text{Ly}\alpha) - 34(\pm 60)\text{km s}^{-1}. \quad (3)$$

The measured FWHM values of our LAEs (LBGs) all fall in the expected range $100 \lesssim \text{FWHM} \lesssim 400 \text{ km s}^{-1}$, and are indicated for each LAE (LBG) in Table B1. Throughout this paper, we use these corrected redshifts for the purpose of computing galaxy–IGM cross-correlations.

3 THE APPARENT CLUSTERING OF GALAXIES AND TRANSMISSION SPIKES FROM EIGHT QUASAR FIELDS

Galaxies are usually thought to be responsible for reionizing the Universe and driving the UVB fluctuations at $z \sim 6$. Having gathered a sample of 21 LAEs and 13 LBGs in the redshift range of the Lyman-α forest of nearby quasars, we are in a position to investigate the direct impact of galaxies on the surrounding IGM. The observational result of our work is summarized in Fig. 6, where we overlay the detected LAEs and LBGs with the transmission features found at the same redshift in the Lyman-α forest of the background quasar.

The natural corollary to the large-scale underdensity of galaxies found around highly opaque sightlines (Becker et al. 2018; Kashino et al. 2020) would be a close association between overdensities of transmission spikes and detected galaxies. We find that LAEs and LBGs are found close to at least one transmission spike in all quasar sightlines, but it is difficult to conclude at first sight whether they trace local spike overdensities. Moreover, this is not a reciprocal relation: some large transmission spikes are not associated with any detected galaxy. Two of our quasars, J1030 and J2032, illustrate this complicated relationship very well. The two sightlines both have a transparent region at $z \sim 5.5$, followed by a relatively opaque one at $z \sim 6$, and a similar *MUSE* exposure time. The transparent region in J1030 is associated with a large overdensity of LAEs and LBGs. In contrast, it is the few transmission spikes in the high-redshift opaque region in the sightline of J2032 that are associated with detected LAEs, whereas only one LAE is detected in the transparent region at $z \sim 5.5$. The detection of LAEs across $5 \lesssim z \lesssim 6$ in both quasar fields implies that we do not miss existing LAEs in these sightlines.

Hence, studying the correlations between galaxies and the IGM must been conducted in a more quantitative manner. In the following section, we compute the cross-correlation of the galaxies' positions with the transmission and the position of selected transmission spikes in the Lyman-α forest of the background quasar.

4 THE CORRELATION OF GALAXIES WITH THE SURROUNDING IGM TRANSMISSION

In this section, we introduce two indicators of the link between galaxies and the ionization state of the surrounding IGM: the cross-correlation of galaxies with the transmitted flux, and the 2-point correlation function (2PCCF) of galaxies with selected transmission spikes. We present the mean transmitted flux around galaxies (the

quantity used in Paper I) in Appendix C (online material) given that method has been superceded by the transmitted flux cross-correlation. Although the mean transmission around galaxies is the most intuitive measure of an enhanced UVB due to faint LyC leakers, in practice this measurement is dominated by the cosmic variance between sightlines and the redshift evolution of the IGM opacity, as noted in Paper II. For the purpose of these cross-correlation measurements, we only consider the Lyman-α forest between 1045 Å (to avoid the intrinsic Lyman-β/O VI quasar emission) and the start of the near-zone of the quasar, and consider only galaxies whose Lyman-α emission would fall in the same observed wavelength range.[4] These boundaries are plotted in dashed black lines in Fig. 6.

4.1 The cross-correlation of the IGM transmission with field galaxies

We first compute the cross-correlation of the IGM transmission with galaxies. We measure the transmission in the Lyman-α forest at a comoving distance r of observed galaxies (DD) and of random mock galaxies (DR). The distance $r = (r_\perp + r_\parallel)^{1/2}$ is computed from the angular diameter distance $r_\perp$ of the galaxy to the quasar sightline and the comoving distance parallel to the quasar line-of-sight $r_\parallel$. For each quasar field, we compute the probability of detecting an LAE at a given redshift in the quasar Lyman-α forest considering the LAE LF (de La Vieuville et al. 2019; Herenz et al. 2019) and the depth of the *MUSE* data. The redshifts of random galaxies are sampled from this probability distribution and the angular distances from the quasar sightlines are chosen uniformly up to 1 arcmin to mimic the *MUSE* FoV. For LBGs, we sample the UV LF (Bouwens et al. 2015; Bowler et al. 2015; Finkelstein et al. 2015; Ono et al. 2018) at the 2σ depth of the photometry of our three fields with an appropriate k-correction ($2.5(\alpha - 1)\log_{10}(1 + z)$, with $\alpha = 2$) and the angular separation from the quasar is drawn uniformly in the 4×16 arcmin^2 DEIMOS FoV. As noted in Section 2.2, the number of LBGs detected in each field depends primarily on the number of slitmasks observed in each field, rather than the depth of the photometric data used for selection. If we sampled each field down to the 3σ detection limit in the z band, we would thus predict similar numbers of predicted LBGs per field. However, a random sample created in this way would have a mean number of transmission spikes around the LBGs larger than is actually observed around our spectroscopically confirmed galaxies. Indeed, those observations that targeted higher redshift fields with reduced IGM transmission (e.g. J1030 and J1148) have greater spectroscopic coverage (due to the use of more than one DEIMOS mask) than for the lower redshift field of J0836. We therefore construct random LBGs by sampling the UV LF to the limiting depth (z') of each field matching the observed redshift distribution, but oversampling by the number of spectroscopic confirmations in each field. This procedure still reproduces the decline of the number of galaxies with redshift in each individual field.

The cross-correlation is then estimated using the standard estimator

$$\xi_{Gal-Ly\alpha}^{\exp(-\tau_\alpha)}(r) = \frac{DD(r)}{DR(r)} - 1. \quad (4)$$

Normalizing the transmitted flux in this way removes the bias introduced by the evolving IGM opacity, and allows us to average

---

[4]In the following, we loosely describe these galaxies as 'being in the (redshift range) of the Lyman-α forest of the quasar'.
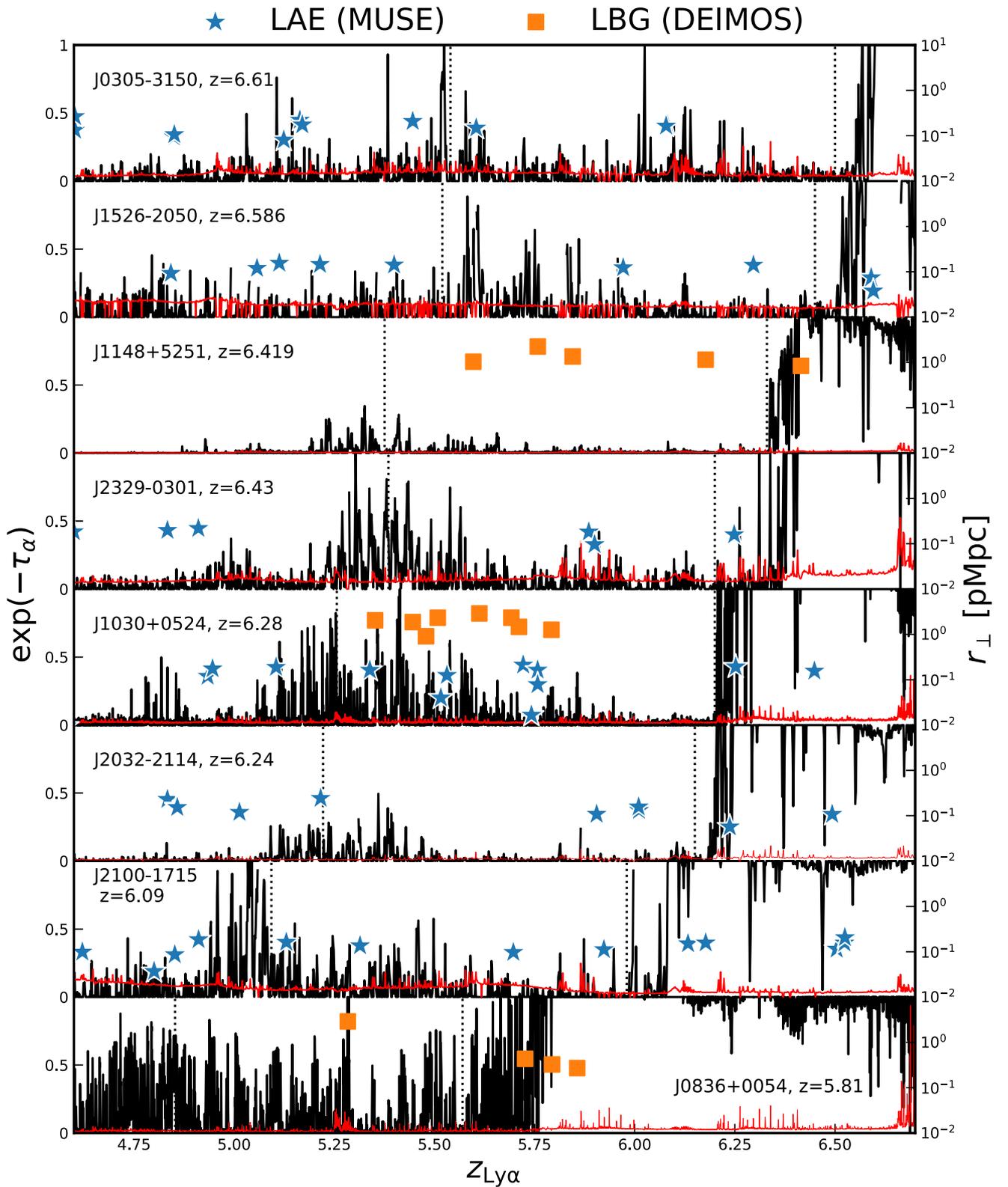
Figure 6. The Lyman-α forests of our high-redshift quasar sample and galaxies detected in the quasar fields. Whilst the average transmission is clearly decreasing with increasing redshift, the galaxies (LAEs in particular) seem to cluster with transmission spikes in some sightlines. The transmission in the quasar Lyman-α forest (black) is indicated on the left axis, whilst the right axis indicates the transverse distance of foreground galaxies to the quasar sightline. LBGs detected with DEIMOS are indicated in orange squares. *MUSE* LAE detections are indicated with blue stars. Only J1030 displays both LAEs and LBGs as it is the only field with DEIMOS and *MUSE* data. The vertical black dotted lines indicate the redshift range of the Lyman-α forests, as defined in Section 4. A colour version is available online.
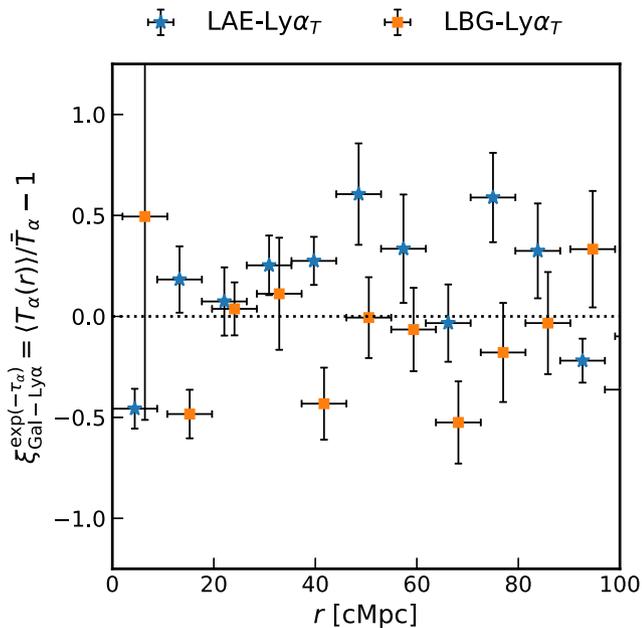
Figure 7. Cross-correlation of the position of LAEs (blue stars) and LBGs (orange squares) with the IGM transmission in the Lyman-α forest of the background quasar. The errorbars are computed by bootstrap resampling of the sample of detected galaxies. The significant deficit of transmission in the first bin of the LAE transmission cross-correlation is consistent with that measured around high-redshift C IV absorbers (Paper II). A colour version is available online.

sightlines without being biased by the most transparent ones (see Appendix C, online material).

