Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **485,** 3947–3969 (2019)
Advance Access publication 2019 March 05

doi:10.1093/mnras/stz632

Inferences on the timeline of reionization at $z \sim 8$ from the KMOS Lens-Amplified Spectroscopic Survey

Charlotte A. Mason,[1,2★†] Adriano Fontana,[3] Tommaso Treu,[1] Kasper B. Schmidt,[4] Austin Hoag,[1,5] Louis Abramson,[6] Ricardo Amorin,[7,8] Maruša Bradač,[5] Lucia Guaita,[3,9] Tucker Jones,[5] Alaina Henry,[10] Matthew A. Malkan,[1] Laura Pentericci,[3] Michele Trenti[11,12] and Eros Vanzella[13]

[1]*Department of Physics and Astronomy, UCLA, Los Angeles, CA 90095-1547, USA*
[2]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden St, Cambridge, MA 02138, USA*
[3]*INAF Osservatorio Astronomico di Roma, Via Frascati 33, I-00040 Monteporzio (RM), Italy*
[4]*Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany*
[5]*Department of Physics, University of California, Davis, CA 95616, USA*
[6]*Carnegie Observatories, 813 Santa Barbara St., Pasadena, CA 91101-1292, USA*
[7]*Instituto de Investigación Multidisciplinar en Ciencia y Tecnología, Universidad de La Serena, Raúl Bitrán 1305, La Serena, Chile*
[8]*Departamento de Física y Astronomía, Universidad de La Serena, Av. Juan Cisternas 1200 Norte, La Serena, Chile*
[9]*Núcleo de Astronomía, Facultad de Ingeniería, Universidad Diego Portales, Av. Ejército 441, Santiago, Chile*
[10]*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*
[11]*School of Physics, University of Melbourne, Parkville, 3010 Victoria, Australia*
[12]*ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D)*
[13]*INAF – OAS, Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Gobetti 93/3, I-40129 Bologna, Italy*

Accepted 2019 March 1. Received 2019 March 1; in original form 2018 May 31

ABSTRACT

Detections and non-detections of Lyman alpha (Lyα) emission from $z > 6$ galaxies (<1 Gyr after the big bang) can be used to measure the timeline of cosmic reionization. Of key interest to measuring reionization's mid-stages, but also increasing observational challenge, are observations at $z > 7$, where Lyα redshifts to near infra-red wavelengths. Here we present a search for $z > 7.2$ Lyα emission in 53 intrinsically faint Lyman Break Galaxy candidates, gravitationally lensed by massive galaxy clusters, in the KMOS Lens-Amplified Spectroscopic Survey (KLASS). With integration times of ∼7–10 h, we detect no Lyα emission with signal-to-noise ratio (S/N) > 5 in our sample. We determine our observations to be 80 per cent complete for 5σ spatially and spectrally unresolved emission lines with integrated line flux $>5.7 \times 10^{-18}$ erg s^{-1} cm^{-2}. We define a photometrically selected sub-sample of 29 targets at $z = 7.9 \pm 0.6$, with a median 5σ Lyα EW limit of 58 Å. We perform a Bayesian inference of the average intergalactic medium (IGM) neutral hydrogen fraction using their spectra. Our inference accounts for the wavelength sensitivity and incomplete redshift coverage of our observations, and the photometric redshift probability distribution of each target. These observations, combined with samples from the literature, enable us to place a lower limit on the average IGM neutral hydrogen fraction of > 0.76 (68 per cent), > 0.46 (95 per cent) at $z \sim 8$, providing further evidence of rapid reionization at $z \sim 6$–8. We show that this is consistent with reionization history models extending the galaxy luminosity function to $M_{\rm UV} \lesssim -12$, with low ionizing photon escape fractions, $f_{\rm esc} \lesssim 15$ per cent.

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

1 INTRODUCTION

The reionization of intergalactic hydrogen in the universe's first billion years is likely linked to the formation of the first stars

★ E-mail: cmason@astro.ucla.edu
† Hubble Fellow

and galaxies: considered to be the primary producers of hydrogen-ionizing photons (e.g. Bouwens et al. 2003; Lehnert & Bremer 2003; Bunker et al. 2004; Yan & Windhorst 2004; Shull et al. 2012; Bouwens et al. 2015). Accurately measuring the timeline of reionization enables us to constrain properties of these first sources (e.g. Robertson et al. 2013, 2015; Greig & Mesinger 2017).

Measurements of the reionization timeline are challenging, however, due to the rarity of bright quasars at $z > 6$ (Fan et al. 2001; Manti et al. 2016; Parsa, Dunlop & McLure 2018), which have historically provided strong constraints on the end stages of reionization (e.g. Fan et al. 2006; McGreer, Mesinger & D'Odorico 2014; Bañados et al. 2017; Greig et al. 2017). In the coming decade, 21 cm observations are expected to provide information about the $z > 6$ intergalactic medium (IGM) and the nature of the first galaxies (e.g. Liu & Parsons 2016; Mirocha, Furlanetto & Sun 2016), but current progress has been driven by observations of Lyman alpha (Lyα) (rest-frame 1216 Å) emission in galaxies, using near infra-red (NIR) spectroscopy.

Lyα is a highly resonant line, and strongly scattered by intervening neutral hydrogen as it travels to our telescopes. Whilst young star-forming galaxies, selected with a Lyman Break (Lyman Break Galaxies – LBGs) show Lyα emission in abundance up to $z \sim 6$ (e.g. Hayes et al. 2011; Stark, Ellis & Ouchi 2011; Curtis-Lake et al. 2012; Cassata et al. 2015; De Barros et al. 2017), at higher redshifts the fraction of galaxies detected with Lyα emission, and the scale length of the Lyα rest-frame equivalent width (EW) distribution, decreases rapidly (e.g. Fontana et al. 2010; Pentericci et al. 2011; Caruana et al. 2012; Ono et al. 2012; Treu et al. 2012, 2013; Caruana et al. 2014; Faisst et al. 2014; Pentericci et al. 2014; Schenker et al. 2014; Tilvi et al. 2014; Jung et al. 2018). This rapid decline of detected Lyα emission is most plausibly due to absorption in an increasingly neutral IGM (Dijkstra, Mesinger & Wyithe 2011; Dijkstra 2014; Mesinger et al. 2015).

Large spectroscopic surveys of LBG candidates are being assembled out to $z \sim 7$ (Pentericci et al. 2011; Pentericci et al. 2014; Pentericci et al. 2018; Hoag et al. 2019) but exploring the earliest stages of reionization requires us to observe Lyα at even higher redshifts. Only a handful of Lyα emitters have been confirmed at $z \gtrsim 7.5$ (Oesch et al. 2015; Zitrin et al. 2015; Roberts-Borsani et al. 2016; Hoag et al. 2017; Stark et al. 2017), where the dominance of sky emission in the NIR makes observations of faint sources even more challenging. Additionally, because Lyα emission can be spatially extended and/or offset from the UV continuum emission (Wisotzki et al. 2016; Leclercq et al. 2017), it is likely that slit-based spectroscopy is not capturing the full Lyα flux. Hence, the observed decline in Lyα emission at $z > 6$ could be partially due to redshift-dependent slit-losses as well as reionization.

In this paper we present a search for $z \gtrsim 7.2$ Lyα emission in NIR spectroscopy of 53 intrinsically faint LBG candidates ($M_{\rm UV} \gtrsim -20$), gravitationally lensed behind six massive galaxy clusters, including four of the Frontier Fields (Lotz et al. 2017), selected from the Grism Lens-Amplified Survey from Space (hereafter GLASS; Schmidt et al. 2014a; Treu et al. 2015). We also present observations of C IV emission in three images of a previously confirmed multiply imaged $z = 6.11$ galaxy (Balestra et al. 2013; Boone et al. 2013; Monna et al. 2014)

The observations presented in this work were carried out with the ESO Very Large Telescope (*VLT*) *K*-band Multi Object Spectrometer (hereafter KMOS; Sharples et al. 2013). This work presents the first results of $z > 3.8$ observations with KMOS. KMOS is an integral field unit (IFU) instrument, and we demonstrate here that our observations are more complete to spatially extended and/or offset Lyα emission than traditional slit spectrographs.

We use our new deep spectroscopic observations to infer the average IGM neutral hydrogen fraction ($\overline{x}_{\rm H\ I}$) at $z \sim 8$. Mason et al. (2018a, hereafter M18a) presented a flexible Bayesian framework to directly infer $\overline{x}_{\rm H\ I}$ from detections and non-detections of Lyα emission from LBGs. The framework combines realistic inhomogeneous reionization simulations and models of galaxy properties. That work measured $\overline{x}_{\rm H\ I} = 0.59^{+0.11}_{-0.15}$ ($16-84$ per cent confidence intervals) at $z \sim 7$. Building on Treu et al. (2012) and M18a, we extend this framework to use the full spectra obtained in our observations for the Bayesian inference, accounting for the incomplete wavelength coverage and spectral variation of the noise, and marginalizing over emission linewidth. Our framework uses the photometric redshift probability distribution, obtained from deep photometry including new Spitzer/IRAC data, of each object to robustly account for uncertainties in redshift determination.

The paper is structured as follows: Section 2 describes our KMOS observations and the target selection from the GLASS parent sample; Section 3 describes the search for Lyα emission in our KMOS data cubes, and the purity and completeness of our survey; and Section 4 describes the Bayesian inference of the neutral fraction and presents our limit on $\overline{x}_{\rm H\ I}$ at $z \sim 8$. We discuss our findings in Section 5, including an assessment of the performance of KMOS for background-limited observations using our deep observations, and summarize in Section 6.

We use the Planck Collaboration VIII (2015) cosmology where $(\Omega_\Lambda, \Omega_{\rm m}, \Omega_{\rm b}, n, \sigma_8, H_0) = (0.69, 0.31, 0.048, 0.97, 0.81, 68\ {\rm km\ s^{-1}\ Mpc^{-1}})$. All magnitudes are given in the AB system.

2 OBSERVATIONS

2.1 The KMOS Lens-Amplified Spectroscopic Survey

The KMOS Lens-Amplified Spectroscopic Survey (KLASS) is an ESO VLT KMOS Large Program (196.A-0778, PI: A. Fontana), which targeted the fields of six massive galaxy clusters: Abell 2744 (hereafter A2744); MACS J0416.1-2403 (M0416); MACS J1149.6+2223 (M1149); MACS J2129.4-0741 (M2129); RXC J1347.5-1145 (RXJ1347); and RXC J2248.7-4431 (RXJ2248, aka Abell S1063). A2744, M0416, M1149, and RXJ2248 are all Frontier Fields (hereafter HFF; Lotz et al. 2017). Observations were carried out in Service Mode during Periods 96−99 (2015 October–2017 October).

KMOS is a multi-object IFU spectrograph, with 24 movable IFUs, split between three different spectrographs (Sharples et al. 2013). Each IFU is $2\farcs8 \times 2\farcs8$ field of view, with pixel size $0\farcs2 \times 0\farcs2$, and 2048 pixels along the wavelength axis.[1]

The key science drivers of KLASS are:

(i) To probe the internal kinematics of galaxies at $z \sim 1-3$, with superior spatial resolution compared to surveys in blank fields (Mason et al. 2017).

(ii) To investigate $z \gtrsim 7$ Lyα emission from the GLASS sample, independently of the *Hubble Space Telescope* (*HST*) spectroscopic observations, providing validation and cross-calibration of the

[1]We use the following definitions for describing 3D spectra in this paper. Pixel: 2D spatial pixel (size $0\farcs2 \times 0\farcs2$). Spaxel: the 1D spectrum in a single spatial pixel (spanning the spectral range $\sim 1 - 1.35\,\mu$m, in 2048 spectral pixels). Voxel: 3D pixel in the data cube with both spatial and spectral indices.

results and enabling us to constrain the timeline and topology of reionization (Treu et al. 2012, 2013; Schmidt et al. 2016; Mason et al. 2018a).

Mason et al. (2017) addressed the first science driver by presenting spatially resolved kinematics in four of the six KLASS clusters from our early data, including five of the lowest mass galaxies with IFU kinematics at $z > 1$, and provided evidence of mass-dependent disc settling at high redshift (Simons et al. 2017). The KLASS kinematic data were combined with metallicity gradients from the *HST* GLASS data to enable the study of metallicity gradients as a diagnostic of gas inflows and outflows (Wang et al. 2016).

This paper addresses the second science driver by presenting our $z > 7$ candidate targets with complete exposures. We use the YJ observing band, giving us access to Lyα emission at $z \sim 7.2$–10.1.

The choice of an IFU instrument for high-redshift Lyα observations was motivated by indications that ground-based slit-spectroscopy measures lower Lyα flux than *HST* slit-less grism spectroscopy (Tilvi et al. 2016; Huang et al. 2016b; Hoag et al. 2017), which, as well as reionization, could contribute to the observed decline in Lyα emission at $z > 6$. Lyα emission can be spatially extended and/or offset from the UV continuum emission (Feldmeier et al. 2013; Momose et al. 2014; Wisotzki et al. 2016; Leclercq et al. 2017), so it is likely that slit-based spectrographs do not capture the full Lyα flux.

By using IFUs, our observations should be more complete to spatially extend and/or offset Lyα than traditional slit spectrographs. Mason et al. (2017) showed that only ∼60 per cent of emission line flux was contained in ∼0.″7 simulated slits (a typical slit-width used for Lyα observations; e.g. Hoag et al. 2017) on KMOS spectra, whereas the full flux is captured within the 2.″8 × 2.″8 KMOS field of view. Thus we expect most Lyα flux to be captured within the KMOS IFUs. The 2.″8 wide IFUs cover ∼14 proper kpc at $z \sim 8$, while the UV effective radii of galaxies at these redshifts is only $\lesssim 1$ proper kpc (Shibuya, Ouchi & Harikane 2015). We demonstrate in Section 3.3 that our KMOS observations have good completeness for spatially extended and/or offset Lyα emission.

2.2 Target selection

KLASS targets were selected from the GLASS survey[2] (Schmidt et al. 2014a; Treu et al. 2015), a large *HST* slit-less grism spectroscopy program. GLASS obtained spectroscopy of the fields of 10 massive galaxy clusters, including the HFF and 8 CLASH clusters (Postman et al. 2012). The Wide Field Camera 3 (WFC3) grisms G102 and G141 were used to cover the wavelength range 0.8–1.6 μm with spectral resolution $R \sim 150$. We refer the reader to Schmidt et al. (2014a) and Treu et al. (2015) for full details of GLASS.

KLASS observations aimed to provide the high spectral resolution necessary to measure the purity and completeness of the grism spectra, to measure lines that were unresolved in *HST*, and to obtain velocity information for $z \sim 1$ targets that the low-resolution grisms cannot provide. In combination with additional GLASS follow-up observations at Keck (Huang et al. 2016b; Hoag et al. 2017; Hoag et al. 2019), we will address the purity and completeness of the *HST* grisms in a future work. In this work we present our high-redshift candidate targets and our inferences about reionization obtained from the KLASS data.

[2]http://glass.astro.ucla.edu

Two categories of high-redshift candidate KLASS targets were selected from the GLASS data:

(i) Category 1: 14 objects with marginal (signal-to-noise ratio [S/N] ∼ 3) candidate Lyα emission in the *HST* grisms, identified by visual inspection of the GLASS data, which fall within the KMOS YJ spectral coverage (∼1–1.35 μm). Four candidates were selected from a list of candidates in a preliminary census of GLASS data by Schmidt et al. (2016). The remaining candidates were selected in a similar method to the procedure followed by Schmidt et al. (2016).

(ii) Category 2: 39 LBG candidates selected with $z_{\rm phot} > 7.2$, from an ensemble of photometric catalogues described by Schmidt et al. (2016). This includes three LBGs, which were spectroscopically confirmed via sub-mm emission lines after our survey began: A2744_YD4 (A2744_2248 in this paper), at $z = 8.38$ (Laporte et al. 2017a, discussed in Sections 4.2 and 4.3), M0416_Y1 (M0416_99 in this paper), at $z = 8.12$ (Tamura et al. 2018, discussed in Section 4.2), and M1149_JD1, at $z = 9.11$ (Hashimoto et al. 2018, discussed in Section 5.2).

An additional three targets were multiple images of the $z = 6.11$ system in RXJ2248 (Balestra et al. 2013; Boone et al. 2013; Monna et al. 2014) where we targeted C IVλ1548,1551 emission. This object is discussed in Appendix A.

We ranked objects in order of the number of inspectors who reported a candidate emission line for our Category 1 targets, and then by the number of independent photometric catalogues the target appeared in (for both categories). Our observations were planned prior to the release of the full HFF data sets, so the photometric catalogues we used to select candidates did not contain the full photometry now available. In particular, deep Spitzer/IRAC data did not exist, which can be useful for distinguishing between high-redshift star-forming galaxies and $z \sim 1$–2 passive galaxies. Nor were sophisticated intra-cluster light (ICL) removal techniques developed at that point (e.g. Merlin et al. 2016; Livermore, Finkelstein & Lotz 2017; Morishita et al. 2017).

Thus our LBG selection was heterogeneous, but in this paper we now add in the new deep and extended photometry to define a homogeneous photometric selection. We expect some faint candidates may have been spurious in the initial photometry and may not appear in the final deep catalogues. Additionally we expect that with the inclusion of Spitzer/IRAC photometry some of the objects originally selected to be $z > 7$ may be low-redshift contaminants. In our reionization analysis we use catalogues built using the final HFF data sets to define a selection function for a photometrically selected sample for our inference (described in Section 4.2). We demonstrate that this KLASS sub-sample is not a biased sample of the final parent catalogues in Appendix B.

The GLASS median 1σ flux limit is 5×10^{-18} erg s^{-1} cm^{-2} (Schmidt et al. 2016), and we tried to be as inclusive as possible when assigning candidates to the KMOS IFUs from the GLASS parent catalogue. Most of the candidates were only 3σ significance in GLASS data and we were aiming to provide confirmation of those tentative targets our deep KMOS observations – though for our ground-based observations, at least 50 per cent of the wavelength range is dominated by sky emission and the low spectral resolution ($R \sim 100$) of the *HST* grisms means that the line position is uncertain by ±25 Å so the lines could be in bad sky regions. Additionally, in planning our observations we likely overestimated the sensitivity of KMOS YJ using the online exposure time calculator, especially at the blue end of the detectors. We discuss this in more detail in Section 5.4.

Table 1. KLASS cluster targets.

Cluster	Run ID	DIT [s]	NDITs[a]	Exposure [hrs]	Number of targets	
					Category 1	Category 2
A2744[b,c]	A	900	25	6.25	3	7
M0416[c]	B	900	43	10.75	2	5
M1149[c]	C	900	40	10.00	2	6
M2129	E	450	85	10.625	3	7
RXJ1347	D	450	88	11.00	3	8
RXJ2248[d]	F	300	93	7.75	1	6

Notes: [a]The number of Detector Integration Times (DITs) used in this analysis: we discarded DITs if the seeing was $>0\farcs8$ as measured by stars observed in each DIT. The total exposure time = DIT × NDITs. [b]We had to discard our initial 4 h of observations of A2744 due to irreparable flexure issues due to rotating the instrument between science and sky DITs. All subsequent observations were performed with no rotation of the instrument between science and sky DITs. [c]Target selection in these clusters was primarily done from preliminary versions of the ASTRODEEP catalogues (Castellano et al. 2016; Merlin et al. 2016; Di Criscienzo et al. 2017), which did not include Spitzer/IRAC photometry. [d]Due to a high proper motion reference star, some of the observations of RXJ2248 were taken at a slight offset from the required target centre, reducing the total exposure at that position. RXJ2248 also included 3 $z = 6.11$ targets (Appendix A).

As we describe below in Section 3.4, our KLASS observations are ∼80 per cent complete for lines with flux $>5.7 \times 10^{-18}$ erg s^{-1} cm^{-2}, which suggests that we should have confirmed the majority of the 14 GLASS candidate Lyα emission targets we observed (Category 1, described in Section 2.2). However, we did not detect any emission in the cubes containing these candidates, suggesting that at least some of the GLASS candidates were spurious noise peaks in the *HST* grisms. A more thorough comparison of the GLASS *HST* grism and ground-based follow-up observations (including KLASS and Keck observations; Huang et al. 2016b; Hoag et al. 2017; Hoag et al. 2019) to recover the grism purity and completeness will be left to a future work.

53 $z_{\rm phot} > 7.2$ candidate targets across the six clusters were assigned to 51 KMOS IFUs (two IFUs contained two nearby candidates). The cluster list and number of high-redshift candidate targets per cluster are shown in Table 1.

2.3 KLASS observing strategy and reduction

KLASS observations were carried out with KMOS YJ (∼1–1.35 μm). The spectral resolution $R \sim 3400$ is sufficient to distinguish Lyα from potential low-redshift contaminants with the [O II] λ3726, 3729 emission doublet at $z \sim 2$.

Observations were carried out in a service mode and executed in 1 h observing blocks with repeating ABA science–sky–science integration units (detector integration times – DITs). Each observing block comprised 1800 s of science integration, and 900 s on sky. Pixel dither shifts were included between science frames. A star was observed in 1 IFU in every observing block to monitor the point spread function (PSF) and the accuracy of dither offsets. The PSF was well-described by a circular Gaussian and the median seeing of our observations was FWHM $\sim 0\farcs6$.

In each cluster, the three top priority targets were observed for 1.5 × the average exposure time by assigning 2 IFUs per target and nodding between them during A and B modes.

2.4 Reduction

Data were reduced using the ESO KMOS pipeline v.1.4.3 (Davies et al. 2013). We apply a correction for known readout channel level offsets before running the pipeline. We run the pipeline including optimized sky subtraction routines `sky_tweak` (Davies 2007) and `sky_scale`.

To improve the sky subtraction in the pipeline-reduced 'A–B' cubes, we produced master sky residual spectra by median combining all IFUs on each spectrograph into a 1D master sky residual spectrum for each DIT, excluding cubes containing $z \lesssim 2$ targets with bright emission lines and/or continua. We then subtract these 1D sky residual spectra from the 'A–B' cubes on the same spectrograph for each DIT, rescaling the 1D spectra in each spaxel so that the resulting 1D spectrum in each spaxel is minimized.

Similar techniques to improve sky subtraction are described by Stott et al. (2016). This method worked best for our survey design. We note that this method performed better than in-IFU sky residual subtraction (i.e. subtracting a median sky residual spectrum produced from 'empty' spaxels in each IFU) as it preserved emission line flux in the modestly sized KMOS IFUs.

Cube frames from each DIT are combined via sigma clipping, using spatial shifts determined by the position of the star observed in the same IFU in each DIT, to produce flux and noise cubes. For this work we used only frames with seeing $\leq 0\farcs8$ (as measured by the star observed in our science frames). The median seeing was $\sim 0\farcs6$. DIT length, observing pattern, and total integration times used for this paper are listed in Table 1. We note that due to the failure of one of the KMOS arms, no star was observed in the A2744 observations. We used a bright $z \sim 1$ target to estimate the dither offsets for this cluster.

For pure Gaussian noise, the pixel distribution of S/N should be normally distributed. We tested this by selecting non-central regions of cubes containing high-redshift candidate targets (i.e. where we expect very little source flux) and found the pixel distribution of S/N to have standard deviation >1, suggesting that the noise is underestimated by the pipeline.

We therefore apply a rescaling to the noise of the combined cubes. We create an average 1D noise spectrum in a single spaxel for each cluster by taking the root mean square (RMS) at every wavelength of every spaxel from the cubes containing high-redshift candidate targets. Since the cubes are predominantly noise, taking the RMS of the flux at each wavelength across multiple cubes should give the appropriate noise. We find that this RMS spectrum is ∼1.2 × higher than the pipeline average 1D noise spectrum (taking the average of the noise cubes across the same set of high-redshift targets). We

rescale the pipeline noise in every cube by this ratio of the cluster RMS noise spectrum to the cluster average noise spectrum.

Finally, we rescale the noise in each cube by a constant value so that the S/N distribution of all pixels has standard deviation 1 (clipping pixels within 99.9 per cent to remove spurious peaks). We find that the S/N distribution is well-described by a Gaussian distribution, with non-Gaussian tails only beyond the $\gtrsim 7\sigma$ confidence regions, due to bad sky subtraction residuals.

3 EMISSION LINE SEARCH, PURITY, AND COMPLETENESS

In this section we describe our search for Lyα emission in our KMOS observations. We give our algorithm for line detection in Section 3.1, and calculate the purity and completeness of our observations in Sections 3.2 and 3.3. Given that we detect no convincing Lyα emission lines in our sample, we present our flux and EW upper limits in Section 3.4.

3.1 Emission line detection technique

To search for emission lines in the KMOS cubes, robustly determine the completeness and purity of our survey, and determine the flux limits of our observations, we used the following algorithm to flag potential lines:

(i) Create a circular aperture with $r = 2\sigma_{\rm PSF} \sim 0\farcs5 \sim 2.5$ pixels (using our median seeing $\rm FWHM_{psf} = 0\farcs6$), which will capture 86 per cent of the total flux for spatially unresolved emission line at the centre of the aperture.

(ii) Sum the flux, and take the RMS of the noise of all spaxels in the aperture to create 1D data and noise spectra.

(iii) Rescale the 1D noise spectrum so the S/N in all pixels (excluding the 0.1 per cent most extreme S/N values) is Normal.

(iv) Scan through in wavelength and flag a detection if three adjacent wavelength pixels have $S/N > 3$. This corresponds to an $S/N \gtrsim 5$ detection of the integrated line flux.

(v) Iterate over 25 apertures centred within three pixels ($0\farcs6$) of the IFU centre, i.e. $x = [-3, -1.5, 0, 1.5, 3]$, $y = [-3, -1.5, 0, 1.5, 3]$ where $(x, y) = (0, 0)$ is the IFU centre.

Our search covers $\sim 25 \times 2000 = 50\,000$ potential emission line positions in each cube. As our detection threshold is 5σ, we would expect a false-positive rate of 6×10^{-7}, i.e. ∼0.03 false detections per cube for Gaussian noise. As discussed in Section 2.4, the S/N has small non-Gaussian tails due to sky subtraction residuals so we expect a slightly higher false detection rate than this.

3.2 Candidate emission lines and sample purity

We ran the detection algorithm described in Section 3.1 on the 54 cubes containing our high-redshift candidate targets (including the three cubes containing the $z = 6.11$ images). Nine unique candidate lines were flagged (combining candidates at the same wavelength identified in different apertures). Each of these candidate lines was then visually inspected to determine whether it was a true emission line or a spurious noise peak. For our inspections, we use both 1D spectra extracted in the detection apertures as well as 2D collapsed images of the candidate line obtained by summing cube voxels in the wavelength direction. The 2D images are helpful for determining plausible spatially compact emission from the uniform emission produced by sky residuals.

Our algorithm correctly identifies the C IVλ1551 emission at 11023.7 Å in the brightest image of the multiply imaged $z = 6.11$ system, demonstrating the depth of our KMOS observations and the fidelity of our algorithm. Another detection is flagged in this object at 13 358.6 Å but the emission appears diffuse and the wavelength is not consistent with other expected UV emission lines so we deem this spurious. We describe this object in more detail in Appendix A.

Of the remaining seven lines flagged, six are deemed to be spurious detections as they are at the spectral edges of the detector, or immediately adjacent to strong skylines and appear to have P-Cygni profiles, indicating extreme sky subtraction failures. Whilst it could be possible to add a cut to e.g. downweight flagged lines adjacent to skylines, given the relatively low spectral resolution of our observation ($R \sim 3400$), we were wary that many true emission lines could be overlapping with skylines, thus visual inspection was necessary. This is clearly demonstrated in our detection of C IV emission where both doublet components overlap with sky lines (see Fig. A1).

The remaining candidate emission line at 12 683.7 Å is spatially offset from the $z > 7$ LBG candidate in the cube. We determine the detected emission to be associated with a nearby (∼1.1 arcsec) galaxy with $z_{\rm phot} = 4.2$, which has bright continuum emission in the GLASS data. The candidate line appears in a particularly bad spectral region of telluric absorption, and we determine the detection to be due to inadequate continuum subtraction of the $z \sim 4$ source.

In our reductions we subtract a sky residual spectrum to minimize the flux in each spaxel of the high-redshift candidate cubes (Section 2.4). During that process most of the continuum emission from the $z \sim 4$ object was poorly subtracted by scaling the sky residual spectrum to high values. Some residual flux is left, which correlates with the positions of sky residuals. We note that the LBG candidate targeted in this IFU is not present in the final deep photometric catalogues and is excluded from our reionization (it was likely a spurious detection in the original shallow photometry; Section 4.2). We remove this cube from further analysis.

Thus we determine our algorithm has detected one real emission line, and seven spurious detections (excluding the $z \sim 4$ continuum object described above), allowing us to define the purity of our spectral sample:

$$P = 1 - \frac{N_{\rm spurious}}{N_{\rm pos}} \quad (1)$$

where $N_{\rm spurious} = 17$ is the total number of spurious flags (eight unique false detections that were sometimes flagged in multiple apertures) and $N_{\rm pos} = 101\,763 \times 25$ is the number of possible emission line positions in the 53 useful cubes, removing wavelength pixels not covered by certain detectors, in 25 apertures. We measure $P = (1-7) \times 10^{-6}$. Our spurious detection rate is ∼10 × higher than that expected for 5σ fluctuations in the noise, which was expected due to the non-Gaussian tail in our S/N distribution due to sky subtraction residuals. To verify that the S/N distribution is symmetrical, we also ran the detection algorithm to look for negative peaks ($\rm S/N \lesssim -5$) which should occur at the same rate. We found 12 flagged negative S/N detections, comparable to our 7 flagged spurious detections with positive S/N.

We ran the algorithm on our Category 1 sources with a lower S/N threshold: $\rm S/N > 2.5$ per wavelength pixel, corresponding to $\rm S/N \gtrsim 4$ in the integrated line. We found no convincing detections with this lower threshold and are thus unable to confirm any of the candidate GLASS emission lines. Given that most of the GLASS Lyα candidates were of low significance in the GLASS *HST* data,

these candidates may have been spurious noise peaks in the grism data.

In Section 4.2 below we list the Lyα flux and EW limits for our most likely $z_{\rm phot}$ LBGs candidates. We discuss our limits on other UV lines in Section 5.3.

3.3 Completeness

To evaluate the completeness of our emission line search, we carry out comprehensive Monte Carlo simulations: inserting simulated lines into cubes with varied total flux, spectral FWHM$_{\rm spec}$, spatial position, spatial extent FWHM$_{\rm spat}$, and wavelength, and testing whether they are detected by our detection algorithm (Section 3.1). Traditionally, these types of simulations are carried out by inserting simulated lines into real raw data and then running through the full reduction pipeline (Fontana et al. 2010; Pentericci et al. 2014; De Barros et al. 2017), however, due to the complexity of the KMOS pipeline that constructs 3D cubes from 2D frames we instead create simulated cubes and add Gaussian noise drawn from an average noise cube for each cluster, mimicking completeness simulations traditionally done in imaging.