We present the cross-correlation independently for LAEs and LBGs in Fig. 7. We do not find any evidence for an excess transmission, unlike that seen around lower redshift C IV absorbers in Paper II. The signal is still dominated by the small number of objects and sightlines as the large uncertainties show. The errors are estimated by bootstrapping the sample of detected galaxies, and thus they might be even underestimated given the small sample of sightlines and the large cosmic variance seen between Lyman-α forest at that redshift (Bosman et al. 2018). The issue is potentially more acute for LBGs as the selection is not completely down to a given luminosity as (1) only a fraction of drop-out candidates could be observed per field due to the instrument and time constraints (2) only a fraction of LBGs have a bright Lyman-α line (e.g. Stark et al. 2010; Ono et al. 2012; De Barros et al. 2017; Mason et al. 2019). Finally, we did not remove completely opaque parts of the Lyman-α where the flux is below the noise level unlike in Paper II, as it would greatly reduce our sample. The measured fluxes are therefore dominated by noise in some sections of the Lyman-α forest, which decreases the signal. None the less, we find that the absorption on small scales $\lesssim 1$ pMpc around LAEs is similar to that seen for C IV absorbers in Paper II. The direct association or not of C IV with LAEs is outside the scope of this paper and is studied in Diaz et al. (2020)

4.2 The two-point correlation of galaxies with selected transmission spikes

At $z \gtrsim 5.5$, the opacity of the IGM has increased sufficiently that the Lyman-α forest resembles more a 'savannah' than a forest: a barren landscape of opaque troughs occasionally interrupted by

transmission spikes. At these redshifts, the effective opacity can only be constrained with an upper limit ($\tau_{\mathrm{eff}} \gtrsim 3$–4), and the average opacity in large sections of the Lyman-α forest falls below this limit. It is thus not surprising that the previous transmission cross-correlation fails to capture the link between galaxies and the reionizing IGM. Indeed, the normalization term $DR(r)$ is often ill-defined at $z \gtrsim 6$ since the average flux measured is below or at the level of the noise of the spectrograph. To circumvent this issue, we examine the extrema of the opacity distribution rather than its mean by utilizing the 2PCCF of galaxies with selected transmission spikes in the Lyman-α forest of quasars. We expect the observed Lyman-α transmission to be more sensitive to fluctuations of the extrema of the distribution, making the 2PCCF theoretically more suited to capturing small perturbations due to clustered faint contributors to reionization.

We thus measure the 2PCCF between galaxies and selected transmission spikes in the Lyman-α forest. We identify transmission spikes with a Gaussian matched-filtering technique (e.g. Hewett et al. 1985; Bolton et al. 2004). We use Gaussian kernels with $\sigma =$ [10, 14, 20] km s^{-1} to pick individual small spikes and compute the SNR for each kernel at each pixel. We keep the best SNR at each pixel, and we select local peaks at SNR > 3, with $T >$ 0.02 (corresponding to $\tau_\alpha \lesssim 4$) as the positions of our transmission spikes. The transmission threshold ($T > 0.02$) was chosen to balance recovery of the small transmission spikes in J1148 and J2032 whilst limiting spurious detections in sightlines with worse SNR (J1526, J2100), and the SNR threshold as a compromise between the purity of the spike sample and large enough numbers to have reasonable bootstrap error estimates. We present an example of the successful recovery of transmission spikes in the high-redshift Lyman-α forest in Fig. 8.

We then estimate the 2PCCF with the estimator

$$\xi^{\mathrm{2PCCF}}_{\mathrm{Gal-Ly\alpha}} = \frac{D_{\mathrm{G}}D_\lambda(r)}{R_{\mathrm{G}}D_\lambda(r)} - 1 \,, \qquad (5)$$

where $D_{\mathrm{G}}D_\lambda$ is the number of transmission spikes–galaxy pairs normalized by the number of pixels in each radial bin and $R_{\mathrm{G}}D_\lambda$ is the normalized number of transmission spikes–random galaxy pairs, and r the 3D co-moving distance. As for the transmission cross-correlation, the redshift of random galaxies are sampled from the LAE or UV LF for LAEs and LBGs, respectively, the angular separation drawn from adequate uniform distributions, and the errors are computed by bootstrapping the sample of detected galaxies.

We show in Fig. 9 the 2PCCFs for both LAEs and LBGs. We detect a positive signal at 3.2σ as an excess of transmission spikes on $r \sim 10$–60 cMpc scales around LAEs and a deficit of transmission spikes at $r \lesssim 10$ cMpc. We also find an excess (1.9σ) of transmission spikes on large scales around LBGs. The significance of the LAE(LBG) 2PCCF excess is decreased by $-1.5\sigma (-0.2\sigma)$ if the redshift correction is not applied (Section 2.4). We compare in Fig. D1 (online) the 2PCCFs with and without correction, with the excesses being reduced in the latter case. The absence of correlation (or even anticorrelation) on the smaller scales stems both from the increased absorption by dense gas around the central LAE (which we model in Section 5.2) and the redshift error, which dampens the signal (~ 200 km s^{-1} corresponding to ~ 1.8 cMpc at $z \sim 6$). The reduced significance of the LBG 2PCCF could stem from an inappropriate normalization due to the difficulty of creating randomly samples of LBGs. As detailed previously, we have conservatively decided to scale the number of random galaxies to that of the observed ones. However, if some of the fields are indeed
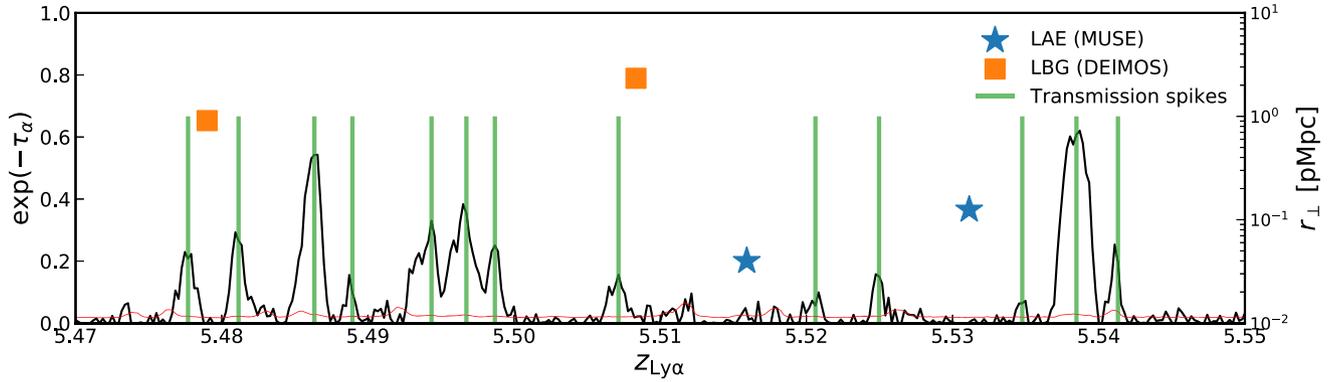
Figure 8. Zoom on the Lyman-α forest of J1030 (Fig. 6). LAEs and LBGs are indicated with blue stars and orange squares, respectively, and the location of transmission spikes identified with a Gaussian-matched filter with green vertical lines. A colour version is available online.
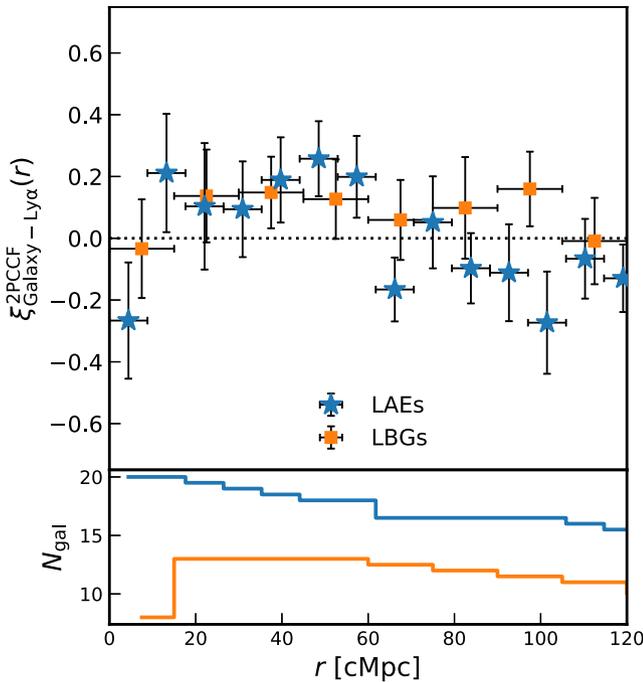
Figure 9. Upper panel: the 2PCCF of LBGs (orange squares) and LAEs (blue stars) with selected Lyman-α transmission spikes at $z \sim 6$. The errorbars are bootstrap errors on the number of detected galaxies. We find a significant excess of transmission spikes on scales 10–60 cMpc around LAEs (3.2σ). The excess of transmission spikes around LBGs is significant at 1.9σ, but might extend to larger scales. We point out, however, that the LBG selection is less complete than the LAEs due to the DEIMOS mask design and that one of the three quasar fields (J0836) has only one detected LBG. On the smaller scales ($\lesssim 1$ pMpc or $\lesssim 7$ cMpc), a deficit of transmission spikes is possibly present. The scales of the excess and the deficit are in good agreement with the measurements of Paper I and Paper II. Lower panel: number of galaxies contributing to the 2PCCF in each radial bin. Note that due to the redshift distribution of galaxies and the limits of the Lyman-α forests, at larger distances some galaxies can only be correlated with transmission spikes at lower or higher redshift. In that case, we count these as contributing $N = 0.5$ instead of $N = 1$ to the total number of galaxies. A colour version is available online.

slightly overdense, we would be overestimating the normalization in the cross-correlation and thus decrease the significance of the signal. The difference in the strength of the signal between the transmission cross-correlation and the 2PCCF can be attributed to the uncertainty

in the mean flux at high redshift. We defer the discussion of this difference to Section 7. where we examine the impact of noise on our measurements of the flux transmission and transmission spikes cross-correlation with galaxies.