We create simulated flux cubes with a 3D Gaussian emission line with varied properties and add noise to each voxel drawn from a Gaussian distribution with mean zero and standard deviation $\sigma_{x,y,\lambda}$ for each cluster. The $\sigma_{x,y,\lambda}$ cubes are constructed by taking the RMS at every voxel of all the final sky-subtracted cubes that do not contain bright $z \lesssim 2$ sources ($\sim$10 cubes per cluster). As each 'empty' cube is expected to be pure noise, taking the RMS at each voxel across the cubes should give an estimate of the noise per voxel, $\sigma_{x,y,\lambda}$.

We calculate completeness as a function of flux, spatial offset from the IFU centre, spectral linewidth, and spatial extent. For each simulation we vary the parameter of interest and wavelength, and fix the other three parameters. Our fiducial values for the parameters are: line flux $= 1 \times 10^{-17}$ erg s^{-1} cm^{-2}, observed line FWHM$_{\rm spec}$ = 4 Å (the spectral resolution, i.e. unresolved lines), line centred at the IFU centre, with source spatial extent FWHM$_{\rm spat} = 0$ arcsec (i.e. unresolved point source, the emission will have observed spatial extent with FWHM$_{\rm spat,tot} = \sqrt{\rm FWHM_{PSF}^2 + FWHM_{spat}^2}$). We draw 1000 realizations of an emission line with noise at every tested value of a parameter. The resulting completeness is the fraction of these simulated lines detected by our detection algorithm.

Our fiducial simulations assume Lyα emission will be spatially unresolved. These assumptions are reasonable for the intrinsically UV faint LBGs we are observing (Schmidt et al. 2016; Marchi et al. 2018). Typical slit spectrograph observations of Lyα emission centre slits on the UV continuum and use slit-widths $\sim$0.″7, thus in KLASS we are more complete to Lyα emission that may be spatially extended and/or offset from the UV continuum.

Fig. 1 shows the results of our completeness simulations for all clusters. We reach 80 per cent completeness over the full wavelength range for lines $\gtrsim 5.7 \times 10^{-18}$ erg s^{-1} cm^{-2}, centred within $<$0.″8 of the IFU centre and with intrinsic line FWHM$_{\rm spec}$ $\lesssim$ 250 km s^{-1}, assuming $z = 8$ to calculate FWHM$_{\rm spec}$ (median over all clusters). For wavelength ranges where the noise level is below the median across the whole spectrum, we reach 80 per cent completeness for 5σ lines $\gtrsim 3.2 \times 10^{-18}$ erg s^{-1} cm^{-2}, centred with $<$0.″9 of the IFU centre and with intrinsic line FWHM$_{\rm spec}$ $\lesssim$ 550 km s^{-1}. The completeness is fairly flat for Lyα spatial extent $\lesssim$ 0.″6 (total extent $\lesssim$ 0.″8) demonstrating our good completeness for spatially extended Lyα emission, with the normalization of the completeness as a function of FWHM$_{\rm spat}$ scaling with the completeness at a given total line flux.

3.4 Flux and equivalent width limits

To calculate average flux limits for each cluster, we take the average 3D noise spectrum for each cluster, $\sigma_{x,y,\lambda}$ (created by taking the RMS at every voxel across the $\sim$10 IFUs observing high-redshift candidates in each cluster). We then create a 1D noise spectrum, σ_λ, by summing the average noise at each wavelength pixel in a circular aperture with radius $r = 2\sigma_{\rm PSF}$ (where we use our median seeing FWHM$_{\rm PSF}$ = 0.″6). At each wavelength pixel, i the flux limit in erg s^{-1} cm^{-2} is given by

$$f_{{\rm lim},i} = 5 \times \frac{1}{1 - {\rm e}^{-\frac{r^2}{2\sigma_{\rm PSF}^2}}} \sqrt{\frac{2{\rm FWHM_{res}}}{\Delta\lambda}} \sigma_i \times \Delta\lambda \quad (2)$$

Here we obtain an estimate of the integrated noise for an emission line with observed FWHM$_{\rm res}$ = 4 Å or $\approx$110 km s^{-1} (the instrumental resolution), and use a threshold integrated S/N = 5. The term in the denominator accounts for the fact that the apertures only capture a fraction of the flux. For $r = 2\sigma_{\rm PSF}$ this results in a rescaling of 1.16. The spectral pixel width of KMOS YJ is $\Delta\lambda$ = 1.75 Å. The above calculation assumes the emission is spatially and spectrally unresolved by KMOS, which is reasonable given the expectation that Lyα emission from UV faint galaxies is likely to be more spatially compact and have lower linewidth than Lyα from UV bright galaxies (e.g. Schmidt et al. 2016; Marchi et al. 2018). We note that the flux limit for wider lines can be estimated as $f_{\rm lim} \propto \sqrt{\rm FWHM/4\,Å}$.

The 5σ flux limits for all clusters are shown in Fig. 2. The median flux limit is 4.5 $\times$ 10^{-18} erg s^{-1} cm^{-2}, and the range of medians for each cluster is 3.9–6.4 $\times$ 10^{-18} erg s^{-1} cm^{-2}.

Rest-frame Lyα equivalent widths are $W = (1+z)^{-1} f(\lambda)/f_{\rm cont}$, where $z = \lambda/\lambda_\alpha - 1$ (with λ_α = 1216 Å), and we define the continuum flux:

$$f_{\rm cont}(m, z) = f_0 10^{-0.4m} \frac{c}{\lambda_\alpha^2 (1+z)^2} \left(\frac{\lambda_{\rm UV}}{\lambda_\alpha}\right)^{-\beta-2} \quad (3)$$

where $f_0 = 3.631 \times 10^{-20}$ erg s^{-1} Hz^{-1} cm^{-2}, m is the apparent magnitude of the UV continuum, c is the speed of light, $\lambda_{\rm UV}$ is the rest-frame wavelength of the UV continuum (usually 1500 Å), and β is the UV slope. We assume $\beta = -2$, consistent with $z \sim 7$ observations (e.g. Stanway, McMahon & Bunker 2005; Blanc et al. 2011; Wilkins et al. 2011; Castellano et al. 2012; Bouwens et al. 2012, 2014). We use the magnitude measured in *HST* WFC3/F160W for the apparent magnitude (`automag`). Example EW limits for objects with a given apparent magnitude, using the RXJ1347 average flux limit, are plotted in Fig. 3.

4 REIONIZATION INFERENCE

In this section we describe the extension to the M18a Bayesian inference framework to include the full spectra, robustly including the uncertainties in redshift via the photometric redshift distribution (Section 4.1), and marginalizing over the linewidth of potential emission lines. Using the observations described above, we now define a clear selection function for a photometrically selected sample of LBGs within our survey (Section 4.2), and perform the inference of the IGM neutral fraction using these data (Section 4.3).

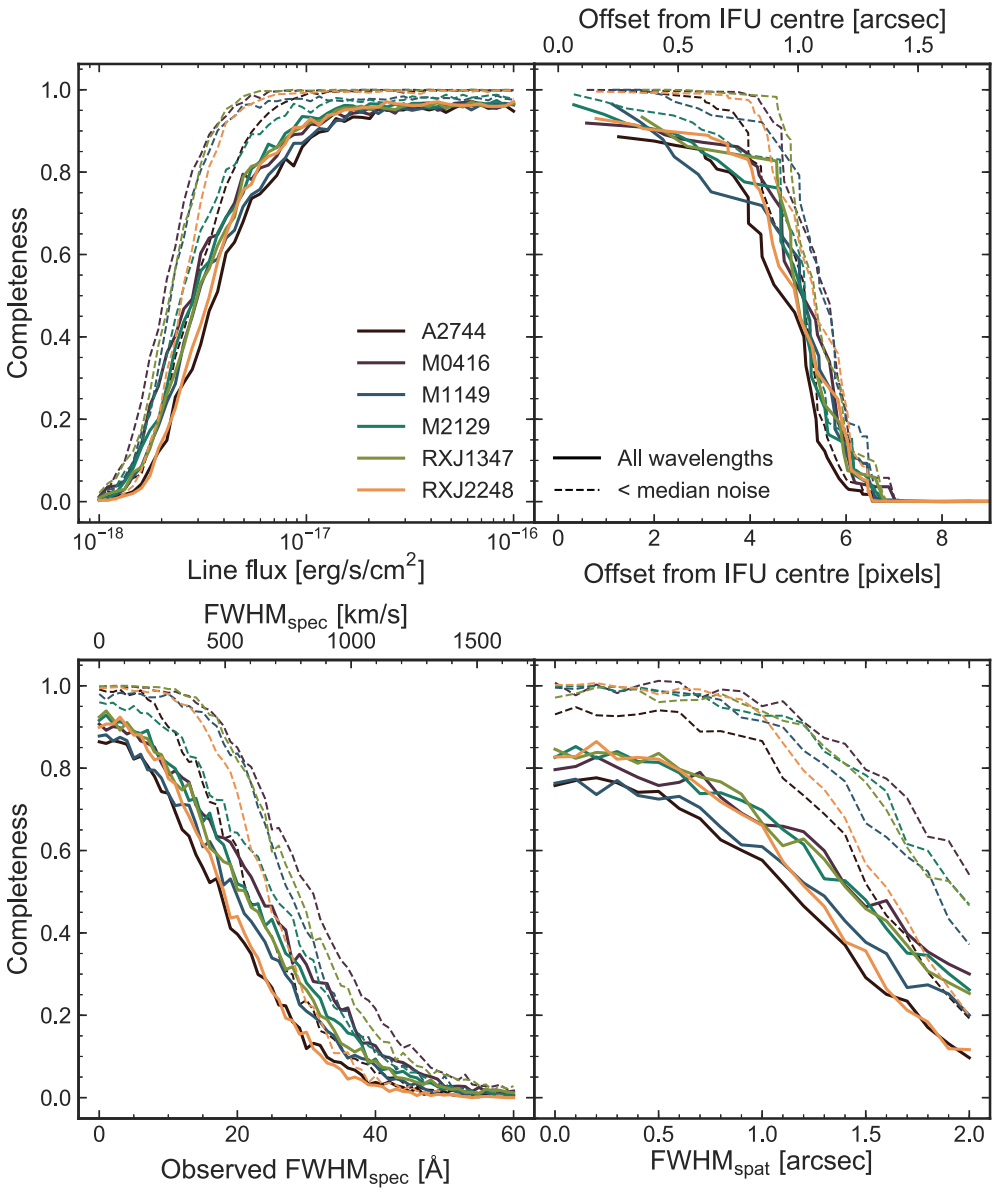

Figure 1. Completeness as a function of line flux (*top left*), spatial offset from IFU centre (*top right*), spectral linewidth (*lower left*) and spatial extent (*lower right*). Each colour corresponds to a separate cluster target. Dashed lines show the completeness across the entire wavelength range, and solid lines show the completeness in wavelength regions where the noise level is below the median. FWHM$_{spec}$ velocities were calculated assuming $z = 8$. Spatial extent, FWHM$_{spat}$ is the extent of the source (excluding the PSF). We create simulated lines with total spatial extent $\mathrm{FWHM}_{spat,tot} = \sqrt{\mathrm{FWHM}^2_{PSF} + \mathrm{FWHM}^2_{spat}}$. In each plot the parameter of interest and wavelength are varied, while the other parameters are held constant. The fiducial parameters are line flux $= 1 \times 10^{-17}$ erg s^{-1} cm^{-2}, observed FWHM$_{spec} = 4$ Å (i.e. unresolved), line centred at the IFU centre, and FWHM$_{spat} = 0''$ (i.e. unresolved point source).

4.1 Bayesian inference framework

To use our observations to make inferences about the neutral hydrogen fraction at $z \sim 8$, we use the method described by M18a. This forward-models the observed rest-frame Lyα EW distribution as a function of the neutral fraction and galaxy UV magnitude, $p(W \,|\, \overline{x}_{\rm HI}, M_{\rm UV})$, using a combination of reionization simulations with realistic inhomogeneous IGM structure (Mesinger, Greig & Sobacchi 2016), and empirical and semi-analytic models of galaxy properties.

The models assume the observed $z \sim 6$ Lyα EW distribution is the 'emitted' distribution (i.e. the distribution without IGM attenuation due to reionization) and use that to forward-model the observed distribution, including the impact of Lyα velocity offsets. Here, as in M18a, we use the recent comprehensive $z \sim 6$ Lyα EW observations from De Barros et al. (2017). We use the public Evolution of 21cm Structure (EoS) suite of reionization simulations described by Mesinger et al. (2015, 2016)[3] to generate Lyα optical depths along millions of sightlines in simulated IGM cubes for a grid of volume-averaged $\overline{x}_{\rm H\,I}$ values. As the size of ionized regions during reionization is expected to be nearly independent of redshift at fixed $\overline{x}_{\rm H\,I}$ (as there is little difference in the matter power spectrum from $z \sim 7-11$; McQuinn et al. 2007), we use the

[3] http://homepage.sns.it/mesinger/EOS.html

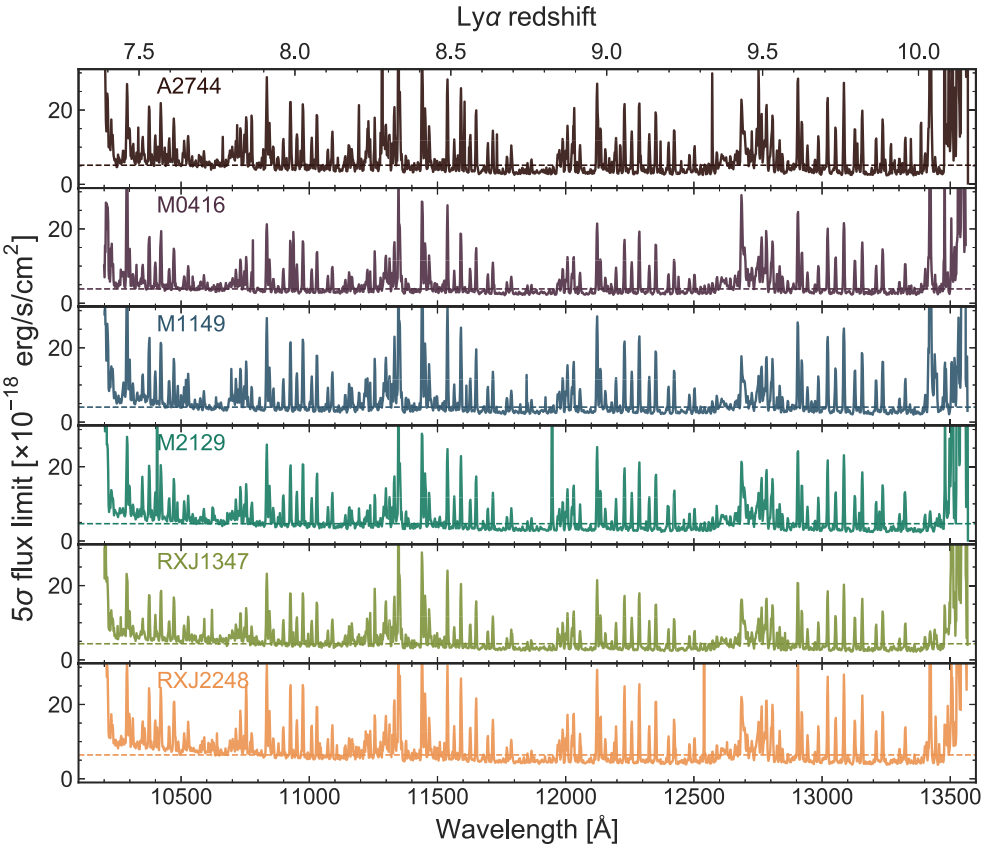

Figure 2. Average 5σ flux limits for each cluster as a function of wavelength, assuming emission lines are spatially unresolved. We use the 1D RMS noise spectrum for each cluster as described in Section 3.4 to obtain the flux limits. Each plot corresponds to the median across all IFUs containing high-redshift candidates, for the different cluster targets. The dashed horizontal lines mark the median flux limit for each cluster.

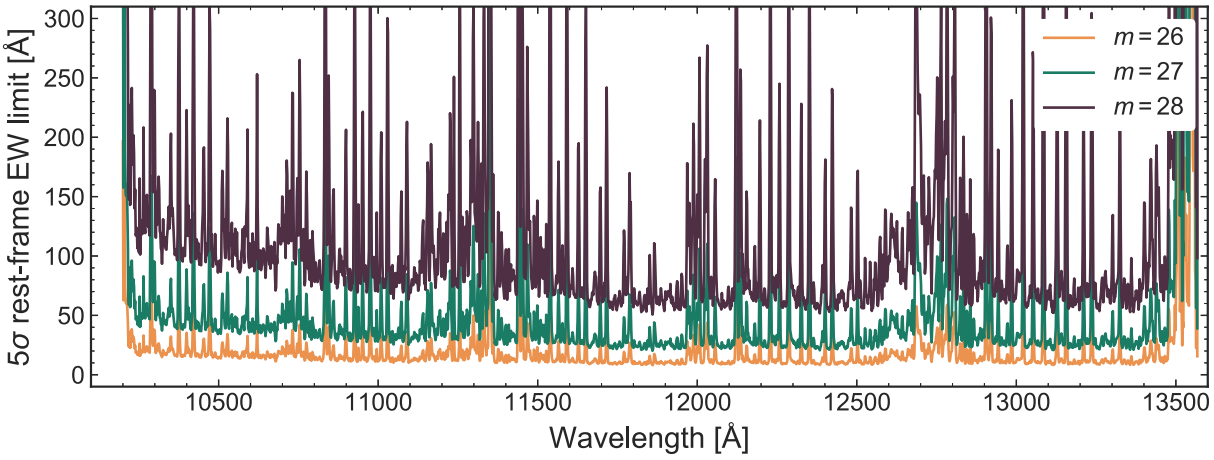

Figure 3. 5σ rest-frame Lyα EW limits in RXJ1347 as a function of wavelength, for three values of UV apparent magnitude m, assuming emission lines are spatially unresolved. We use the 5σ flux limit for RXJ1347 shown in Fig. 2 and divide by the continuum flux and $(1 + z_{\mathrm{Ly}\alpha})$ at each wavelength to obtain the EW limit.

same $z \sim 7$ cubes as used by M18a rather than generating new $z \sim 8$ cubes.

We refer the reader to M18a for more details of the forward-modelling approach. Here we describe the modifications we have made to our Bayesian inference to make use of the spectral coverage and sensitivity of our observations. We account for the incomplete redshift coverage and for the gravitational lensing magnification of the objects by the foreground clusters. We marginalize over a range of potential linewidths for the Lyα emission lines. We also marginalize over the photometric redshift distribution for each galaxy, which we obtain from comprehensive photometry (Section 4.2), to robustly account for uncertainties and degeneracies in redshift determination.

We want to obtain the posterior distribution for the neutral fraction: $p(\overline{x}_{\rm H\,I}\,|\,\{f\}, m, \mu)$ for each galaxy, where $\{f\}$ is an observed flux density spectrum as a function of wavelength, m is the observed apparent UV magnitude, and μ is the magnification. A full derivation of the posterior is shown in Appendix C, and we summarize it here.

Our inference framework calculates the likelihood of an emission line emitted at redshift z_d with observed rest-frame EW = W, being present in an observed flux density spectrum. To calculate this likelihood, we must assume a lineshape for the observed emission line. In previous inferences by Treu et al. (2013), Pentericci et al. (2014) and Tilvi et al. (2014) treated emission lines as unresolved: lines were modelled as Dirac Delta functions, with all the flux contained in a single spectral pixel. However, motivated by recent observations of $z > 6$ Lyα emission with linewidth FWHM$_{\rm spec}$ $\sim 200 - 450\,{\rm km\ s^{-1}}$, several times greater than the instrumental resolution (Vanzella et al. 2011; Ono et al. 2012; Finkelstein et al. 2013; Oesch et al. 2015; Zitrin et al. 2015; Song et al. 2016b), here we improve the method by including the effect of linewidth.

The inference is quite sensitive to linewidth as at fixed EW a broader line will have lower S/N in our observations. By assuming unresolved emission lines the lower limits on the reionization 'patchiness' parameter inferred by Treu et al. (2013), Pentericci et al. (2014) and Tilvi et al. (2014) will be slightly overestimated compared to a more realistic treatment of linewidth. We note that the $z \sim 7$ neutral fraction inference by (M18a) used EW limits calculated assuming a range of realistic linewidths so their result does not need revision. We discuss the impact of linewidth in more detail in Appendix C4 but note that our results are robust for FWHM in a realistic range $\sim$100−400 km s^{-1}.

To modify our inference to account for linewidth, we assume Gaussian emission lines for simplicity so can write the model emission line flux density as a function of EW:

$$f_{\rm mod}(\lambda, W, m, z_d, {\rm FWHM}) = \frac{W f_{\rm cont}(m, z_d)(1+z_d)}{\sqrt{2\pi}\sigma_\lambda} {\rm e}^{-\frac{1}{2}\left(\frac{\lambda-\lambda_d}{\sigma_\lambda}\right)^2} \quad (4)$$

where $z_d = \lambda_d/\lambda_\alpha - 1$, with $\lambda_\alpha = 1216\,$Å, is the redshift of an emission line, W is the rest-frame equivalent width of the emission line, $f_{\rm cont}$ is the flux density of the continuum calculated using equation (3) using the observed continuum apparent magnitude m, and $\sigma_\lambda = {\rm FWHM}/2.355$ is the spectral linewidth.

The likelihood of observing a 1D flux density spectrum $\{f\} = f(\lambda_i)$ for an individual galaxy (where i is the wavelength pixel index), given our model where the true EW is drawn from the conditional probability distribution $p(W\,|\,\overline{x}_{\rm H\,I}, m, \mu, z_d)$ is

$$p(\{f\}\,|\,\overline{x}_{\rm H\,I}, m, \mu, z_d, {\rm FWHM}) = \prod_i^N \int_0^\infty {\rm d}W \left[\frac{1}{\sqrt{2\pi}\sigma_i} {\rm e}^{-\frac{1}{2}\left(\frac{f_i - f_{\rm mod}(\lambda_i, W, m, z_d, {\rm FWHM})}{\sigma_i}\right)^2} \times p(W\,|\,\overline{x}_{\rm H\,I}, m, \mu, z_d)\right] \quad (5)$$

where σ_i is the uncertainty in flux density at wavelength pixel i and there are a total of N wavelength pixels in the spectrum. $p(W\,|\,\overline{x}_{\rm H\,I}, m, \mu, z_d)$ is the probability distribution for the observed rest-frame EW as a function of the neutral fraction, and galaxy properties – UV apparent magnitude, magnification, and redshift. This PDF is obtained by convolving the $p(W\,|\,\overline{x}_{\rm H\,I}, M_{\rm UV})$ model outputs from M18a with the probability distribution for each galaxy's absolute UV magnitude, including errors on m and μ (equation C5).

We note that the range of neutral fraction in the EoS simulations is $\overline{x}_{\rm H\,I} = 0.01 - 0.95$. In order to correctly calculate posteriors and confidence intervals, we set the likelihood at $\overline{x}_{\rm H\,I}$ such that we expect to observe no Lyα flux at all (in a fully neutral universe). That is, $p(\{f\}\,|\,\overline{x}_{\rm H\,I} = 1, m, \mu, z_d, {\rm FWHM}) = \prod_i^N \frac{1}{\sqrt{2\pi}\sigma_i} \exp(-f_i^2/2\sigma_i^2)$.

Given our relatively small sample size, we choose to restrict our inference to $z \sim 8$, thus for ease of computation we evaluate $p(W\,|\,\overline{x}_{\rm H\,I}, m, \mu, z_d)$ at $z_d = 8$; this has a negligible impact on the final likelihood. We keep z_d free in the rest of the inference. This product of likelihoods over the wavelength range of the spectrum accounts for the wavelength sensitivity of our observations, i.e. high-noise regions are weighted lower than low-noise regions.

We also note that EW is independent of magnification. Therefore, our inferences should be quite robust to magnification, which enters only through the dependency on $M_{\rm UV}$ of the assumed intrinsic EW distribution.

Using Bayes' theorem, the posterior distribution for $\overline{x}_{\rm H\,I}$, z_d and FWHM is

$$p(\overline{x}_{\rm H\,I}, z_d, {\rm FWHM}\,|\,\{f\}, m, \mu) \propto p(\{f\}\,|\,\overline{x}_{\rm H\,I}, m, \mu, z_d, {\rm FWHM}) \times p(\overline{x}_{\rm H\,I})\, p(z_d)\, p({\rm FWHM}) \quad (6)$$

We use a uniform prior on $\overline{x}_{\rm H\,I}$ between 0 and 1, $p(\overline{x}_{\rm H\,I})$, and use the photometric redshift distribution for the prior $p(z_d)$. As we are interested only in the posterior probability of $\overline{x}_{\rm H\,I}$ we can marginalize over FWHM and z_d for each galaxy. We use a lognormal prior on FWHM with mean depending on $M_{\rm UV}$ derived through empirical relations and 0.3 dex width; we discuss our choice of FWHM priors in more detail in Appendix C4 but find our results to be negligibly changed if we had used a uniform prior spanning the range of observed Lyα FWHM at $z > 7$ ($\sim$100−400 km s^{-1}). To account for the incomplete wavelength coverage, we use the fact that if the object has Lyα outside of the KMOS wavelength range (covering $[z_{\rm min} = 7.2, z_{\rm max} = 10.1]$) we would measure a non-detection in our data. Thus the posterior for $\overline{x}_{\rm H\,I}$ from one galaxy is

$$p(\overline{x}_{\rm H\,I}\,|\,\{f\}, m, \mu) \propto \int_{z_{\rm min}}^{z_{\rm max}} {\rm d}z_d\ p(\{f\}\,|\,\overline{x}_{\rm H\,I}, m, \mu, z_d) p(z_d) + \prod_i^N \frac{1}{\sqrt{2\pi}\sigma_i} {\rm e}^{-\frac{f_i^2}{2\sigma_i^2}} \left(1 - \int_{z_{\rm min}}^{z_{\rm max}} {\rm d}z_d\ p(z_d)\right) \quad (7)$$

We assume all galaxies observed are independent, so that the final posterior is the product of the normalized posteriors (equation 7) for each object.

Using the photometric redshift distributions as a prior on the redshift allows us to incorporate the probability of each galaxy truly being at high redshift (rather than a low-redshift contaminant) in a statistically rigorous way. In combining the posteriors in equation (7) for each galaxy, the photometric redshift distribution weights the individual posteriors based on the probability of the source being within our redshift range. LBGs usually have degeneracies in their photometry, which make it difficult to determine whether they are high-redshift star-forming galaxies or mature $z \sim 1$−2 galaxies. Thus with our method we are able to obtain reionization inferences from sources even when the photometric redshift distribution has multiple and/or broad peaks.

Whilst here we have carried out the inference at $z \sim 8$ only, with larger samples, it will be possible to measure $\overline{x}_{\rm H\,I}(z)$ directly, for

example by parametrizing its evolution with redshift and inferring the values of its redshift-dependent parameters, or in a Markov Chain Monte Carlo exploration of IGM simulations to also infer relevant astrophysical parameters (Greig & Mesinger 2015; Greig & Mesinger 2017).

4.2 Defining a selection function for a photometric sample

To make accurate inferences for reionization, it is important to have uniform and well-understood target selection functions for the sources we use. At the time of target selection for KLASS, not all deep HFF data were available, nor were sophisticated ICL removal techniques developed (e.g. Merlin et al. 2016; Livermore et al. 2017; Morishita et al. 2017). This led to heterogeneous target selections. However, for this analysis we now use the most up-to-date photometry available to create a sub-sample for analysis with a homogeneous selection function. We demonstrate in Appendix B that this sub-sample is not a biased selection from the final parent catalogues.

Deep, multiband *HST*, Spitzer-IRAC, and HAWK-I photometry is now available for all our targets through the CLASH, SURFSUP, and HFF programs (Postman et al. 2012; Bradač et al. 2014; Huang et al. 2016a; Lotz et al. 2017). For A2744, M0416, and M1149 we used the ASTRODEEP photometric catalogue that removed foreground ICL (Castellano et al. 2016; Merlin et al. 2016; Di Criscienzo et al. 2017). For M2129, RXJ1347 and RXJ2248 we created our own catalogues based on the ASTRODEEP methodology (M. Bradač et al., submitted). Of the 56 high-redshift candidate targets, we assigned to KMOS IFUs, 46 have matches in these final deep catalogues (including the three images of a $z = 6.11$ multiply-imaged system in RXJ2248).

To determine why 10 targets had no match in the final photometric catalogues, we examined our target selection catalogues. We used preliminary versions of the ASTRODEEP catalogues for A2744, M0416, and M1149 in our initial selection, so all the objects targeted in A2744 and M0416 have matches in the final catalogues. Three targets do not appear in the final M1149 catalogue; these objects were never in the preliminary ASTRODEEP catalogue but were selected from alternative preliminary HFF catalogues. Three targets from M2129, three targets from RXJ1347, and one target from RXJ2248 have no matches in the final catalogues, which was expected as they were selected from an ensemble of preliminary photometric catalogues with shallower photometry, and narrower wavelength coverage compared to our final catalogues. Three of the unmatched objects were Category 1 targets. These missing targets were likely faint in the initial photometry and so turn out to be spurious in deep photometry.

Photometric redshift distributions were obtained from the final catalogues with the EAzY code (Brammer, van Dokkum & Coppi 2008). We perform the EAzY fit to the entire photometric data set, and obtain photometric redshift posteriors without the magnitude prior (which weights bright objects to lower redshifts based on observations of field galaxies and may be inappropriate for our lensed sources). As described in Section 4.1, our inference framework uses the full photometric redshift distribution, thus we can robustly use all objects with non-zero probability of being in our redshift range of interest for our inferences.

Taking the 43 high-redshift KMOS targets matched in the catalogues (excluding the three images of the $z = 6.11$ galaxy described in Appendix A), we then use the photometric redshift distributions to select objects that could be in the KMOS YJ range ($7.2 < z_{\rm phot} < 10.1$). We calculate $P(7.2 < z_{\rm phot} < 10.1) = \int_{7.2}^{10.1} p(z_{\rm phot}) {\rm d}z_{\rm phot}$ using the normalized EAzY photometric redshift distribution for each object to find the total probability of the object being within that redshift range. We select 30 objects with $P(7.2 < z_{\rm phot} < 10.1) > 0.01$ (though the majority have a much higher probability of being in that redshift range). The photometric redshift distributions of these objects within the KMOS YJ range are plotted in Fig. 4.