5 MODELLING THE GALAXY–LYMAN-α TRANSMISSION AND TWO-POINT CROSS-CORRELATIONS

In order to interpret the observed galaxy–Lyman-α forest cross-correlations, we use a radiative transfer model based on the halo occupation distribution (HOD) framework introduced in Paper I. Here, we summarize the key ingredients and extensions used in this paper.

Paper I derived the average H I photoionization rate at a distance r from a galaxy due to the clustered faint population

$$\langle \Gamma_{\mathrm{H\,I}}^{\mathrm{CL}}(r|M_{\mathrm{h}}, z) \rangle = \frac{\bar{\Gamma}_{\mathrm{H\,I}}}{\lambda_{\mathrm{mfp}}(z)} \int \frac{e^{-|r-r'|/\lambda_{\mathrm{mfp}}(z)}}{4\pi|r-r'|^2} \left[1 + \langle \xi_{\mathrm{g}}(|r'|) \rangle_L \right] \mathrm{d}^3 r',$$

$$= \bar{\Gamma}_{\mathrm{H\,I}} \left[1 + \int_0^\infty \frac{k^2 \mathrm{d}k}{2\pi^2} R(k\lambda_{\mathrm{mfp}}(z)) \right.$$

$$\left. \times \langle P_{\mathrm{g}}(k|M_{\mathrm{h}}, z) \rangle_L \frac{\sin kr}{kr} \right], \quad (6)$$

where $\lambda_{\mathrm{mfp}}(z) = 6.0(1 + z/7)^{-5.4}$ (Worseck et al. 2014) is the mean free path of ionizing photons and $R(k\lambda_{\mathrm{mfp}}) = \arctan(k\lambda_{\mathrm{mfp}})/(k\lambda_{\mathrm{mfp}})$ is the Fourier transform of the radiative transfer kernel $e^{-r/\lambda_{\mathrm{mfp}}}/(4\pi r^2 \lambda_{\mathrm{mfp}})$. The luminosity-weighted galaxy power spectrum is

$$\langle P_{\mathrm{g}}(k|M_{\mathrm{h}}, z) \rangle_L = \frac{\int_{L_{\mathrm{UV}}^{\min}}^\infty L_{\mathrm{UV}} \Phi(L_{\mathrm{UV}}|z) P_{\mathrm{g}}(k, L_{\mathrm{UV}}|M_{\mathrm{h}}, z) \mathrm{d}L_{\mathrm{UV}}}{\int_{L_{\mathrm{UV}}^{\min}}^\infty L_{\mathrm{UV}} \Phi(L_{\mathrm{UV}}|z) \mathrm{d}L_{\mathrm{UV}}}, \quad (7)$$

where $P_g(k, L_{\mathrm{UV}}|M_{\mathrm{h}}, z)$ is the Fourier transform of the correlation function between bright tracers (i.e. detected LBGs and LAEs) with host-halo mass $>M_{\mathrm{h}}$ and galaxies with luminosity L_{UV}. We assume only central galaxies will be detected as LBGs or LAEs and therefore populate each halo with an HOD using a step function, $\langle N|M_{\mathrm{h}} \rangle = 1$ for halo mass $>M_{\mathrm{h}}$ and zero otherwise. Fainter galaxies are populated using the conditional LF pre-constrained by simultaneously fitting the $z \sim 6$ UV LF (Bouwens et al. 2015; Bowler et al. 2015; Finkelstein et al. 2015; Ono et al. 2018) and the galaxy autocorrelation function (Harikane et al. 2016) as in Paper I.

5.1 From the cross-correlation of galaxies with transmitted flux to the 2PCCF

As in Paper I and Paper II, the enhanced UVB can be used to compute the mean Lyman-α forest transmission at a distance r of galaxy,

$$
\langle \exp(-\tau_\alpha)(r|M_\mathrm{h}, z) \rangle = \int \exp\left[-\tau_\alpha\left(\Delta_\mathrm{b}, \langle\Gamma_\mathrm{HI}^\mathrm{CL}(r|M_\mathrm{h}, z)\rangle\right)\right]
$$
$$
\times\, P_V(\Delta_\mathrm{b}|r, M_\mathrm{h})\mathrm{d}\Delta_\mathrm{b}, \tag{8}
$$

where $P_V(\Delta_\mathrm{b}|r, M_\mathrm{h})$ is the volume-averaged PDF of the baryon overdensities Δ_b at a distance r from our galaxy tracer with a halo of mass M_h at redshift z, and $\langle\Gamma_\mathrm{HI}^\mathrm{CL}(r)\rangle$ is the clustering-enhanced photoionization rate modelled previously. The optical depth τ_α is derived using the fluctuating Gunn–Peterson approximation (see, e.g. Becker, Bolton & Lidz 2015, for a review),

$$
\tau_\alpha \simeq 11\Delta_\mathrm{b}^{2-0.72(\gamma-1)}\left(\frac{\Gamma_\mathrm{HI}}{10^{-12}\,\mathrm{s}^{-1}}\right)^{-1}\left(\frac{T_0}{10^4\,\mathrm{K}}\right)^{-0.72}\left(\frac{1+z}{7}\right)^{9/2}, \tag{9}
$$

where Δ_b is the baryon overdensity and T_0 is the temperature of the IGM at mean density. We include thermal fluctuations of the IGM using the standard power-law scaling relation (Hui & Gnedin 1997; McQuinn & Upton Sanderbeck 2016),

$$
T(\Delta_b) = T_0\Delta_\mathrm{b}^{\gamma-1}, \tag{10}
$$

assuming the fiducial values $T_0 = 10^4$ K and $\gamma = 1.3$.

We now expand this framework to predict a new statistic: the probability of seeing a transmission spike in the Lyman-α forest. Given a transmission threshold over which a detection is considered secure, we can derive an equivalent optical depth threshold. We fix the transmission threshold at $\exp(-\tau_\alpha) \gtrsim 0.02$, corresponding to $\tau_\alpha^\mathrm{th} \simeq 4$, to match our measurement of the 2PCCF. By substituting the predicted clustering-enhanced photoionization rate and the threshold optical depth in equation (9), we obtain the maximum baryon underdensity $\Delta_\mathrm{b}^\mathrm{max}$ required to produce a detected transmission spike in the Lyman-α forest at a distance r of a tracer galaxy,

$$
\Delta_\mathrm{b} \leq \Delta_\mathrm{b}^\mathrm{max}(\Gamma_\mathrm{HI}) \simeq 0.57\left(\frac{\tau_\alpha^\mathrm{th}}{4}\right)^{0.56}\left(\frac{\Gamma_\mathrm{HI}}{10^{-12}\,\mathrm{s}^{-1}}\right)^{0.56}
$$
$$
\times\left(\frac{T_0}{10^4\,\mathrm{K}}\right)^{0.4}\left(\frac{1+z}{7}\right)^{-2.52}. \tag{11}
$$

Thus, the occurrence probability of Lyman-α transmission spike at a location with H I photoionization rate Γ_HI is given by the probability to reach such an underdensity:

$$
P\left[< \Delta_\mathrm{b}^\mathrm{max}(\Gamma_\mathrm{HI})\,|r, M_\mathrm{h}\right] = \int_0^{\Delta_\mathrm{b}^\mathrm{max}(\Gamma_\mathrm{HI})} P_V(\Delta_\mathrm{b}|r, M_\mathrm{h})\mathrm{d}\Delta_\mathrm{b}. \tag{12}
$$

The cross-correlation between galaxies and the Lyman-α transmission spikes can therefore be modelled as the excess occurrence probability, $P[< \Delta_b^\mathrm{max}(\langle\Gamma_\mathrm{HI}^\mathrm{CL}(r)\rangle)|r, M_\mathrm{h}]$, of transmission spikes around an object with host halo mass M_h and an enhanced photoionization rate $\langle\Gamma_\mathrm{HI}^\mathrm{CL}\rangle$ relative to one at mean photoionization rate $\bar{\Gamma}_\mathrm{HI}$ and average density fluctuations, i.e. $P[< \Delta_b^\mathrm{max}(\bar{\Gamma}_\mathrm{HI})|r \to \infty, M_\mathrm{h}]$. It is then straightforward to deduce the cross-correlation between galaxies and the transmitted Lyα spikes as

$$
\xi_\mathrm{Gal-Ly\alpha}^\mathrm{2PCCF}(r) = \frac{P\left[< \Delta_b^\mathrm{max}\left(\langle\Gamma_\mathrm{HI}^\mathrm{CL}(r)\rangle\right)|r, M_\mathrm{h}\right]}{P\left[< \Delta_b^\mathrm{max}(\bar{\Gamma}_\mathrm{HI})|r \to \infty, M_\mathrm{h}\right]} - 1. \tag{13}
$$

The advantage of such a statistic over the transmission cross-correlation is that given the large number of pixels in high-resolution spectra of high-redshift quasars, a very low probability of transmission can still be measured with acceptable significance, whereas often only an upper limit on the mean flux can be measured at $z \gtrsim 6$.

5.2 Extending our UVB model with varying mean free path and gas overdensities

We now proceed to extend the model of UVB enhancement due to galaxy clustering by adding a varying mean free path and taking into account the gas overdensities associated with LAEs and LBGs on scales of several cMpc.

We first consider the effect of gas overdensity using the relevant probability distribution function. We derive the conditional PDF of overdensities around suitable haloes $P_V(\Delta_\mathrm{b}|r, M_\mathrm{h})$ from the IllustrisTNG simulations (Nelson et al. 2018). We utilize the TNG100-2 simulation for host halo masses $10^{10.5} < \mathrm{M}_\mathrm{h} < 10^{11.7}\,\mathrm{M}_\odot$, whereas for larger host halo masses $(M_\mathrm{h} > 10^{11.7}\,\mathrm{M}_\odot)$ we use TNG300-3 in order to get a higher number of such haloes at the cost of larger gas and dark-matter particle masses. We present in Appendix D (online material) the extracted conditional PDFs for a range of halo masses and radii.

Following Miralda-Escude, Haehnelt & Rees (2000) and Pawlik, Schaye & van Scherpenzeel (2009), we then fit each conditional PDF with a parametrization of the form

$$
P_V(\Delta_\mathrm{b}|r, M_\mathrm{h})\mathrm{d}\Delta_\mathrm{b} = A(r, M_\mathrm{h})\exp\left[-\frac{\left(\Delta_b^{-2/3} - C_0(r, M_\mathrm{h})\right)^2}{2(2\delta_0(r, M_\mathrm{h})/3)^2}\right]
$$
$$
\times\, \Delta_b^{-\beta(r, M_\mathrm{h})}\mathrm{d}\Delta_\mathrm{b}, \tag{14}
$$

with the parameter $A(r, M_\mathrm{h})$ being determined by requiring that the integral of the PDF is unity $(\int P_V(\Delta_\mathrm{b}|r, M_\mathrm{h})\mathrm{d}\Delta_\mathrm{b} = 1)$. The fitted values of A, C_0, δ_0, and β are listed in Table D1 (online material) for a choice of the relevant range of (r, M_h) whilst the rest are given in Appendix D (online material). We show in Fig. 10 the good agreement of the fits with the simulated PDF in a snapshot at $z = 5.85$ and a chosen central halo mass $M_\mathrm{h} \sim 10^{11.2\pm0.1}\,\mathrm{M}_\odot$ corresponding to the one derived from the clustering of LAEs (Ouchi et al. 2018).