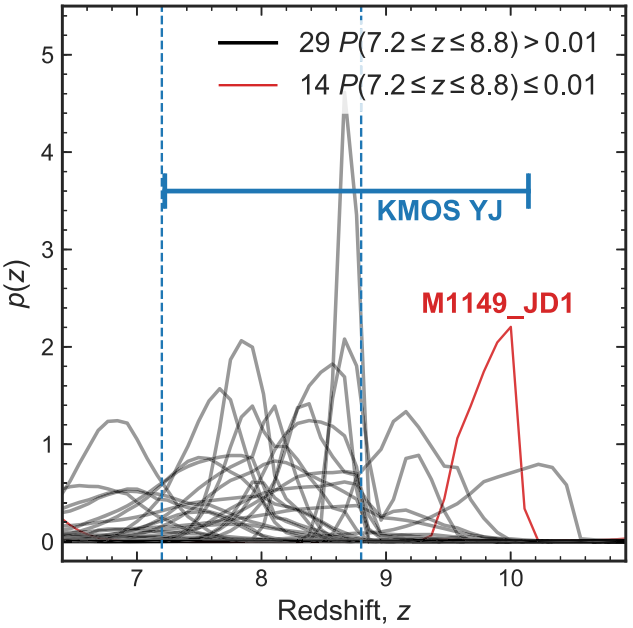

Figure 4. Photometric redshift distributions centred on the KMOS observable range. We show the KMOS YJ range for Lyα with the solid blue horizontal line. Black lines show the $p(z_{\rm phot})$ for the 29 sources that have >1 per cent probability of $7.2 \le z_{\rm phot} \le 8.8$ (marked by blue dashed vertical lines) that we use for the inference. 14 sources have $P(7.2 \le z_{\rm phot} \le 8.8) < 0.01$, including the galaxy M1149_JD1, recently spectroscopically confirmed at $z = 9.11$ with ALMA by Hashimoto et al. (2018). In our photometric catalogue, this galaxy is correctly found to be outside of our redshift range of interest (shown here as the red curve with $z_{\rm phot} > 9$), so we do not use it for our reionization analysis but discuss it in Section 5.2. Note that the remaining 13 objects have photometric redshift distributions outside of the range plotted here.

We examined the final deep photometry of the 13 objects that dropped out of the KMOS YJ range in this selection, which include six Category 1 targets. As expected, the selection of these objects shifts to lower redshifts now the full photometry is available. The majority of them have detections in the bluest bands that would negate a $z > 7$ Lyman Break, and several are clearly $z \sim 1$ passive galaxies when the IRAC bands are included.

Due to the relatively small sample size, we choose to perform our inference at $z \sim 8$, so we select only objects with some probability to have $7.2 < z_{\rm phot} < 8.8$. We calculate $P(7.2 < z_{\rm phot} < 8.8) = \int_{7.2}^{8.8} p(z_{\rm phot}) {\rm d}z_{\rm phot}$. We select 29 objects with >1 per cent probability of being within this redshift range (21 have >10 per cent probability, and 13 >60 per cent probability). One object has $z_{\rm phot} > 9$ and is excluded from our inference. This is M1149_JD1, recently spectroscopically confirmed at $z = 9.11$ by Hashimoto et al. (2018) in ALMA, who also show a tentative Lyα detection from X-shooter. As this galaxy's photometric redshift distribution clearly puts it at $z > 9$, we do not include it in our $z \sim 8$ reionization inferences. Its $p(z_{\rm phot})$ can be seen in Fig. 4 (red line) and we discuss our observations of it in Section 5.2.

Table 2. KLASS targets with $P(7.2 \leq z_{\rm phot} \leq 8.8)$ solutions.

Object ID[a]	RA [deg]	Dec. [deg]	$m_{\rm F160W}$	μ	$M_{\rm UV}^{\rm b}$	$P(z_{\rm phot})^{\rm c}$	$f_{\rm lim}^{\rm d} \times 10^{-18}$ [erg s^{-1} cm^{-2}]	$EW_{{\rm Ly}\alpha}^{\rm d,e}$ [Å]
A2744_2036	3.596 087	−30.385 836	26.95 ± 0.07	$2.4^{+7.4}_{-0.5}$	-19.27 ± 0.89	0.97	<12.1	<96
A2744_2346	3.606 460	−30.380 995	26.78 ± 0.06	$1.6^{+0.8}_{-0.5}$	-19.89 ± 0.41	1.00	<10.6	<70
A2744_2345	3.606 572	−30.380 932	26.49 ± 0.06	$1.6^{+0.8}_{-0.5}$	-20.19 ± 0.41	0.99	<10.6	<54
A2744_2261	3.603 996	−30.382 309	27.29 ± 0.10	$1.7^{+1.1}_{-0.5}$	-19.34 ± 0.47	0.79	<10.6	<113
A2744_2503	3.588 979	−30.378 668	27.27 ± 0.12	$2.2^{+0.9}_{-0.7}$	-19.04 ± 0.39	0.36	<11.4	<120
A2744_2257	3.598 123	−30.382 393	28.62 ± 0.18	$1.9^{+0.8}_{-0.4}$	-17.87 ± 0.36	0.54	<10.7	<392
A2744_20236	3.572 523	−30.413 267	28.61 ± 0.24	$1.8^{+1.0}_{-0.5}$	-17.94 ± 0.48	0.42	<9.5	<342
A2744_1040	3.592 505	−30.401 482	27.52 ± 0.15	$14.2^{+11.2}_{-6.3}$	-16.79 ± 0.65	0.04	<9.7	<129
A2744_2248[f]	3.603 863	−30.382 261	26.57 ± 0.07	$1.7^{+1.1}_{-0.5}$	-20.06 ± 0.47	0.96	<10.6	<58
M0416_99[g]	64.039 162	−24.093 182	26.28 ± 0.05	$1.5^{+0.5}_{-0.3}$	-20.49 ± 0.30	0.78	<3.4	<14
M0416_286	64.037 567	−24.088 116	28.20 ± 0.17	$1.9^{+0.3}_{-0.5}$	-18.29 ± 0.31	0.66	<3.6	<89
M0416_743	64.048 058	−24.081 427	26.56 ± 0.06	$1.7^{+0.3}_{-0.2}$	-20.07 ± 0.19	0.07	<3.1	<17
M0416_1956	64.060 333	−24.064 962	28.16 ± 0.16	$1.9^{+0.2}_{-0.6}$	-18.33 ± 0.28	0.91	<3.0	<72
M0416_1997	64.049 583	−24.064 596	27.56 ± 0.17	$6.3^{+39.3}_{-1.5}$	-17.64 ± 1.23	0.90	<2.9	<40
M0416_22746	64.046 509	−24.061 630	27.77 ± 0.23	$8.1^{+4.3}_{-3.0}$	-17.15 ± 0.53	0.62	<2.9	<49
M1149_23695	177.382 996	22.412 041	28.11 ± 0.14	$3.6^{+0.7}_{-2.1}$	-17.69 ± 0.58	0.77	<4.2	<97
M1149_3343	177.392 715	22.384 718	28.64 ± 0.28	$1.7^{+0.4}_{-0.5}$	-17.96 ± 0.42	0.04	<5.3	<201
M1149_1428	177.412 216	22.394 894	28.34 ± 0.17	$7.5^{+0.9}_{-2.8}$	-16.67 ± 0.36	0.25	<3.1	<87
M1149_945	177.412 079	22.389 055	27.92 ± 0.13	$9.2^{+14.4}_{-3.2}$	-16.87 ± 0.76	0.16	<3.3	<63
M2129_2633	322.345 232	−7.671 373	25.65 ± 0.12	$1.6^{+0.1}_{-0.1}$	-21.06 ± 0.13	0.20	<3.4	<8
M2129_2661	322.350 848	−7.675 239	26.38 ± 0.17	$1.7^{+0.0}_{-0.0}$	-20.25 ± 0.17	0.07	<3.4	<15
M2129_1556	322.344 535	−7.688 473	27.53 ± 0.26	$4.2^{+0.2}_{-0.2}$	-18.11 ± 0.27	0.01	<3.4	<45
RXJ1347_1831	206.896 270	−11.742 338	26.30 ± 0.26	$9.2^{+0.4}_{-0.4}$	-18.49 ± 0.26	0.06	<3.3	<14
RXJ1347_656	206.891 246	−11.752 607	26.43 ± 0.24	$20.4^{+1.6}_{-1.2}$	-17.49 ± 0.25	0.72	<3.7	<18
RXJ1347_101	206.880 973	−11.769 816	25.16 ± 0.15	$43.9^{+10.2}_{-5.4}$	-17.92 ± 0.26	0.20	<3.6	<5
RXJ1347_1368	206.893 076	−11.760 230	27.92 ± 0.43	$16.6^{+1.1}_{-1.1}$	-16.22 ± 0.43	0.34	<3.1	<60
RXJ1347_1280	206.896 921	−11.763 833	27.28 ± 0.28	$4.8^{+0.7}_{-0.5}$	-18.22 ± 0.31	0.03	<2.8	<29
RXJ2248_1006	342.208 379	−44.537 520	25.83 ± 0.17	$1.6^{+0.4}_{-0.4}$	-20.88 ± 0.32	0.92	<4.6	<13
RXJ2248_2086	342.179 829	−44.525 664	26.88 ± 0.13	$41.0^{+72.3}_{-25.5}$	-16.28 ± 1.09	0.48	<3.9	<28

Note: [a]IDs for A2744, M0416, and M1149 match the ASTRODEEP catalogue IDs (Merlin et al. 2016; Di Criscienzo et al. 2017). [b]These listed intrinsic magnitudes are calculated using $z = 8$ and the listed magnifications and errors. [c]This is the photometric redshift from EAzY integrated between $z = 7.2$ and $z = 8.8$, i.e. the total probability of the object to have. [d]Flux and EW limits are 5σ. [e]All EW are rest-frame. We stress that the EW limits only hold if the Lyα is actually in the KMOS range, which has probability given by $P(7.2 \leq z_{\rm phot} \leq 8.8)$. [f]This object was spectroscopically confirmed by Laporte et al. (2017a) at $z = 8.38$. [g]This object was spectroscopically confirmed by Tamura et al. (2018) at $z = 8.31$.

Our inference uses the full $p(z_{\rm phot})$ distribution, to robustly account for any probability of an object being a lower redshift contaminant. The median and standard deviation of best-fitting photometric redshifts over this range for the sub-sample of 29 objects are $z_{\rm phot} = 7.9 \pm 0.6$. These objects and their observed properties, including $P(7.2 < z_{\rm phot} < 8.8)$, are listed in Table 2. We demonstrate that this sub-sample is not a biased sample of the final photometric catalogues in Appendix B.

We also cross-checked our Table 2 with publicly available spectroscopic catalogues from ground-based follow-up at optical wavelengths for clusters A2744 (Mahler et al. 2018), M0416 (Balestra et al. 2016; Caminha et al. 2017), M1149 (Grillo et al. 2016), M2129 (Monna et al. 2017), and RXJ2248 (Karman et al. 2015, 2016). We found no matches in those catalogues for any of the objects in Table 2. Non-detections of these objects in optical spectroscopy lends credence to their selection as $z > 7$ candidates.

Two objects have been spectroscopically confirmed at $7.2 < z < 8.8$ by groups. A2744_2248 (a.k.a. A2744_YD4) was confirmed at $z = 8.38$ via [O III]88 μm emission in ALMA, and a tentative Lyα emission line was also reported with line flux $(1.82 \pm 0.64) \times 10^{-18}$ erg s^{-1} cm^{-2} and EW $= 10.7 \pm 2.7$ (Laporte et al. 2017a), which is well below our limit for that object. As discussed in Section 4.3, we find that treating the object as a detection in our inference has a negligible impact on our inferred limits on the neutral fraction. M0416_99 (a.k.a M0416_Y1) was also confirmed via [O III]88μm emission in ALMA observations at $z = 8.31$ by Tamura et al. (2018). They also observed the object with X-shooter and found no rest-frame UV emission lines, with a 5σ Lyα flux limit of $<8.0 \times 10^{-18}$ erg s^{-1} cm^{-2} (in the line if offset by up to 250 km s^{-1}). Our KMOS flux median limits are of a comparable depth ($<3.4 \times 10^{-18}$ erg s^{-1} cm^{-2}).

We obtain magnification estimates for each object using the publicly available HFF lens models.[4] We take the best-fitting magnifications from the most recent versions of all available lens models for each object and drop the highest and lowest magnifications to produce an approximate 1σ range of estimated magnifications, $\{\mu\}$. We list the median magnification from this sub-sample, and the upper and lower bounds in Table 2. For the inference, we assume magnifications are log-normally distributed with mean given by

[4]https://archive.stsci.edu/prepds/frontier/lensmodels/

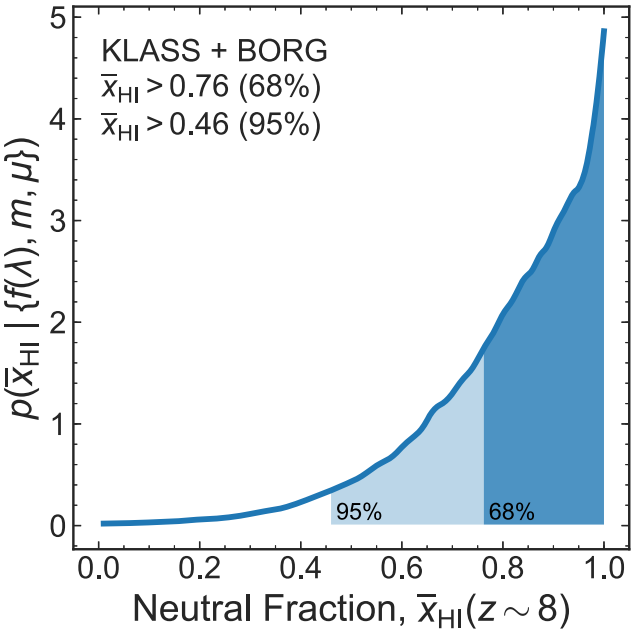

Figure 5. Posterior probability distribution for the IGM neutral fraction $\bar{x}_{\rm H\,I}$ at $z \sim 8$ obtained using equation (7) and the EW spectra from the KLASS sample described in Section 4.2 and the BoRG sample described by Treu et al. (2013). The blue line and shaded regions show the posterior from the combined data sets, and its 68 per cent and 95 per cent confidence regions (the darkest region is the 68 per cent confidence range).

the median $\log_{10}\{\mu\}$ and standard deviation given by half the range of $\log_{10}\{\mu\}$, which is a reasonable fit to the distribution of magnifications from the models. For M2129 and RXJ1347, the only non-HFF clusters, we use the magnification distribution from the Bradač group lens models (Huang et al. 2016b; Hoag et al. 2019) and obtain mean and standard deviation log magnifications. As discussed in Section 4.1 by using the EW in our inference, which is independent of magnification (as opposed to flux), our results are quite robust to magnification uncertainties.

We calculate flux and Lyα EW limits for individual objects as in Section 3.4, using equations (2) and (3). The median intrinsic UV absolute magnitude (i.e. corrected for magnification) of the sample is $M_{\rm UV} = -18.2$. The median observed flux 5σ upper limit in this sub-sample is $<3.6 \times 10^{-18}$ erg s^{-1} cm^{-2}, and the median rest-frame Lyα EW 5σ upper limit is <58 Å.

4.3 Inference on the IGM neutral fraction

We use 1D spectra and uncertainties as a function of wavelength for the 29 objects described above to infer the IGM neutral fraction at $z \sim 8$ using equation (7) to calculate the posterior distribution of $\bar{x}_{\rm H\,I}$.