Paper I considered a constant mean free path for simplicity. Introducing a full self-consistency of the mean free path down to ckpc scales in equation (6) is the realm of numerical simulations if a real distribution of gas and discrete sources is to be considered (and not the average distribution we use here). However, we can approximate variations of the mean free path to first order. Following Miralda-Escude et al. (2000), McQuinn, Peng Oh & Faucher-Giguère (2011), Davies & Furlanetto (2016), and Chardin, Puchwein & Haehnelt (2017), the mean free path of ionizing photons is dependent on the photoionization rate and the mean density of hydrogen,

$$
\lambda_\mathrm{mfp}(r) = \lambda_0\left(\frac{\langle\Gamma_\mathrm{HI}^\mathrm{CL}(r)\rangle}{\bar{\Gamma}_\mathrm{HI}}\right)^{\beta_\mathrm{mfp}}\left[\int \Delta_\mathrm{b} P_V(\Delta_\mathrm{b}|r, M_\mathrm{h})\mathrm{d}\Delta_\mathrm{b}\right]^{-\gamma_\mathrm{mfp}}, \tag{15}
$$

where β_mfp and γ_mfp reflects a simple parametrization of the mean free path dependence on the local UVB and gas overdensity. In this work, we chose to use the values $\beta_\mathrm{mfp} = 2/3$ and $\gamma_\mathrm{mfp} = -1$ for our fiducial model following previous works cited above.

Given the mutual dependence between $\langle\Gamma_\mathrm{HI}^\mathrm{CL}(r)\rangle$ and $\lambda_\mathrm{mfp}(r)$, we iterate the computation until $\langle\Gamma_\mathrm{HI}^\mathrm{CL}(r)\rangle$ is converged at the 1 per cent level at every distance r. As expected, a varying mean free path does not affect the photoionization rate on large scales but decreases it
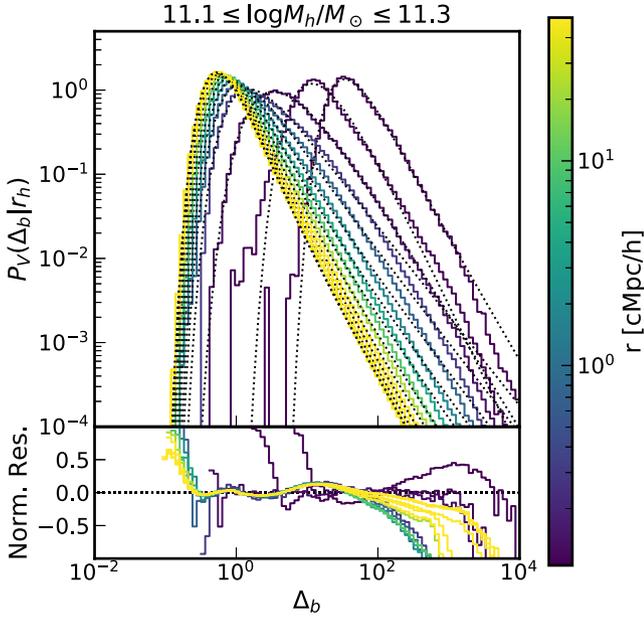
Figure 10. Upper panel: a subsample of the conditional PDFs from $r = 10^{-1}$ to $r = 10^{1.5}$ cMpc h^{-1} in increments of 0.045 dex. The fits with equation (14) are overlaid (dotted black) on top of the PDF extracted from the IllustrisTNG 100-2 simulation box in a snapshot at $z \sim 5.85$ (coloured histograms). Lower panel: residuals of the PDF fit, coloured by distance from the centre of the halo, showing good agreement in the validity limits of the prescribed analytical form between $10^{-1} \leq \Delta_b \leq 10^2$. A colour version is available online.

by a factor of 2–3 on scales $\lesssim 1$ cMpc. We show the impact on the predicted 2PCCF in Fig. 11. We find that any reasonable choice of $(\beta_{\rm mfp}, \gamma_{\rm mfp})$ modifies the 2PCCF only by a factor of <2 on scales $r < 10$ cMpc.

5.3 The observed 2PCCF

We have so far only considered the cross-correlation in real space. However, the observed two-point correlation is distorted by peculiar velocities and infall velocities. We consider here only the impact of random velocities and redshift errors. Following Hawkins et al. (2003) and Bielby et al. (2016), the real-space 2D correlation $\xi'(\sigma, \pi)$ is convolved with a distribution of peculiar velocities along the line-of-sight direction (π),

$$\xi(\sigma, \pi) = \int_{-\infty}^{+\infty} \xi'(\sigma, \pi - v/H(z)) f(v) \mathrm{d}v , \quad (16)$$

with an Gaussian kernel for the velocity distributions $f(v) = (2\pi\sigma_v^2)^{-1} \exp(-v^2/2\sigma_v^2)$. We use $\sigma_v = 200$ km s^{-1}, which is the observed scatter in the difference between Lyman-α and systemic redshifts at $z \sim 2-3$ (Steidel et al. 2010), encapsulating both redshift errors and the random velocities of galaxies. We finally take the monopole of the 2D correlation function,

$$\xi_0(s) = \frac{1}{2} \int_{-1}^{-1} \xi(\sigma, \pi) P_0(\mu) \mathrm{d}\mu , \quad (17)$$

where $s = \sqrt{\sigma^2 + \pi^2}$, $\mu = \pi/s$, and $P_0(\mu) = 1$ is the zeroth-order Legendre polynomial. As shown in Fig. 12, the peculiar velocities reduce slightly the signal on small scales.

We show in Fig. 12 various realizations of our model of the 2PCCF. We present here the impact of the modelling improvements
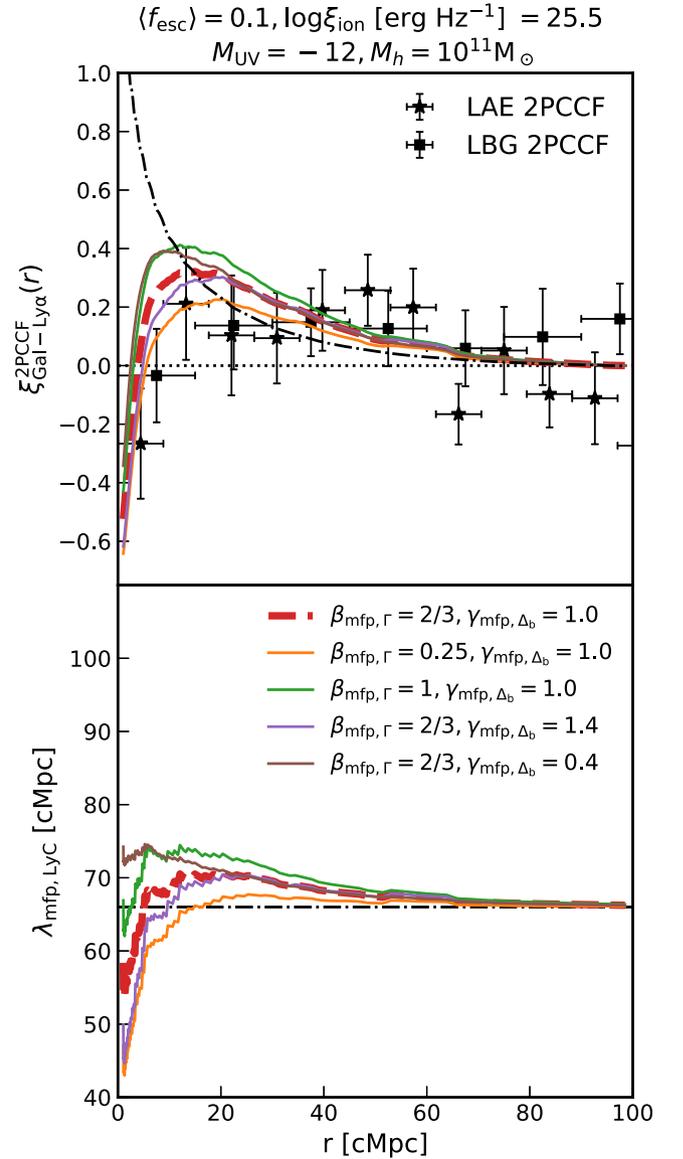
Figure 11. The impact of a spatially varying mean free path on the modelled 2PCCF of galaxies with transmission spikes in the Lyman-α forest. Variation of either the mean free path power-law dependence on the photoionization rate $(\beta_{\rm mfp})$ or the gas overdensity $(\gamma_{\rm mfp})$ do not affect significantly the predicted 2PCCF. The models are generated with the fiducial parameters, $\langle f_{\rm esc} \rangle = 0.1$, $M_{\rm h} = 10^{11}$ M$_\odot$, $\log \xi_{\rm ion}/[{\rm erg}^{-1}{\rm Hz}] = 25.5$, and $\gamma_{\rm mfp} = 1.3$. The black dash–dotted line in the upper panel shows a model with a fixed mean free path.

that we described previously. The addition of gas overdensities decreases the correlation on the smallest scales ($r \lesssim 20$ cMpc). The varying mean free path has little impact on the final shape of the predicted two-point correlation function, but boosts it slightly at $r > 20$ cMpc. Finally, the redshift errors and random velocities have a negligible impact on scales larger than few cMpc.

We conclude this modelling section by comparing the 2PCCF to the data for various fiducial parameters of the limiting luminosity of contributors to reionization $M_{\rm UV}^{\rm lim}$, escape fraction of the LyC photons $\langle f_{\rm esc} \rangle$ and host halo mass $M_{\rm h}$ of the detected bright galaxy in Fig. 13. We adopt a fiducial $\log \xi_{\rm ion}/[{\rm erg}^{-1}{\rm Hz}] = 25.5$, $\beta_{\rm mfp} = 2/3$, $\gamma_{\rm mfp} = -1$. Increasing the minimum UV luminosity increases the

$$\langle f_{\rm esc}\rangle = 0.1,\ \log \xi_{\rm ion}/[\rm erg\ Hz^{-1}] = 25.5$$
$$M_{\rm UV}^{\rm lim} = -12,\ M_h = 10^{11} M_\odot$$
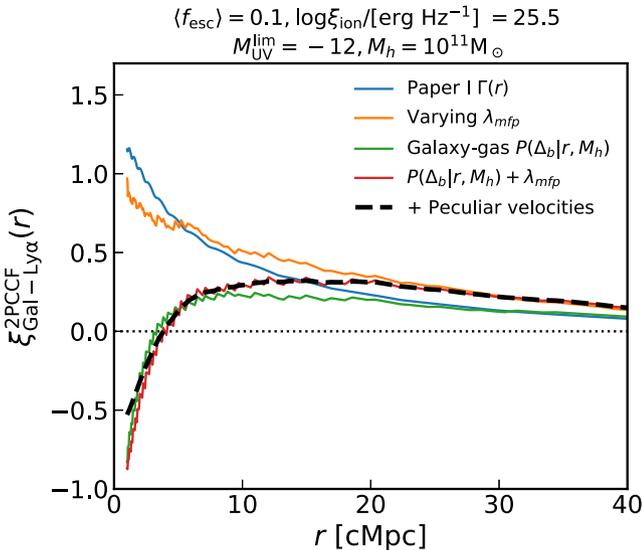
Figure 12. Successive versions of our model showcasing the increasing levels of refinement detailed throughout Section 5. The resulting predicted 2PCCF is mainly sensitive to the original UVB computed as in Paper I, but is dampened on small scales by the addition of a realistic gas overdensity PDF. The implementation of a variable mean free path enhances the signal on large scales.

correlation as more sources contribute to the local photoionization rate. We find that the host halo mass of the tracer galaxy is correlated positively with the 2PCCF signal strength, as they cluster more strongly with other galaxies. Finally, the escape fraction has a non-trivial effect on the cross-correlation: because it affects both the local and the overall UVB, an increase in the escape fraction decreases the total 2PCCF. Indeed, excesses of ionized gas close to clustered galaxies become harder to detect as the Universe becomes fully ionized and transmission spikes are ubiquitous. We defer to Appendix E (online material) for a full mathematical derivation of the role of $f_{\rm esc}$ in our modelled 2PCCF.

6 CONSTRAINTS ON THE IONIZING CAPABILITIES OF $z \sim 6$ CONTRIBUTORS CLUSTERED AROUND LAES

Our model of the statistical proximity effect of galaxies based on their correlation with Lyman-α transmission spikes can be applied at different redshifts, across absorbed and transparent sightlines, and to galaxy populations with different halo masses. We have detected a signal in the 2PCCF of high-redshift LAEs and LBGs with Lyman-α transmission spikes, which we will now proceed to fit.

The median redshift of our LAE (LBG) sample is $\langle z \rangle = 5.82(5.597)$. We therefore use the gas overdensity PDF from the Illustris TGN100-2 (TNG300-3 for LBGs) at $z = 5.85$ (the closest snapshot in redshift to the larger LAE sample), and we fix the redshift at the same value for the computation of the CLF and our 2PCCF model in general for consistency. We use the fiducial values of $\beta_{\rm mfp} = 2/3, \gamma_{\rm mfp} = -1$ for the mean free path of ionizing photons, and a temperature density relation $T \propto \Delta_{\rm b}^{\gamma-1}$ with $\gamma = 1.3$.

Our model constrains the number of ionizing photons emitted around galaxies, and thus the luminosity-weighted-average

contribution[5] of sources to reionization

$$\langle f_{\rm esc}\xi_{\rm ion}\rangle_L = \frac{\int_{M_{\rm UV}^{\rm lim}}^{\infty} f_{\rm esc}(L_{\rm UV})\xi_{\rm ion}(L_{\rm UV})L_{\rm UV}\Phi(L_{\rm UV}){\rm d}L_{\rm UV}}{\int_{M_{\rm UV}^{\rm lim}}^{\infty} L_{\rm UV}\Phi(L_{\rm UV}){\rm d}L_{\rm UV}},\quad (18)$$

which for simplicity we have recast with a fixed $\log \xi_{\rm ion}/[\rm erg^{-1}Hz] = 25.5$, such that our main results will the luminosity-averaged escape fraction. We emphasize that the limiting luminosity of contributing sources simply marks the truncation of the UV LF. A Gaussian prior to the turnover of the UV LF at $M_{\rm UV}^{\rm lim} \sim -12 \pm 1$ encompasses the scatter between different studies (e.g. Bouwens et al. 2015; Livermore et al. 2017; Atek et al. 2018) and the recent constraint via the extragalactic background light measurement (Fermi-LAT Collaboration 2018).

We fit the LAE 2PCCF signal with the EMCEE Monte Carlo sampler (Foreman-Mackey et al. 2013) using a flat prior in the range $0 \le \langle f_{\rm esc}\rangle \le 1$, a Gaussian prior over $M_{\rm UV}^{\rm lim} \simeq -12$ with $\sigma_{M_{\rm UV}} = 1$, and another Gaussian prior for the host halo masses based on the angular clustering measurements of LAEs (Ouchi et al. 2018). We use the values of $\log M_{\rm h}^{\rm LAE}/[{\rm M}_\odot] = 11.1_{-0.4}^{+0.2}$ derived at $z = 5.7$ for all our LAE detections at $5.5 < z < 6.2$. We marginalize over LAE host mass and minimum luminosity priors get our final constraint from the LAE-spike 2PCCF

$$\langle f_{\rm esc}\rangle_{M_{\rm UV}\lesssim-12} = 0.14_{-0.05}^{+0.28}\quad (\log \xi_{\rm ion}/[\rm erg^{-1}Hz] = 25.5),\quad (19)$$

where the errors represent a 1σ credible interval. The LBG-spike 2PCCF, where the host halo mass prior of LBGs at $z \sim 6$ ($M_{\rm h}^{\rm LBG}/[{\rm M}_\odot] = 12.02_{-0.01}^{+0.02}$) is based on the clustering measurement with Hyper Suprime-Cam at the Subaru telescope (Harikane et al. 2018), gives the following constraint :

$$\langle f_{\rm esc}\rangle_{M_{\rm UV}\lesssim-12} = 0.23_{-0.12}^{+0.46}\quad (\log \xi_{\rm ion}/[\rm erg^{-1}Hz] = 25.5).\quad (20)$$

These average constraints on the entire luminosity range can be of course rearranged to test any given functional form of the escape fraction and/or the ionizing efficiencies, and accommodate other fiducial values of $\xi_{\rm ion}$. For example, we present in Fig. 14, the average escape fraction of galaxies as a function of the minimum UV luminosity of contributors between $-20 < M_{\rm UV}^{\rm min} < -10$. Our results are in good agreement with literature estimates derived from neutral fraction histories (e.g. Robertson et al. 2015; Naidu et al. 2020), especially for models invoking a substantive contribution of faint galaxies to reionization. Both LAE–IGM and LBG–IGM 2PCCF constraints are in agreement with the C IV–IGM transmission cross-correlation of Paper II. Although the three measurements' maximum-likelihood values differ, the uncertainties are still too large to conclude yet on any significant tension between the escape fraction of the galaxies traced by LAEs, LBGs, and C IV absorbers.

7 DISCUSSION

7.1 Relative contribution of sub-luminous sources

As the cross-correlation slope is sensitive to the minimum UV luminosity of contributing sources (Fig. 13), it is theoretically possible to measure simultaneously the luminosity-averaged escape fraction of reionization sources and their minimum or maximum luminosity. We now proceed to extend our analysis to test whether we can infer the relative contribution of bright and faint sources to reionization.

[5]Which we shorten to luminosity averaged for convenience.
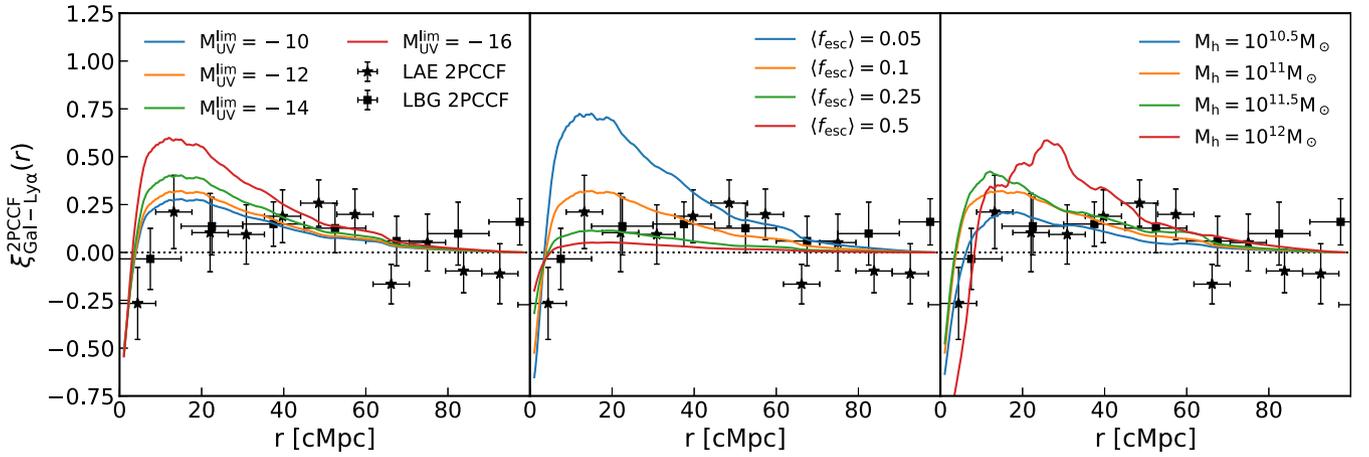
Figure 13. Examples of our model of the 2PCCF given a range of parameters (limiting luminosity of contributors to reionization M_{UV}^{lim}, escape fraction of LyC photons f_{esc} and host halo mass of the detected LAE/LBG M_h). In each panel, one parameter is varying, while the others are kept fixed at the fiducial values of $M_{UV}^{lim} = -12$, $\langle f_{esc} \rangle = 0.1$, $M_h = 10^{11} M_\odot$. All models plotted here assume a redshift of $z = 5.85$ and ionizing efficiency $\log \xi_{ion}/\mathrm{erg}^{-1}\mathrm{Hz} = 25.5$, and a mean free path dependence on the mean overdensity with exponent $\gamma_{mfp} = -1$.
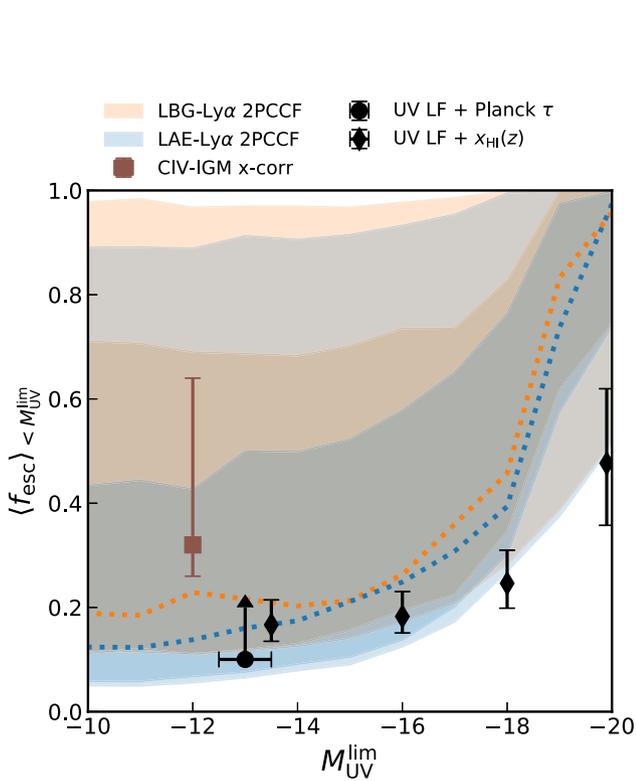
Figure 14. Luminosity-averaged escape fraction of galaxies contributing to reionization as a function of their minimum UV luminosity. The two measurements of the 2PCCF from this work (LAE/LBG 1σ, 2σ blue/orange contours) are in agreement with the C IV–IGM transmission cross-correlation from Paper II (brown square). The irregular shape of the LBG 2PCCF posterior is due to the gas overdensities PDF, which is poorly constrained by Illustris TNG due to the large mass and limited volume, and thus varies quickly from one mass bin to another, making the convergence of the MCMC chain difficult. The LAE–IGM 2PCCF is in better agreement with the average escape fraction derived from the UV LF and the *Planck* optical depth measurement (Robertson et al. 2015, black circle) or the neutral fraction history when the minimum UV luminosity of contributors is small (Naidu et al. 2020, black diamonds). The escape fractions are (re-)derived assuming $\log \xi_{ion}/[\mathrm{erg}^{-1}\mathrm{Hz}] = 25.5$. A colour version is available online.
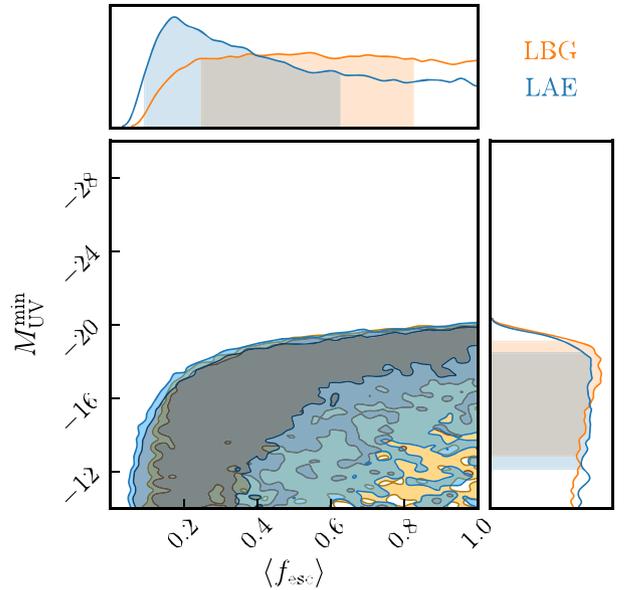
Figure 15. Posterior distributions of our fiducial model parameters with reionization dominated by more luminous galaxies fitted to the LAE/LBG-transmission spikes 2PCCF. The 2PCCF constrain the minimum UV luminosity of contributors to be at least $M_{UV} \lesssim -20$. Our fiducial models have the following fiducial parameters $\gamma = 1.3$, $\beta_{mfp} = 2/3$, $\gamma_{mfp} = -1$, and $\log \xi_{ion}/[\mathrm{erg}^{-1}\mathrm{Hz}] = 25.5$.