We obtain the flux density spectra using the cubes for each object, extracting flux and noise in a circular aperture with $r = 2\sigma_{\rm psf}$, and apply a rescaling to both to account for the incomplete recovery of flux in the aperture, and a constant rescaling to the noise spectrum to ensure that the S/N distribution of pixels in each spectrum is a Normal distribution.

In Fig. 5 we plot the posterior distribution for $\bar{x}_{\rm H\,I}$ obtained using our observations of the 29 $z \sim 8$ KLASS targets, as well as Keck/MOSFIRE observations of 8 $z \sim 8$ LBGs from the Brightest of Reionizing Galaxies survey (BoRG; Trenti et al. 2011; Bradley et al. 2012; Schmidt et al. 2014b) described by Treu et al. (2013). Using the BoRG sample allows us to cover a broader range in intrinsic magnitudes spanning opposite ends of the galaxy UV luminosity function: the IGM attenuation of Lyα from UV bright and UV faint galaxies is expected to be different due to differing Lyα escape paths through their interstellar and circumgalactic media (e.g. Stark et al. 2010, 2017; Mason et al. 2018b).

These two sets of independent observations indicate a predominantly neutral IGM at $z \sim 8$. The BoRG data alone produce a lower limit of $\bar{x}_{\rm H\,I} > 0.34$ (68 per cent) and for the KLASS data alone $\bar{x}_{\rm H\,I} > 0.76$ (68 per cent). Lower limits from the combined data set are $\bar{x}_{\rm H\,I} > 0.76$ (68 per cent) and $\bar{x}_{\rm H\,I} > 0.46$ (95 per cent).

By exploiting gravitational lensing, the KLASS sample sets much lower limits on the Lyα EW for intrinsically UV faint galaxies (which produce the strongest constraints on reionization's mid-stages, M18a) than is possible in blank fields. Our new KLASS sample also demonstrates how increasing the number of sources for the inference produces much tighter constraints on the IGM neutral fraction compared to the eight BoRG sources.

To test whether the inclusion of objects with candidate Lyα emission in GLASS data biased our sample, we tested the inference with and without including the Category 1 targets (which were specifically targeted in KLASS because they had candidate Lyα emission in the *HST* data). We found no significant difference in the posteriors. We also tested the inference with and without including the $z = 8.38$ marginal detection of Lyα in object A2744_2248 by Laporte et al. (2017a) with spectroscopic confirmation from O III emission in ALMA observations. We use the EW reported by Laporte et al. (2017a) $W = 10.7 \pm 2.7$ Å, which is well below our 5σ limit for that object (<53 Å). Despite the potential detection, the posterior distribution for this single object strongly favours a mostly neutral IGM due to its very low EW and low significance. We did our inference using both our KMOS spectra and the Laporte et al. (2017a) measurement for this object and found it to have a negligible impact on our final posterior (changing the inferred limit by only $\Delta\bar{x}_{\rm H\,I} \sim 0.01$), demonstrating that deep limits on non-detections have a lot of power in our inferences. Our quoted posterior limits include the object as a non-detection.

5 DISCUSSION

We discuss our new lower limit on the neutral fraction and the implications for the timeline of reionization in Section 5.1, and show it favours reionization driven by UV faint galaxies with a low ionizing photon escape fraction. In Section 5.2 we discuss the recent tentative detection of Lyα at $z = 9.11$ by Hashimoto et al. (2018) and show it is not inconsistent with our results. In Section 5.3 we discuss our EW limits on NV and C IV emission. Finally, in Section 5.4 we present a comparison of the KMOS ETC and our achieved S/N for background-limited observations.

5.1 The timeline of reionization

We plot our new limit on the reionization timeline in Fig. 6. We also plot other statistically robust constraints from Ouchi et al. (2010), McGreer et al. (2014), Sobacchi & Mesinger (2015), Mesinger et al. (2015), Davies et al. (2018), Greig et al. (2019), Mason et al. (2018a) and Planck Collaboration XLVII (2016). While no increase in $\bar{x}_{\rm H\,I}$ compared to the Mason et al. (2018a) constraint is statistically possible within the 95 per cent confidence interval, our new limit, combined with the other recent $\bar{x}_{\rm H\,I}$ statistical measurements at $z \sim 7$, and other estimates (e.g. Caruana et al. 2014; Zheng et al. 2017), provides increasing evidence for the bulk of hydrogen reionization

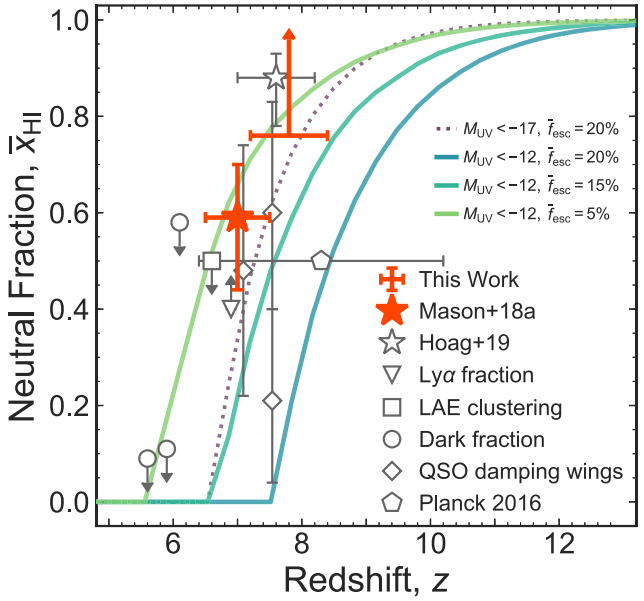

Figure 6. The redshift evolution of the volume average neutral hydrogen fraction of the IGM. Our new lower limit is shown in orange, with the horizontal errorbar at the 68 per cent confidence level. We also plot measurements derived from observations of the evolving Lyα EW distribution at $z \sim 7$ (orange-filled star M18a) previous estimates from the fraction of LBGs emitting Lyα (open black star; Mesinger et al. 2015); the clustering of Lyα emitting galaxies (square; Ouchi et al. 2010; Sobacchi & Mesinger 2015); Lyα and Lyβ forest dark fraction (circle – 68 per cent limits; McGreer et al. 2014); and QSO damping wings (diamond; Davies et al. 2018; Greig, Mesinger & Bañados 2019). We offset the constraints at $z \sim 7$ by $\Delta z = 0.1$ for clarity. We also plot the Planck Collaboration XLVII (2016) redshift range of instantaneous reionization (black pentagon). We show median model reionization histories derived from the Mason, Trenti & Treu (2015) UV luminosity function models as coloured lines. We plot models obtained from integrating the luminosity function down to two magnitude limits – $M_{UV} = -17$ (purple dashed line) and $M_{UV} = -12$ (darkest blue solid line) and drawing from uniform distributions for the ionizing photon escape fraction 10−30 per cent ($\langle f_{esc} \rangle = 20$ per cent) and clumping factor $C = 1-6$, and lognormal distribution for the ionizing efficiency ξ_{ion} with mean 25.2 and standard deviation 0.15 dex. Comparing reionization histories with ionizing escape fraction drawn from a uniform distribution $1-10$ per cent (light green, $\langle f_{esc} \rangle \approx 5$ per cent) and $10-20$ per cent (medium teal, $\langle f_{esc} \rangle = 15$ per cent), integrating LFs down to $M_{UV} = -12$ in both cases and using the same distribution for the clumping factor and ξ_{ion} as above.

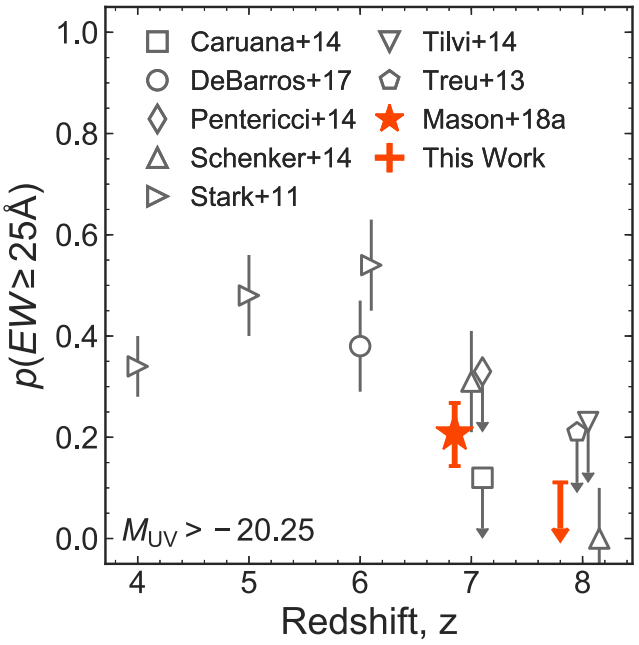

Figure 7. The redshift evolution of the 'Lyα fraction' for UV faint galaxies, the fraction of LBGs observed with Lyα EW ≥ 25 Å. We plot literature measurements from Stark et al. (2011), Pentericci et al. (2014), Treu et al. (2013), Tilvi et al. (2014), Schenker et al. (2014), Caruana et al. (2014), and De Barros et al. (2017). We add small offsets in redshift for measurements at the same redshifts to ease the display of the data. We also plot the predicted Lyα fraction from M18a calculating $p(W > 25Å \,|\, \bar{x}_{H\,I}, M_{UV})$ using $M_{UV} = -20$ galaxies and the neutral fraction constraint $\bar{x}_{H\,I} = 0.59^{+0.11}_{-0.15}$ (16 − 84 per cent confidence intervals) as the orange star. We plot the upper limits recovered in this paper as orange lines, with the solid line showing our 68 per cent confidence limit, and the dotted line extending to the 95 per cent confidence limit. We calculate $p(W > 25Å \,|\, \bar{x}_{H\,I} > 0.76, M_{UV})$ again using $M_{UV} = -20$. Our constraint is consistent with literature values at the same redshift.

occurring $z \sim 6-8$ (Bañados et al. 2017; Greig et al. 2017; Davies et al. 2018; Mason et al. 2018a), late in the Planck Collaboration XLVII (2016) confidence range.

Accurate measurements of the reionization timeline can help constrain properties of early galaxies. In Fig. 6 we show model reionization histories obtained from integrating the Mason et al. (2015) UV luminosity functions, varying the typical reionization parameters: the minimum UV luminosity of galaxies, and the average ionizing photon escape fraction. We see that late reionization is most consistent with either a high minimum UV luminosity of galaxies ($M_{UV} < -17$) and moderate escape fraction ($\langle f_{esc} \rangle = 20$ per cent), or with including ultra-faint galaxies $M_{UV} < -12$) with low escape fractions ($\langle f_{esc} \rangle \lesssim 15$ per cent).

There are many degeneracies between these reionization parameters, and certainly the escape fraction is unlikely to be constant for all galaxies at all times (Trebitsch et al. 2017), but non-detections of high-redshift GRB host galaxies, and observations of lensed high-redshift galaxies, and local dwarfs, indicate galaxies fainter than $M_{UV} = -17$ likely exist at $z \sim 8$ (e.g. Kistler et al. 2009; Tanvir et al. 2012; Trenti et al. 2012; Alavi et al. 2014; Bouwens et al. 2017; Livermore et al. 2017; Weisz & Boylan-Kolchin 2017; Ishigaki et al. 2018). If ultra-faint galaxies do contribute significantly to reionization; our result suggests reionization can be completed with low escape fractions, consistent with low-redshift estimates of the average escape fraction (Rutkowski et al. 2017; Marchi et al. 2018; Naidu et al. 2018; Steidel et al. 2018)

For comparison with previous high-redshift Lyα spectroscopic surveys, we plot the so-called 'Lyα fraction', the fraction of LBGs emitting Lyα with EW ≥ 25 Å in Fig. 7. We compare our new upper limits on the Lyα fraction with literature measurements from Stark et al. (2011), Pentericci et al. (2011), Treu et al. (2013), Tilvi et al. (2014), Schenker et al. (2014), Caruana et al. (2014), and De Barros et al. (2017). We also plot the predicted Lyα fraction from M18a. Using the M18a model EW distributions $p(W \,|\, \bar{x}_{H\,I}, M_{UV})$ we can calculate the Lyα fraction as the probability of EW ≥ 25 Å given our constraint on the neutral fraction.

As noted by M18a and Mason et al. (2018b) the Lyα EW distribution is likely a function of at least UV magnitude as well as the neutral fraction (see Oyarzún et al. 2017, for a thorough analysis of Lyα EW dependencies on galaxy properties), so it can be difficult to compare Lyα fraction from samples with different M_{UV}. Hence,

when converting from the neutral fraction measurement in this work and M18a we use the model Lyα EW distribution for $M_{\rm UV} = -20$ galaxies to compare more easily with the literature values for which that is the typical median UV magnitude. For $M_{\rm UV} = -20$ our Lyα fraction limits are $f_{\rm Ly\alpha} < 0.11$ (68 per cent), <0.27 (95 per cent). Using our sample median magnitude, $M_{\rm UV} = -18.2$, the limits are not significantly different: $f_{\rm Ly\alpha} < 0.08$ (68 per cent), <0.24 (95 per cent). Our measurements are consistent with the literature values.

We note that our inference assumes no evolution in the emitted Lyα EW distribution at fixed UV magnitude from $z \sim 6-8$, i.e. the only evolution in the *observed* EW distribution is due to reionization. Whilst there may be evolution in the amount of Lyα escaping the ISM of galaxies with increasing redshift, it is probably increasing as dust masses and HI covering fractions may decrease at higher redshifts and facilitate Lyα escape at fixed galaxy mass (Hayes et al. 2011; Oyarzún et al. 2016). In this case we expect our model to underestimate the observed EW distribution, which would suggest an even higher neutral hydrogen fraction given our non-detections. Our model also assumes no significant evolution in the dust spatial distribution and/or CGM opacity between $z \sim 6-8$, which could both reduce the Lyα EW before the photons reach the IGM. If these effects do significantly decrease Lyα EW between $z \sim 6-8$, this could lower our constraint on the neutral fraction. In modelling the emitted Lyα EW distribution, we assume a Gaussian plus Dirac Delta function parametrization, which has been shown to describe the Lyα EW distribution well (Oyarzún et al. 2017). However, choosing another functional form for the distribution will not significantly change the results (Treu et al. 2012; Schenker et al. 2014).

More accurate models of Lyα emerging from the $z > 6$ ISM are required to improve our inferences. Whilst it is increasingly difficult to directly observe all of the emitted Lyα from $z > 6$ galaxies, because of the intervening neutral gas, other emission lines could be used as a diagnostic of emerging Lyα. For example, Henry et al. (2018) showed that MgII emission line profiles and escape fractions closely trace those of Lyα in Green Peas, low-redshift analogues of high-redshift galaxies (Jaskot & Oey 2014; Yang et al. 2016). As the IGM optical depth to MgII is much lower than for Lyα, observations of MgII at $z > 6$ (which will be possible with JWST) could be used to infer the nature of Lyα emission at these redshifts.

Additionally, better knowledge of Lyα line profiles at $z \gtrsim 5$ are necessary to provide more informative priors on the observed FWHM for our inferences. In particular, high-resolution spectroscopy ($R > 4000$) is needed to resolve the narrow lines expected for UV faint galaxies (Verhamme et al. 2015), and could provide additional constraints via the evolving Lyα profile (Pentericci et al. 2018) and the prevalence of double-peaked Lyα in the late stages of reionization (Matthee et al. 2018).