We examine two simple cases: a model where all galaxies fainter than a certain UV luminosity solely contribute to reionization and, conversely, a model where such faint galaxies do not contribute at all. To do so, we treat the minimum/maximum UV luminosity as a parameter and fit the model with a flat prior on this quantity. We then fit these two models to the LAE/LBG–IGM 2PCCF.

We present the posterior distribution of our parameters in Fig. 15 for the model where bright galaxies dominate, and the inferred constraints in Table 3. The LAE/LBG 2PCCF were fitted with the parameters described above except for a flat prior on the minimum UV luminosity of contributing sources, $-10 < M_{UV}^{min} < -30$. In both cases, the minimum UV luminosity of the contributing sources

Table 3. Summary of our constraints on the luminosity-weighted average escape fraction of galaxies at $z \sim 6$. For each galaxy overdensity tracer (LAEs or LBGs), we fit the galaxy–Lyman-α transmission spike 2PCCF for three different scenarios: a turnover of the LF at $M_{UV} = -12$ obtained by imposing a Gaussian prior (Section 6), reionization dominated by luminous galaxies (Fig. 15), and the reverse scenario where only low-luminosity galaxies contribute (Fig. 16). Our models assume the following IGM parameters: $\gamma = 1.3$, $\beta_{mfp} = 2/3$, $\gamma_{mfp} = -1$, and $\log \xi_{ion}/[erg^{-1} Hz] = 25.5$.

Tracer	$\langle f_{esc} \rangle$	M_{UV}^{min}	M_{UV}^{max}
	$0.14^{+0.28}_{-0.05}$	$-12.1^{+1.1}_{-0.9}$	-30
LAE	$0.18^{+0.44}_{-0.09}$	$> -19.0 \,(2\sigma)$	-30
	$0.14^{+0.36}_{-0.07}$	-10	$-14.1^{+2.4}_{-6.6}$
	$0.23^{+0.46}_{-0.12}$	$-12.0^{+0.9}_{-1.0}$	-30
LBG	$>0.17 \,(2\sigma)$	$> -19.1 \,(2\sigma)$	-30
	$>0.14 \,(2\sigma)$	-10	$< -17.3 \,(2\sigma)$

Figure 16. Posterior distributions of our fiducial model parameters where faint galaxies dominate, fitted to the LAE/LBG-transmission spikes 2PCCF. The LAE and LBG constraints are in tension, with the LAE 2PCCF favouring a model where only low-luminosity galaxies contribute ($-10 \gtrsim M_{UV} \gtrsim -17$) and the LBG 2PCCF requiring the contribution of more luminous objects up to at least $M_{UV} \lesssim -21$. Our fiducial model has the following IGM parameters: $\gamma = 1.3$, $\beta_{mfp} = 2/3$, $\gamma_{mfp} = -1$, and $\log \xi_{ion}/[erg^{-1} Hz] = 25.5$.

is $M_{UV}^{min} < -20.0 \,(2\sigma)$. In practice, however, a model where only galaxies brighter than $M_{UV} = -20.0$ is implausible because it would require an overwhelmingly high-luminosity-averaged escape fraction of ≈ 1, contradicting existing $z \sim 6$ measurements (Matthee et al. 2018) and marking a stark departure from measurements at lower redshift (e.g. Izotov et al. 2016, 2018; Vanzella et al. 2016; Fletcher et al. 2019; Tanvir et al. 2019). It thus more likely that, if only the brightest objects contribute, they include at least relatively faint galaxies down to $M_{UV} \sim -18(-16)$.

We now examine our model where faint galaxies dominate. The LAE/LBG 2PCCF were fitted with the parameters, as described in Section 6, except for a flat prior on the maximum UV luminosity of contributing sources $-10 < M_{UV}^{max} < -30$, and the minimum UV luminosity of LyC contributing sources was fixed at $M_{UV}^{min} = -10$. We present the posterior distribution of our parameters in Fig. 16, and the inferred constraints in Table 3. The posteriors for the LAE and LBG 2PCCF are strikingly different: whereas the LAE signal is well fitted by a model where faint galaxies ($-17 \lesssim M_{UV} \lesssim -10$) drive reionization, the LBG 2PCCF constrains the maximum luminosity of contributors to be at least $< -18.4 \,(2\sigma)$. In other words, the 2PCCF signal around more luminous tracers (LBGs) of galaxies is consistent with a contribution of brighter objects, whereas faint tracers (LAEs) favour an ionizing environment dominated by faint sources. Because clustering is already included in our model, this is not simply a consequence of LAEs likely sitting in smaller overdensities than LBGs, therefore tracing less massive and fainter objects. This result rather indicates that bright objects ($M \lesssim -20$) traced by LBGs have increased ionizing efficiencies. One natural explanation is that they would create early ionized bubbles, which would, in turn, enhance the confirmation rate with a Lyman-α emission-line detection of such LBG candidates. This result is in agreement with the study by Mason et al. (2018b), which found that the boosted transmission around bright ($M_{UV} < -22$) objects cannot only be explained by their biased environment, and thus they must have increased ionizing efficiencies.

As a conclusion, it is interesting to note that the 2PCCFs can be fitted with two mutually exclusive populations of galaxies: the sources or reionization can either be galaxies fainter or brighter than $M_{UV} \sim -18$. These two results show how the 2PCCF is able to constrain the parameters of a given specific model of escape fraction dependence on luminosity. However, identifying which model is correct is difficult with the current data as the likelihood ratio of the

two best (LAE 2PCCF) fits favours only very marginally (1.3σ) the faint-dominated scenario. Future measurements of galaxy–Lyman-α forest cross-correlations are required to distinguish between the two scenarios, as well the possible differences in the sub-populations traced by LAEs, LBGs, and other potential overdensity tracers.

7.2 Impact of noise on the detection of transmission spikes and the non-detection of a transmission cross-correlation

We now investigate whether or not we can explain the apparently contradictory absence of a transmission cross-correlation but the detection of the transmission spike two-point correlation (Fig. 7 and 9).

In order to do so, we use our improved model of the galaxy–GM cross-correlation, including the varying mean free path and the gas overdensities. We sample $P_V(\Delta_b | r, M_h)$ to generate 1000 values of transmission $\exp(-\tau_\alpha)$ for each distance r from the tracer LAE. We then sample the distribution of errors σ as measured in the Lyman-α forest pixels used in the cross-correlation measurement (i.e. after masking). We then add a flux error drawn from the normal distribution $\Delta_T \sim N(0, \sigma)$ to every computed flux value to mimic the effect of noise. Finally, we bin the data to match the measurement binning using the same number of mock Lyman-α forest 'pixel' points as the ones measured in the real quasar spectra. The transmission cross-correlation is computed as the mean flux value in each bin, whereas the 2PCCF is the fraction of transmission values above $T > 0.02$ (the same threshold used for the previous measurement).

The resulting mock observations are shown in Fig. 17 alongside the original model without errors. Clearly, the mean flux or the transmission cross-correlation is difficult to measure with any cer-
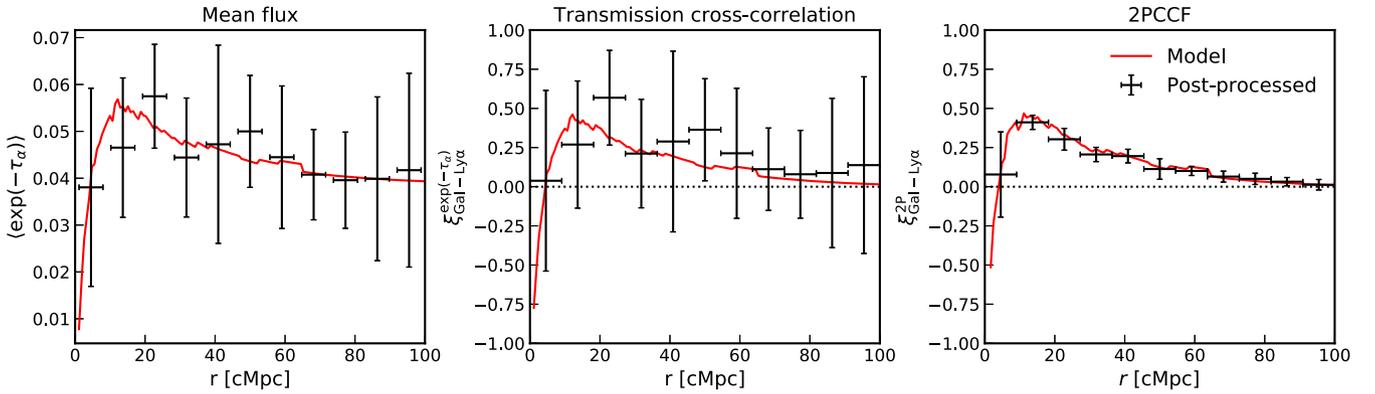
Figure 17. Post-processing of different statistics probing the Lyman-α forest–galaxy correlation from our improved model. The addition of noise directly drawn from the spectrsocopic data of quasars used in this study highlights the extreme difficulty of measuring the mean flux in the Lyman-α forest around galaxies (left Paper I) or the transmission cross-correlation with galaxies (centre Paper II). The 2PCCF is, however, robust to such perturbations, hence its use in this study.

tainty. This also explains why an increase in the mean transmission or a transmission cross-correlation is much harder to detect than the 2PCCF at $z \sim 6$, as we found in Section 4. The addition of noise is crucial because the noise level is comparable to the mean transmission ($T = 0.01 - 0.1$). It is thus no surprise that an increase in the average flux is difficult to measure. The 2PCCF, however, is shown to be rather unaffected by the addition of noise as the spikes we consider are at high enough SNR and transmission. Indeed, because the distribution of transmission pixels is lognormal (Bosman et al. 2018), there will be more pixels with intrinsic transmission below any given threshold ($T > 0.02$) than above. As the observational error is drawn from a normal distribution, there will be more pixels observed to have a higher transmission than the given threshold but with lower intrinsic transmission than the reverse, increasing the number of spurious spike detections. In practice, however, this only slightly decreases the 2PCCF and therefore the addition of noise is neglected in our modelling. We conclude that the 2PCCF is less biased by fluctuations of the mean opacity in different sightlines and should be less affected by continuum normalization uncertainties.

We now conclude by examining whether the observed transmission cross-correlation (Fig. 7) is consistent with the predicted uncertainty on the modelled signal generated using the best-fitting physical parameters of the 2PCCF LAE-transmission spike detection. We find that our LAE transmission cross-correlation measurement is in agreement with the predicted 1σ uncertainty range of the model. (Fig. 18). There appears to be a slight tension between the (LAE) post-processed model and the LBG measurement, but it is not very significant. The potential tension is more likely due to the smaller number of objects (and quasar sightlines) for the LBG transmission measurement, which would lead us to underestimate the errors on the measurement. This is expected as the bootstrap uncertainties are primarily limited by cosmic variance and small sample size, and this measurement might be accurate with a larger sample of quasar sightlines and foreground objects (e.g. Paper II).

8 SUMMARY

We have assembled a new sample of galaxies in the field of eight high-redshift quasars in order to examine various correlations between galaxies and the fluctuations in the Lyman-α forest at the
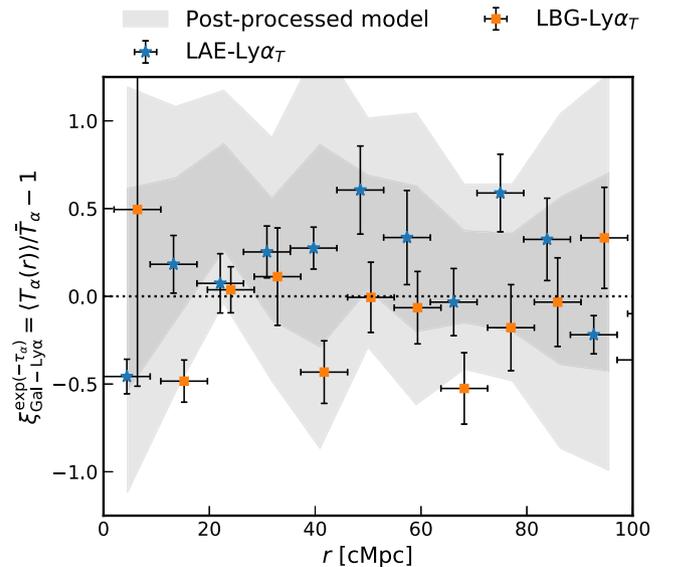
Figure 18. Comparison of the observed LAE/LBG–Lyα transmission cross-correlation (blue stars, orange square) with our post-processed (noisy) model (grey-shaded areas, 1σ and 2σ) generated using the best-fitting parameters for the LAE–Lyα 2PCCF. The observation is in agreement with the predicted uncertainties stemming from the fact that the SNR of the Lyman-α at $z \sim 6$ is extremely low, on average, due to the IGM opacity, and that our quasars were primarily chosen to have *MUSE* coverage rather than deep spectroscopy.

end of reionization. We have extended the approach pioneered in Paper I and Paper II to model the galaxy–Lyman-α flux transmission and the two-point correlation with transmission spikes. We report the following key findings:

(i) We find a 3.2(1.9)σ-significant excess in the 2PCCF of transmission spikes with LAEs(LBGs) at on scales of ∼10–60 cMpc. Our model of the LAE(LBG) 2PCCF constrains sources with $M_{UV} < -12$ to contribute to reionization with a luminosity-averaged escape fraction $\langle f_{esc} \rangle_{M_{UV} < -12} = 0.14^{+0.28}_{-0.05} (0.23^{+0.46}_{-0.12})$, assuming a fixed $\log \xi_{ion}/(erg^{-1}Hz) = 25.5$.

(ii) We present a new model of the 2PCCF of detected Lyman-α transmission spikes with LAEs, which includes a consistent treatment of gas overdensities around detected LAEs and their

peculiar velocities. We find that a spatially varying mean free path does not impact the 2PCCF significantly. We demonstrate that this model is more robust than the transmission cross-correlation at high redshift.

(iii) We show how parametric models of the escape fraction dependence on the galaxy luminosity can be constrained by the LAE–IGM 2PCCF. We find that the LAE 2PCCF is consistent with a local UVB enhanced either by faint galaxies with $M_{UV}^{max} = -14.1^{+2.4}_{-6.6}$ or brighter than $M_{UV} < -19.0\,(2\sigma)$. The LBG 2PCCF favours brighter objects with at least $M_{UV} < -19\,(2\sigma)$ contributing to reionization. Differentiating between these hypotheses will, however, require a larger data set of galaxies in quasar fields.

(iv) We find no evidence for a correlation between the transmission in the Lyman-α forest and LAEs/LBGs at $z \sim 6$. We show how this absence of signal is consistent with scatter and noise of our quasar sightlines. None the less, the deficit of transmission on scales up to ~ 10 cMpc is seen in the Lyman-α forest around LAEs as previously reported around C IV absorbers (Paper II).

REFERENCES

Ajiki M. et al., 2006, PASJ, 58, 499
Atek H., Richard J., Kneib J.-P., Schaerer D., 2018, MNRAS, 479, 5184
Bacon R. et al., 2010, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III. SPIE, Bellingham, p. 773508
Bacon R. et al., 2015, A&A, 575, A75
Bacon R., Piqueras L., Conseil S., Richard J., Shepherd M., 2016, Astrophysics Source Code Library, record ascl:1611.003
Bañados E. et al., 2016, ApJS, 227, 11
Bañados E. et al., 2018, Nature, 553, 473
Becker R. H. et al., 2001, AJ, 122, 2850
Becker G. D., Bolton J. S., Lidz A., 2015, PASA, 32, e045
Becker G. D., Davies F. B., Furlanetto S. R., Malkan M. A., Boera E., Douglass C., 2018, ApJ, 863, 92
Bielby R. M. et al., 2016, MNRAS, 456, 4061
Bolton A. S., Burles S., Schlegel D. J., Eisenstein D. J., Brinkmann J., 2004, AJ, 127, 1860
Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G. D., Haehnelt M. G., 2018, MNRAS, 479, 1055
Bosman S. E. I., Kakiichi K., Meyer R. A., Gronke M., Laporte N., Ellis R. S., 2019, preprint (arXiv:1912.11486)
Bouwens R. J. et al., 2015, ApJ, 803, 34
Bowler R. A. A. et al., 2015, MNRAS, 452, 1817
Chardin J., Puchwein E., Haehnelt M. G., 2017, MNRAS, 465, 3429
Cooper M. C., Newman J. A., Davis M., Finkbeiner D. P., Gerke B. F., 2012, Astrophysics Source Code Library, record ascl:1203.003
Davies F. B., Furlanetto S. R., 2016, MNRAS, 460, 1328
De Barros S. et al., 2017, A&A, 608, A123
de La Vieuville G. et al., 2019, A&A, 628, A3
Díaz C. G., Ryan-Weber E. V., Cooke J., Koyama Y., Ouchi M., 2015, MNRAS, 448, 1240
Diaz C. G., Ryan-Weber E., Karman W., Caputi K., Salvadori S., Crighton N., Ouchi M., Vanzella E., 2020, preprint (arXiv:2001.04453)
Drake A. B., Farina E. P., Neeleman M., Walter F., Venemans B., Banados E., Mazzucchelli C., Decarli R., 2019, ApJ, 881, 131
Eilers A.-C., Davies F. B., Hennawi J. F., Prochaska J. X., Lukić Z., Mazzucchelli C., 2017, ApJ, 840, 24
Eilers A.-C., Davies F. B., Hennawi J. F., 2018, ApJ, 864, 53
Faber S. M. et al., 2003, in Iye M., Moorwood A. F. M., eds, Proc. SPIE Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes. SPIE, Bellingham, p. 1657
Fan X., Narayanan V. K., Strauss M. A., White R. L., Becker R. H., Pentericci L., Rix H. W., 2002, AJ, 123, 1247
Fan X. et al., 2006, AJ, 132, 117
Farina E. P. et al., 2017, ApJ, 848, 78
Farina E. P. et al., 2019, ApJ, 887, 34
Fermi-LAT Collaboration T. F.-L., 2018, Science, 362, 1031
Finkelstein S. L. et al., 2015, ApJ, 810, 71
Fletcher T. J., Tang M., Robertson B. E., Nakajima K., Ellis R. S., Stark D. P., Inoue A., 2019, ApJ, 878, 87
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Garaldi E., Compostella M., Porciani C., 2019, MNRAS, 483, 5301
Giallongo E. et al., 2008, A&A, 482, 349
Giallongo E. et al., 2015, A&A, 578, A83
Greig B., Mesinger A., 2017, MNRAS, 465, 4838
Gunn J. E., Stryker L. L., 1983, ApJS, 52, 121
Harikane Y. et al., 2016, ApJ, 821, 123
Harikane Y. et al., 2018, PASJ, 70,
Hashimoto T. et al., 2018, Nature, 557, 392
Hawkins E. et al., 2003, MNRAS, 346, 78