We also assume the fraction of low-redshift contaminants in our photometric sample is the same as our reference $z \sim 6$ sample from De Barros et al. (2017). Whilst the selection techniques for the two samples are different (ours is based on photometric redshifts; De Barros et al. 2017 use a colour selection), our targets have extensive multiwavelength photometry that help to rule out low-redshift contaminants (e.g. Vulcani et al. 2017; Livermore et al. 2018). Additionally, we use the full photometric redshift distribution from EAzY in our inference which will weight the most convincing high-redshift candidates most strongly in our inference, and robustly account for contamination. With the final GLASS Lyα candidate sample it will be possible to use the same selections for both the $z \sim 6$ reference EW distribution and the $z > 6$ samples for reionization inferences (K. B. Schmidt et al., in preparation).

As our inference weights sources by their photometric redshift distribution, the tightest constraints on $\overline{x}_{\rm H\ I}$ will be obtained from samples with robust redshift estimates or, ideally, spectroscopic redshifts obtained from other emission lines, and deep Lyα EW limits. We note that the objects that constribute the most to our posterior are the objects with the highest probability (> 60 per cent) of having photometric redshift at $z \sim 8$ due to their SEDs, and we expect these to have consistent high-redshift solutions even if the photometric redshift fitting priors are changed. The prospects for large spectroscopic samples at these redshifts are increasing: ALMA is enabling spectroscopic confirmation of $z \gtrsim 7$ galaxies in the sub-mm (e.g. Bradač et al. 2017; Laporte et al. 2017a; Hashimoto et al. 2018; Smit et al. 2018; Tamura et al. 2018), and other UV emission lines have also been confirmed (Stark et al. 2015, 2017; Schmidt et al. 2016; Mainali et al. 2017; Mainali et al. 2018). Future observations with JWST slitless and slit spectroscopy will be able to build large and deep spectroscopic samples of $z \gtrsim 7$ galaxies, ideal for this type of analysis.

Understanding the differing evolution of Lyα emission as a function of galaxy properties and environment will be key to understanding how reionization progresses. Here we have shown that a sample of intrinsically UV faint systems at $z \sim 8$ (more likely to live in low-density environments) show no significant Lyα emission, and favour a mostly neutral IGM. However, Lyα has been observed in a handful of UV bright galaxies at $z \gtrsim 7.5$ (Oesch et al. 2015; Zitrin et al. 2015; Roberts-Borsani et al. 2016; Stark et al. 2017). Mason et al. (2018b) showed that the observed Lyα fraction for UV bright galaxies at $z \sim 8$ could not be reproduced with standard reionization models (using the EoS simulations; Mesinger et al. 2016), even when placing them in overdense regions (which reionize early) and giving them high Lyα velocity offsets to facilitate Lyα IGM transmission. Mason et al. (2018b) proposed those objects have detectable Lyα because they have unusually high emitted Lyα EW (they were certainly selected to have high nebular line EW; Roberts-Borsani et al. 2016).

Fluctuations in the UV background during reionization, for example, due to the inhomogeneous distribution of ionizing sources, could also contribute to the differing evolution of Lyα emission from UV bright and UV faint galaxies by boosting the IGM opacity (transparency) in underdense (overdense) regions (Davies & Furlanetto 2016; Becker et al. 2018). One important missing piece in our inference is the halo environment of the LBGs. This work assumes a simple mapping between UV luminosity and halo mass. This works well in an average sense (Mason et al. 2015), but deep imaging with JWST could measure the clustering strength and scatter of galaxies in the reionization epoch (Ren, Trenti & Mutch 2018), and be used to inform more realistic IGM simulations.

5.2 M1149_JD1 – Lyα emission at $z = 9.11$?

One target in our observations (known as M1149_JD1; Zheng et al. 2012; Hoag et al. 2018) was recently spectroscopically confirmed at $z = 9.11$ via [O III] 88 μm emission with ALMA observations (Hashimoto et al. 2018). Our EAzY photometric redshift distribution for this galaxy put it outside of our inference redshift range (all of the $p(z)$ is at $z > 9$; see Fig. 4), so it was not used in our reionization inference. However, Hashimoto et al. (2018) also report a tentative 4σ detection of Lyα emission from this galaxy in X-shooter observations at 12271.5 Å with total line flux

$(4.3 \pm 1.1) \times 10^{-18}$ erg s^{-1} cm^{-2}. Hoag et al. (2018) also targeted this galaxy with low-resolution *HST* grism spectroscopy, including GLASS data, which covered the Lya wavelength at $z = 9.11$. While they did not claim a detection, their spectra show an $\sim$2.5σ feature at approximately the same wavelength and flux as Hashimoto et al. (2018). We examined our KMOS cube and find no evidence of a feature at this wavelength. Our median 1σ flux limit for $z > 9$ Lyα in the cube is $>1.1 \times 10^{-18}$ erg s^{-1} cm^{-2} and $>0.8 \times 10^{-18}$ erg s^{-1} cm^{-2} at 12271.5 Å.

As noted by Hashimoto et al. (2018), if their candidate line is Lyα, it is blueshifted by $\sim$450 km s^{-1} with respect to the [O III] emission. Hashimoto et al. (2018) suggest that Lyα photons scattered off inflowing gas, causing it to emerge blueshifted from the galaxy's systemic velocity. Whilst blueshifts due to inflows are expected and observed for Lyα (e.g. Dijkstra, Haiman & Spaans 2006; Verhamme, Schaerer & Maselli 2006; Trainor et al. 2015), at $z > 6$ the IGM is opaque to emission $<$ 1216 Å, thus no Lyα emitted bluer than its source galaxy's systemic redshift should be transmitted through the IGM (Dayal, Maselli & Ferrara 2011; Dijkstra et al. 2011).

Observing blueshifted Lyα requires the galaxy to sit in a large ionized bubble ($\gtrsim$500 km s^{-1} or $\gtrsim$400 kpc in radius; Haiman 2002). Alternatively, the Lyα emission could arise in a different component or merging companion of the [O III] emitting galaxy, similar to a $z = 7.1$ galaxy observed by Carniani et al. (2017). The tentative emission we observe in our KMOS cube does appear spatially offset from the predicted position of the UV continuum and [O III] by $\sim$0.″4, which could provide evidence for the multicomponent/merger scenario. This may also account for our slightly lower flux measurement as the emission extends to the edge of the IFU. However, the weakness of the detection and some general astrometric uncertainty in KMOS make a thorough analysis difficult. Deeper near-IR IFU observations of this galaxy would be extremely interesting to confirm and determine the nature of the Lyα emission, and will be possible in the future with JWST NIRSpec.

We calculate the probability of observing Lyα emission from such an object in a mostly neutral IGM using the framework of M18a, which modelled $p(W \,|\, \overline{x}_{\rm H\ I}, M_{\rm UV})$. Using $m_{\rm F160W} = 25.7$ (Zheng et al. 2012), we obtain $M_{\rm UV} = 19.2 - 2.5\log_{10}(10/\mu)$, $EW = 4 \pm 2$ Å for our measured flux and $EW = 11 \pm 3$ Å from the measurement by Hashimoto et al. (2018). Using these measurements, we calculate $p(W = 4 \pm 2\ \text{Å} \,|\, \overline{x}_{\rm H\ I} > 0.76, M_{\rm UV} = -19.2) < 0.05$, while $p(W = 11 \pm 3\ \text{Å} \,|\, \overline{x}_{\rm H\ I} > 0.76, M_{\rm UV} = -19.2) < 0.03$. In fact, the total probability of observing Lyα from this galaxy with EW $>4 \pm 2$ Å if $\overline{x}_{\rm H\ I} > 0.76$ is $\lesssim$ 0.5: low Lyα EW are expected and consistent with a mostly neutral IGM.

We note that our calculations assume the Lyα is emitted close to systemic velocity (i.e. assuming that the Lyα comes from another component). Obviously if the galaxy does sit in an ionized bubble, the probability of seeing emission would be higher. But we note that assuming emission is emitted at systemic velocity the probability of detecting the emission is not negligible, and thus this detection is still consistent with a mostly neutral IGM at $z > 8$.

5.3 Other UV emission lines at $z \sim 8$

With Lyα increasingly suppressed at $z > 6$, rest-frame UV emission lines can be used to spectroscopically confirm high-redshift LBGs. These lines can also be used as diagnostics for the stellar populations and physical conditions present in these high-redshift galaxies. Our KMOS observations cover the wavelength range where NVλ1238, 1242 and C IVλ1558, 1551 can be observed, and we briefly discuss our upper limits on the EW of these lines.

NVλ1238, 1242 can arise due to stellar winds, particularly from very young stars (Shapley et al. 2003; Jones, Stark & Ellis 2012), or from HII regions if powered by an active galactic nucleus (AGN) or radiative shocks. Of the three $z > 7$ galaxies detected to date with tentative NV emission ($S/N \sim 4$) all have been UV bright galaxies, where AGN activity could plausibly be powering NV emission (Tilvi et al. 2016; Laporte et al. 2017b; Mainali et al. 2018). In our KLASS $7.2 < z_{\rm phot} < 8.8$ sub-sample (Section 4.2), the median NV EW upper limit is $<$60 Å. As our sample comprises intrinsically faint galaxies, which are less likely to have strong AGN activity, it is not surprising that we do not detect strong NV emission.

Nebular C IVλ1558, 1551 emission has been observed in two galaxies at $z > 6$ (Stark et al. 2015; Mainali et al. 2017; Schmidt et al. 2017). We observed C IV in the $z = 6.11$ galaxy with KMOS and describe our observations in more detail in Appendix A. The C IV emission can be powered by either AGN activity or extremely metal poor stars. Limits on other UV lines in these objects find low metallicity stars are a more likely source of the hard photons needed to produce C IV emission, rather than AGN. The two galaxies are also both UV faint galaxies ($M_{\rm UV} \lesssim -20.2$) and Mainali et al. (2018) have suggested that there is anticorrelation between UV luminosity and C IV EW, which could arise if the lowest luminosity (mass) systems are more metal-poor.

Our KLASS observation provides a large additional sample of UV faint galaxies that can place new limits on C IV emission. In our KLASS $7.2 < z_{\rm phot} < 8.8$ sub-sample (Section 4.2), the median C IV EW upper limit is $<$74 Å. In the three most UV faint systems with $P(7.2 < z_{\rm phot} < 8.8) > 0.6$, M0416_22746, RXJ1347_656, M0416_1997 (all with $M_{\rm UV} \sim -17.5$), the C IV upper limits are $<$62 Å, $<$22 Å, and $<$51 Å, respectively. These upper limits are comparable to, and in one case below, the C IV detection presented by Stark et al. (2015) in a $M_{\rm UV} \sim -19$ galaxy (with $EW_{\rm CIV} \approx 38$ Å), and so suggest that the proposed anticorrelation between UV luminosity and C IV EW may not be so simple.

5.4 Background limited observations with KMOS

Optical and near-IR IFU observations have provided revolutionary 3D information about the structure and kinematics of galaxies out to $z \sim 2$ (Epinat et al. 2009; Förster Schreiber et al. 2009; Wisnioski et al. 2015; Stott et al. 2016; Genzel et al. 2017) and revealed diffuse Lyα haloes around $z \lesssim 6$ galaxies (Bacon et al. 2014; Karman et al. 2016; Wisotzki et al. 2016; Leclercq et al. 2017).

In KLASS we have provided the first deep NIR IFU observations of $z \gtrsim 7$ galaxy candidates. Whilst we did not make any 5σ detections of Lyα, it is important to understand how this depended not only on the selection of our targets and the opacity of the IGM to Lyα at $z \sim 8$, but on the sensitivity of KMOS. In our long integrations we have pushed KMOS to the limits of its sensitivity to search for faint emission lines in near-IR IFU cubes, in wavelength regions dominated by OH sky emission lines. Using our deep observations, we provide an assessment of the performance of KMOS for background-limited observations.

As described in Section 2.4 we performed additional sky subtraction routines after running the ESO pipeline to reduce residuals around bright OH lines. We also found that the pipeline underestimated the noise in cubes by a factor of $\sim$1.2 and performed additional rescaling of the noise as a function of wavelength using the RMS noise obtained from the flux cubes.

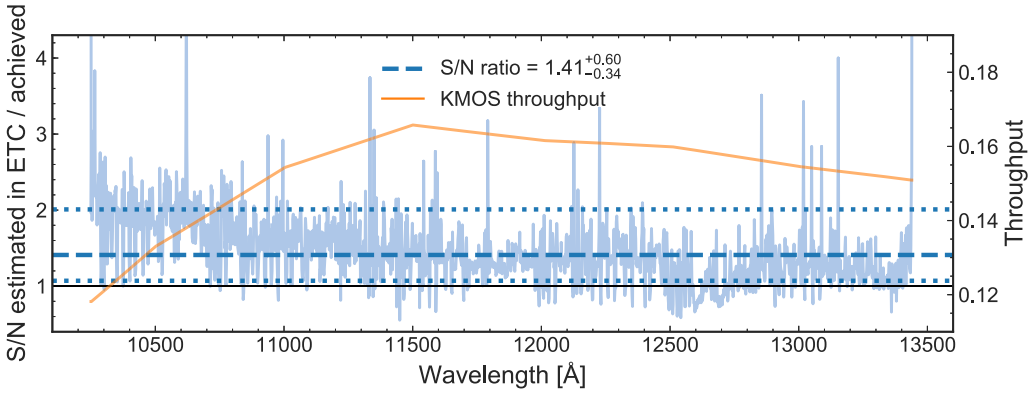

Figure 8. Comparison of our deepest exposure, 11 h in RXJ1347, with 450 s DITs, with the ESO KMOS ETC using the same exposure times. We compare the 5σ flux limits from our data and the ETC as a function of wavelength, assuming emission lines are spatially and spectrally unresolved. We divide the ETC estimated noise by $\sqrt{2}$ to account for the noise introduced by 'A–B' sky subtraction. The pale blue solid line shows the ratio of the ETC estimated S/N to our achieved S/N. The blue dashed (dotted) horizontal lines show the median (16 − 84 per cent range) of the S/N ratio over the whole YJ range. The orange line shows the KMOS throughput for comparison.

One key question is how well the instrument performs relative to the predictions based on its instrumental capabilities. We can compare S/N estimated by the KMOS ETC[5] to our achieved S/N to assess its performance. We take the 5σ flux density limits as a function of wavelength for our deepest exposure, 11 h in RXJ1347, (shown in Fig. 2) and calculate the S/N as estimated by the ETC. We use our flux calibration based on observations of standard stars to convert flux to e^-/s and rescaled by the wavelength-dependent sky transmission and KMOS throughput curve (both obtained through the KMOS ETC webpage). We use the following ETC settings, which are comparable to those of our observations: line $\mathrm{FWHM_{spec}}$ = 4 Å (unresolved); point source; seeing 0.″6; airmass: 1.50, Moon illumination FLI: 0.50, Moon-target separation: 45 deg, PWV: <2.5 mm. We calculate the S/N in an aperture with radius equal to the seeing FWHM ∼0.″6.