Herenz E. C., Wisotzki L., 2017, A&A, 602, A111
Herenz E. C. et al., 2019, A&A, 621, A107
Hewett P. C., Irwin M. J., Bunclark P., Bridgeland M. T., Kibblewhite E. J., He X. T., Smith M. G., 1985, MNRAS, 213, 971
Hill J. M., Salinari P., 2000, in Sebring T. A., Andersen T., eds, Proc. SPIE 4004, Telescope Structures, Enclosures, Controls, Assembly/Integration/Validation, and Commissioning. SPIE, Bellingham, p. 36
Hinton S., Hinton R. S., 2016, J. Open Source Softw., 1, 45
Horne K., 1986, PASP, 98, 609
Hu E. M., Cowie L. L., Songaila A., Barger A. J., Rosenwasser B., Wold I. G. B., 2016, ApJ, 825, L7
Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27
Hunter J. D., 2007, Computi. Sci. Eng., 9, 90
Izotov Y. I., Orlitová I., Schaerer D., Thuan T. X., Verhamme A., Guseva N. G., Worseck G., 2016, Nature, 529, 178
Izotov Y. I., Worseck G., Schaerer D., Guseva N. G., Thuan T. X., Fricke Verhamme A., Orlitová I., 2018, MNRAS, 478, 4851
Kaifu N. et al., 2000, PASJ, 52, 1
Kakiichi K. et al., 2018, MNRAS, 479, 43 (Paper I)
Kashino D., Lilly S. J., Shibuya T., Ouchi M., Kashikawa N., 2020, ApJ, 888, 6
Kulkarni G., Worseck G., Hennawi J. F., 2019, MNRAS, 488, 1035
Le Fèvre O. et al., 2015, A&A, 576, A79
Livermore R., Finkelstein S. L., Lotz J. M., 2017, ApJ, 835, 113
McGreer I. D., Mesinger A., D'Odorico V., 2015, MNRAS, 447, 499
McQuinn M., Upton Sanderbeck P. R., 2016, MNRAS, 456, 47
McQuinn M., Peng Oh S., Faucher-Giguère C.-A., 2011, ApJ, 743, 82
Mainali R. et al., 2018, MNRAS, 479, 1180
Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018a, ApJ, 856, 2
Mason C. A. et al., 2018b, ApJ, 857, L11
Mason C. A. et al., 2019, MNRAS, 485, 3947
Matthee J., Sobral D., Darvish B., Santos S., Mobasher B., Paulino-Afonso A., Röttgering H., Alegre L., 2017, MNRAS, 472, 772
Matthee J., Sobral D., Gronke M., Paulino-Afonso A., Stefanon M., Röttgering H., 2018, A&A, 619, A136
Mazzucchelli C. et al., 2017, ApJ, 849, 91
Mesinger A., 2010, MNRAS, 407, 1328
Meyer R. A., Bosman S. E. I., Kakiichi K., Ellis R. S., 2019, MNRAS, 483, 19 (Paper II)
Miralda-Escude J., Haehnelt M., Rees M. J., 2000, ApJ, 530, 1
Miyazaki S. et al., 2002, PASJ, 54, 833
Morselli L. et al., 2014, A&A, 568, A1
Naidu R. P., Tacchella S., Mason C. A., Bose S., Oesch P. A., Conroy C., 2020, ApJ, 892, 109
Nakajima K., Ellis R. S., Iwata I., Inoue A. K., Kusakabe H., Ouchi M., Robertson B. E., 2016, ApJ, 831, L9
Nakajima K., Fletcher T., Ellis R. S., Robertson B. E., Iwata I., 2018, MNRAS, 477, 2098
Nelson D. et al., 2018, MNRAS, 475, 624
Newman J. A. et al., 2013, ApJS, 208,
Oesch P. A., Bouwens R. J., Illingworth G. D., Labbé I., Stefanon M., 2018, ApJ, 855, 105
Oke J. B. B. J., 1974, ApJS, 27, 21
Ono Y. et al., 2012, ApJ, 744, 83
Ono Y. et al., 2018, PASJ, 70, S10
Ouchi M. et al., 2009, ApJ, 706, 1136
Ouchi M. et al., 2018, PASJ, 70, S13
Parsa S., Dunlop J. S., McLure R. J., 2018, MNRAS, 474, 2904
Pawlik A. H., Schaye J., van Scherpenzeel E., 2009, MNRAS, 394, 1812
Pentericci L. et al., 2018, A&A, 619, A147
Planck Collaboration et al., 2018, preprint (arXiv:1807.06209)
Prochaska J. X. et al., 2019, Astrophysics Source Code Library, record ascl.net:1911.004

Robertson B. E. et al., 2013, ApJ, 768, 71
Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, ApJ, 802, L19
Schenker M. A., Stark D. P., Ellis R. S., Robertson B. E., Dunlop J. S., McLure R. J., Kneib J.-P., Richard J., 2012, ApJ, 744, 179
Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., Erb D. K., 2006, ApJ, 651, 688
Songaila A., Hu E. M., Barger A. J., Cowie L. L., Hasinger G., Rosenwasser B., Waters C., 2018, ApJ, 859, 91
Soto K. T., Lilly S. J., Bacon R., Richard J., Conseil S., 2016, MNRAS, 458, 3210
Stark D. P., Ellis R. S., Chiu K., Ouchi M., Bunker A., 2010, MNRAS, 408, 1628
Stark D. P., Ellis R. S., Ouchi M., 2011, ApJ, 728, L2
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
Steidel C. C., Bogosavljević M., Shapley A. E., Reddy N. A., Rudie G. C., Pettini M., Trainor R. F., Strom A. L., 2018, ApJ, 869, 123
Tang M., Stark D. P., Chevallard J., Charlot S., 2019, MNRAS, 489, 2572
Tanvir N. R. et al., 2019, MNRAS, 483, 5380
Tasca L. A. M. et al., 2017, A&A, 600, A110
The Astropy Collaboration A., 2013, A&A, 558, 9
The Astropy Collaboration A., 2018, AJ, 156, 19
Vanzella E. et al., 2016, ApJ, 821, L27
Vanzella E. et al., 2017, ApJ, 842, 47
van der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Engi. 13, 22
Venemans B. P. et al., 2013, ApJ, 779,
Verhamme A. et al., 2018, MNRAS, 478, L60
Virtanen P. et al., 2019, Nat. Methods, 17, 261
Weilbacher P. M., Streicher O., Urrutia T., Jarno A., Pécontal-Rousset A., Bacon R., Böhm P., 2012, in Radziwill N. M., Chiozzi G., eds, Proc. SPIE Conf. Ser. Vol. 8451, Software and Cyberinfrastructure for Astronomy II. SPIE, Bellingham, p. 9
Weilbacher P. M., Streicher O., Urrutia T., Pécontal-Rousset A., Jarno A., Bacon R., 2015, in Manset N., Forshay P., eds, ASP Conf. Ser. Vol. 485, Astronomical Data Analysis Software and Systems XXIII. Astron. Soc. Pac., San Francisco, p. 451
Willott C. J. et al., 2007, AJ, 134, 2435
Worseck G. et al., 2014, MNRAS, 445, 1745

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Paper_III_MNRAS_OnlineAppendix_FINAL.pdf

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: SUMMARY OF ALL DETECTIONS AND PLOTS FROM DEIMOS

We present the DEIMOS spectroscopic confirmation of new LBGs in the field of J0836 (Fig. A1, online material) and J1030 (Fig. A2, online) used in this work for the cross-correlations. We leave the presentation and analysis of the three objects detected in the near-zone of J0836 to Bosman et al. (2019). Table A1 lists the LBG detections with their coordinates, redshift, Lyman-α FWHM, and corrected redshift.

Table A1. Summary of the detected LBGs in the DEIMOS fields.

Quasar	RA	Dec.	z_{peak}	FWHM	z_{corr}	r (mag)	i (mag)	z (mag)	M_{UV}
J0836	129.09106	1.00954	5.283	95.812	5.281	>27.62	26.35	25.33	−21.16
J1030	157.71105	5.36851	5.508	92.518	5.507	>27.5	25.54	23.95	−22.61
	157.58161	5.46687	5.791	66.852	5.790	>27.50	24.95	23.41	−23.23
	157.58308	5.44516	5.481	69.325	5.480	>27.50	26.51	25.12	−21.43
	157.67004	5.45504	5.712	176.514	5.709	>27.50	26.13	25.18	−21.44
	157.73887	5.46775	5.612	137.348	5.610	>27.50	>26.80	25.45	−21.13
	157.70962	5.36157	5.692	223.848	5.688	>27.50	26.39	25.19	−21.42
	157.52691	5.37737	5.352	118.119	5.351	>27.50	26.38	25.18	−21.33
	157.56116	5.34611	5.446	186.830	5.443	>27.50	25.93	25.49	−21.05

APPENDIX B: INDIVIDUAL LAE DETECTIONS WITH *MUSE* IN THE LYMAN-α FOREST OF OUR QUASARS

We present a summary of all detected LAEs in the redshift range of the Lyman-α forest of the nearby quasar in Table B1. We adopt an identification scheme where each LAE is named 'JXXXX_NBYYYY', where XXXX denotes the hours and minutes of the RA coordinates of the central quasar and YYYY the rounded wavelength of the narrowband frame in which *MUSE*LET or LSDCAT found the highest signal of the detection, which is often very close to the wavelength of the emission peak. Individual plots similar to Fig. 4 for each LAE can be found online in Fig. B1 (J0305, online), B2 (J1030, online), B3 (J1526, online), B4 (J2032, online), B5 (J2100, online), and B6 (J2329, online). Finally, we provide an example of common misdetections that are removed by visual inspection in Fig. B7 (online) such as low-redshift O II or continuum emitters, bright nearby foreground objects or defects or cosmic rays impacting only one of the exposures.

Table B1. Summary of the detected LAEs in the *MUSE* fields (and in the suitable redshift range for the cross-correlation). The last column gives the corrected redshift using the method of Verhamme et al. (2018), as described in Section 2.4.

ID	RA	Dec.	λ_{Lya}	z_{Lya}	FWHM (km s^{-1})	z_{corr}
J0305_NB8032	46.32776	−31.84569	8034.7	5.607	186.7	5.604
J0305_NB8609	46.31154	−31.85152	8609.7	6.081	174.2	6.078
J0305_NB8612	46.31095	−31.85202	8612.2	6.084	174.2	6.081
J1030_NB7707	157.61238	5.40784	7707.2	5.340	97.3	5.339
J1030_NB7927	157.61109	5.41578	7927.2	5.520	236.5	5.516
J1030_NB7942	157.61054	5.40995	7942.2	5.533	141.6	5.531
J1030_NB8177	157.61534	5.40556	8177.2	5.725	229.3	5.722
J1030_NB8202	157.61366	5.41512	8202.2	5.746	228.6	5.742
J1030_NB8220a	157.62069	5.41484	8220.9	5.761	228.1	5.758
J1030_NB8220b	157.61321	5.41900	8220.9	5.761	228.1	5.758
J1526_NB8476	231.66377	−20.83180	8475.9	5.972	88.5	5.971
J1526_NB8874	231.65771	−20.82652	8874.7	6.299	169.0	6.296
J2032_NB8396	308.04785	−21.23293	8396.4	5.907	134.0	5.905
J2032_NB8524	308.04240	−21.22620	8523.9	6.012	132.0	6.010
J2032_NB8525	308.03598	−21.23630	8525.2	6.013	132.0	6.011
J2100_NB7454	315.23399	−17.26017	7454.8	5.132	150.9	5.130
J2100_NB7678	315.23219	−17.26062	7678.6	5.316	146.5	5.314
J2100_NB8146	315.22375	−17.25901	8146.1	5.701	230.2	5.697
J2100_NB8419	315.22404	−17.26045	8419.8	5.925	133.6	5.923
J2329_NB8372	352.28913	−3.04041	8372.7	5.887	134.4	5.885
J2329_NB8390	352.28769	−3.03636	8390.2	5.902	134.1	5.900

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