At every wavelength, the estimated S/N is:

$$\frac{\mathrm{S}}{\mathrm{N}} = \frac{\sqrt{\mathrm{NDIT}} \times S_{\mathrm{source}}}{\sqrt{S_{\mathrm{source}} + S_{\mathrm{bkg}} + n_{\mathrm{spat}}(\mathrm{DC} \times \mathrm{DIT} + \mathrm{RON}^2)}} \quad (8)$$

where for RXJ1347 NDIT = 88 is the number of DITs, of length DIT =450 s. The KMOS dark current (DC) is 0.01 e^-/pixel/s and the read-out noise is 3.5 e^-/pixel/DIT. The aperture corresponds to $n_{\mathrm{spat}} = 25$ spatial pixels and the calculation is done at the peak wavelength pixel. We use the online ETC to generate the background flux S_{bkg} in e^-/DIT as a function of wavelength, convolved with the instrumental resolution, given our input settings described above. We then calculate the estimated S/N using equation (8) at every wavelength using our 5σ flux density limits as the source flux.

In Fig. 8 we show a comparison of the ETC estimated S/N as a function of wavelength for the line fluxes corresponding to our 5σ limits. We plot the S/N estimated by the pipeline divided by 5 to show how the achieved S/N compares to the predicted S/N from the ETC. The public ETC does not account for noise due to sky subtraction routines. Assuming all DITs have equal noise σ, for 'A–B' frames the noise should be $\sqrt{2}\sigma$. Thus in Fig. 8 we also divide the ETC estimate by a factor of $\sqrt{2}$ for a fairer comparison with our data. We find that the ETC S/N is a median ∼1.4 × higher than our achieved values, and this overestimate is highest for wavelengths $\lesssim$ 11500 Å, where the ETC estimate can be ∼(1.6−1.8) × higher.

As shown in Fig. 8, the KMOS YJ throughput is known to decrease at $\lesssim$11 500 Å but our results suggest that the YJ grating is less sensitive in the blue for background-limited observations than expected.

Unfortunately this corresponds to Lyα redshifts $z \lesssim 8.5$, where we expect to find the majority of our targets. Using the S/N estimated from the ETC in planning our observations likely led us to overestimate the line sensitivity of KMOS for our targets. Most of the GLASS Lyα candidates we assigned to KMOS IFUs had tentative detections in the *HST* grisms. Thus a key aim of the deeper KMOS observations was to confirm these emission lines. While our deepest 1σ flux limit in our KMOS sample is 0.8×10^{-18} erg s^{-1} cm^{-2}, deeper than the 1σ flux limit in GLASS (5×10^{-18} erg s^{-1} cm^{-2}), we did not detect any emission from the tentative GLASS Lyα candidates with KMOS, suggesting that some of the *HST* grism lines were spurious noise fluctuations. A thorough comparison of the GLASS and KLASS observations, in combination with other follow-up at Keck, to determine the *HST* grism purity and completeness will be discussed in a future paper.

We advise any future KMOS users planning observations of faint targets to take into consideration both the additional noise from sky subtraction when using the KMOS ETC, and the lower than expected performance at the blue end of YJ. However, we find better agreement with the ETC estimates at redder wavelengths, demonstrating that KMOS YJ is performing well at $\gtrsim$ 11 500 Å.

6 SUMMARY AND CONCLUSIONS

We have presented an analysis of reionization epoch targets from KLASS, a large ground-based ESO VLT/KMOS program following up sources studied in the *HST* grism survey GLASS. Our main conclusions are as follows:

(i) The median 5σ flux limit of our survey is 4.5×10^{-18} erg s^{-1} cm^{-2}. We determine our spectroscopic survey to be 80 per cent complete over the full wavelength range for $7.2 \lesssim z \lesssim 10.1$ spatially unresolved Lyα emission lines with flux $\gtrsim 5.7 \times 10^{-18}$ erg s^{-1} cm^{-2}, centred within 0.″8 of the IFU centre and with intrinsic line $\mathrm{FWHM_{spec}} \lesssim 250$ km s^{-1}. Our observations are more complete

[5] https://www.eso.org/observing/etc/bin/gen/form?INS.NAME=KMOS+INS.MODE=lwspectr

to Lyα emission that may be spatially offset and/or extended compared to the UV continuum than typical slit-spectroscopy surveys.

(ii) Of the 52 $z \gtrsim 7$ candidate targets observed, none have confirmed Lyα emission, including those with candidate lines detected in the *HST* grisms. No other UV emission lines are detected at $z > 7$. We detect C IV emission in one image of a previously known C IV emitter at $z = 6.11$.

(iii) We define a sub-sample of 29 targets with a homogeneous photometric selection of $7.2 < z_{\rm phot} < 8.8$ for a Bayesian inference of the IGM neutral hydrogen fraction. The median Lyα flux limit for our sample is 3.6×10^{-18} erg s^{-1} cm^{-2} and the median Lyα EW upper limit is 58 Å. Combining our sub-sample with eight previously observed $z \sim 8$ LBGs from the BoRG survey (Trenti et al. 2011; Treu et al. 2013; Schmidt et al. 2014b), we obtain a lower limit on the IGM neutral hydrogen fraction at $z = 7.9 \pm 0.6$, $\overline{x}_{\rm H\,I} > 0.76$ (68 per cent) and $\overline{x}_{\rm H\,I} > 0.46$ (95 per cent).

(iv) Our constraint favours a late reionization consistent with models where ultrafaint galaxies contribute significantly to reionization, with an ionizing photon escape fraction $\langle f_{\rm esc} \rangle \lesssim 15$ per cent.

Our KMOS observations provide more evidence of a predominantly neutral IGM at $z \sim 8$. To make more precise constraints on the timeline of reionization will require larger samples of LBGs with precise photometric (or even better, spectroscopic) redshift estimates, more informative priors on Lyα FWHM, and deep spectroscopic limits on Lyα. Forthcoming deep spectroscopic observations with JWST (e.g. Treu et al. 2017) will provide ideal samples for future inferences on reionization.

ACKNOWLEDGEMENTS

The authors thank the referee for their constructive comments. We thank Trevor Mendel and Owen Turner for useful discussions related to KMOS reductions for faint sources, and T. Mendel for sharing the readout correction code. We thank Andrei Mesinger for providing Lyα optical depths from the EoS simulations.

C.M. acknowledges support by NASA Headquarters through the NASA Earth and Space Science Fellowship Program Grant NNX16AO85H, and through the NASA Hubble Fellowship grant HST-HF2-51413.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. This work was supported by the *HST* GLASS grant GO-13459, the *HST* BoRG grants GO-12572, 12905, 13767, and 15212, and *HST*-AR-13235 and *HST*-AR-14280.

This work was based on observations collected at the European Organisation for Astronomical Research in the Southern Hemisphere under ESO program 196.A-0778; and on observations made with the NASA/ESA *Hubble Space Telescope*, obtained at STScI. We are very grateful to all ESO and STScI staff who have assisted in these observations.

This work utilizes gravitational lensing models produced by PIs Bradač, Natarajan & Kneib (CATS), Merten & Zitrin, Sharon, Williams, Keeton, Bernstein and Diego, and the GLAFIC group. This lens modelling was partially funded by the *HST* Frontier Fields program conducted by STScI. The lens models were obtained from the Mikulski Archive for Space Telescopes (MAST).

Software: IPYTHON (Pérez & Granger 2007), MATPLOTLIB (Hunter 2007), NUMPY (Van Der Walt, Colbert & Varoquaux 2011), SCIPY (Oliphant 2007), ASTROPY (Robitaille et al. 2013), and QFITSVIEW (http://www.mpe.mpg.de/~ott/QFitsView/).

REFERENCES

Alavi A. et al., 2014, ApJ, 780, 143
Bacon R. et al., 2014, A&A, 575, A75
Balestra I. et al., 2013, A&A, 559, L9
Balestra I. et al., 2016, ApJS, 224, 33
Bañados E. et al., 2017, Nature, 553, 473
Becker G. D., Davies F. B., Furlanetto S. R., Malkan M. A., Boera E., Douglass C., 2018, ApJ, 863, 92
Blanc G. A. et al., 2011, ApJ, 736, 31
Blanton M. R., Roweis S., 2007, AJ, 133, 734
Boone F. et al., 2013, A&A, 559, L1
Bouwens R. J. et al., 2003, ApJ, 595, 589
Bouwens R. J. et al., 2012, ApJ, 752, L5
Bouwens R. J. et al., 2014, ApJ, 793, 115
Bouwens R. J., Illingworth G. D., Oesch P., Caruana J., Holwerda B., Smit R., Wilkins S., 2015, ApJ, 811, 140
Bouwens R. J., Illingworth G. D., Oesch P. A., Atek H., Lam D., Stefanon M., 2017, ApJ, 843, 41
Bradač M. et al., 2014, ApJ, 785, 108
Bradač M. et al., 2017, ApJ, 836, L2
Bradley L. D. et al., 2012, ApJ, 760, 108
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Bunker A. J., Stanway E. R., Ellis R. S., McMahon R. G., 2004, MNRAS, 355, 374
Caminha G. B. et al., 2016, A&A, 595, A100
Caminha G. B. et al., 2017, A&A, 600, A90
Carniani S. et al., 2017, A&A, 605, A42
Caruana J., Bunker A. J., Wilkins S. M., Stanway E. R., Lacy M., Jarvis M. J., Lorenzoni S., Hickey S., 2012, MNRAS, 427, 3055
Caruana J., Bunker A. J., Wilkins S. M., Stanway E. R., Lorenzoni S., Jarvis M. J., Ebert H., 2014, MNRAS, 443, 2831
Cassata P. et al., 2015, A&A, 573, A24
Castellano M. et al., 2012, A&A, 540, A39
Castellano M. et al., 2016, A&A, 590, A31
Christensen L. et al., 2012, MNRAS, 427, 1973
Curtis-Lake E. et al., 2012, MNRAS, 422, 1425
Davies F. B. et al., 2018, ApJ, 864, 2
Davies F. B., Furlanetto S. R., 2016, MNRAS, 460, 1328
Davies R. I. et al., 2013, A&A, 558, A56
Davies R. I., 2007, MNRAS, 375, 1099
Dayal P., Maselli A., Ferrara A., 2011, MNRAS, 410, 830
De Barros S. et al., 2017, A&A, 608, A123
Dijkstra M., 2014, PASA, 31, e040
Dijkstra M., Haiman Z., Spaans M., 2006, ApJ, 649, 37
Dijkstra M., Mesinger A., Wyithe J. S. B., 2011, MNRAS, 414, 2139
Di Criscienzo M. et al., 2017, A&A, 607, A30
Epinat B. et al., 2009, A&A, 546, A118
Faisst A. L., Capak P., Carollo C. M., Scarlata C., Scoville N., 2014, ApJ, 788, 87
Fan X. et al., 2001, AJ, 122, 2833
Fan X. et al., 2006, AJ, 132, 117
Feldmeier J. J. et al., 2013, ApJ, 776, 75
Finkelstein S. L. et al., 2013, Nature, 502, 524
Flower D. R., Nussbaumer H., Schild H., 1979, A&A, 72, L1
Fontana A. et al., 2010, ApJ, 725, L205
Förster Schreiber N. M. et al., 2009, ApJ, 706, 1364
Genzel R. et al., 2017, Nature, 543, 397
Greig B., Mesinger A., 2015, MNRAS, 449, 4246
Greig B., Mesinger A., 2017, MNRAS, 465, 4838
Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017, MNRAS, 466, 4239
Greig B., Mesinger A., Bañados E., 2019, MNRAS, 484, 4
Grillo C. et al., 2016, ApJ, 822, 78
Haiman Z., 2002, ApJ, 576, L1
Hashimoto T. et al., 2018, Nature, 557, 392
Hayes M., Schaerer D., Östlin G., Mas-Hesse J. M., Atek H., Kunth D., 2011, ApJ, 730, 8

Henry A., Berg D., Scarlata C., Verhamme A., Erb D., 2018, ApJ, 855, 96
Hoag A. et al., 2017, Nat. Astron., 1, 0091
Hoag A. et al., 2018, ApJ, 854, 39
Hoag A. et al., 2019, ApJ, preprint (arXiv:1901.09001)
Huang K.-H. et al., 2016a, ApJ, 817, 11
Huang K.-H. et al., 2016b, ApJ, 823, L14
Hunter J. D., 2007, Comput. Sci. Eng., 9, 99
Ishigaki M., Kawamata R., Ouchi M., Oguri M., Shimasaku K., Ono Y., 2018, ApJ, 854, 73
Jaskot A. E., Oey M. S., 2014, ApJ, 791, L19
Jones T., Stark D. P., Ellis R. S., 2012, ApJ, 751, 51
Jung I. et al., 2018, ApJ, 864, 103
Karman W. et al., 2015, A&A, 574, A11
Karman W. et al., 2016, A&A, 599, 43
Kistler M. D., Yüksel H., Beacom J. F., Hopkins A. M., Wyithe J. S. B., 2009, ApJ, 705, L104
Laporte N. et al., 2017a, ApJ, 837, L21
Laporte N., Nakajima K., Ellis R. S., Zitrin A., Stark D. P., Mainali R., Roberts-Borsani G. W., 2017b, ApJ, 851, 40
Leclercq F. et al., 2017, A&A, 608, A8
Lehnert M. D., Bremer M., 2003, ApJ, 593, 630
Liu A., Parsons A. R., 2016, MNRAS, 457, 1864
Livermore R. C., Finkelstein S. L., Lotz J. M., 2017, ApJ, 835, 113
Livermore R. C., Trenti M., Bradley L. D., Bernard S. R., Holwerda B. W., Mason C. A., Treu T., 2018, ApJ, 861, L17
Lotz J. M. et al., 2017, ApJ, 837, 97
Mahler G. et al., 2018, MNRAS, 473, 663
Mainali R. et al., 2018, MNRAS, 479, 1180
Mainali R., Kollmeier J. A., Stark D. P., Simcoe R. A., Walth G., Newman A. B., Miller D. R., 2017, ApJ, 836, L14
Manti S., Gallerani S., Ferrara A., Greig B., Feruglio C., 2016, MNRAS, 466, 1160
Marchi F. et al., 2018, A&A, 614, A11
Mason C. A. et al., 2017, ApJ, 838, 14
Mason C. A. et al., 2018b, ApJ, 857, L11
Mason C. A., Trenti M., Treu T., 2015, ApJ, 813, 21
Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018a, ApJ, 856, 2
Matthee J., Sobral D., Gronke M., Paulino-Afonso A., Stefanon M., Röttgering H., 2018, A&A, 619, A136
McGreer I. D., Mesinger A., D'Odorico V., 2014, MNRAS, 447, 499
McQuinn M., Lidz A., Zahn O., Dutta S., Hernquist L., Zaldarriaga M., 2007, MNRAS, 377, 1043
Merlin E. et al., 2016, A&A, 590, A30
Mesinger A., Aykutalp A., Vanzella E., Pentericci L., Ferrara A., Dijkstra M., 2015, MNRAS, 446, 566
Mesinger A., Greig B., Sobacchi E., 2016, MNRAS, 459, 2342
Mirocha J., Furlanetto S. R., Sun G., 2016, MNRAS, 6, 1365
Momose R. et al., 2014, MNRAS, 442, 110
Monna A. et al., 2014, MNRAS, 438, 1417
Monna A. et al., 2017, MNRAS, 466, 4094
Morishita T., Abramson L. E., Treu T., Schmidt K. B., Vulcani B., Wang X., 2017, ApJ, 846, 139
Naidu R. P., Forrest B., Oesch P. A., Tran K.-V. H., Holden B. P., 2018, MNRAS, 478, 791
Oesch P. A. et al., 2015, ApJ, 804, L30
Oliphant T. E., 2007, Comput. Sci. Eng., 9, 10
Ono Y. et al., 2012, ApJ, 744, 83
Ouchi M. et al., 2010, ApJ, 723, 869
Oyarzún G. A. et al., 2016, ApJ, 821, L14
Oyarzún G. A., Blanc G. A., González V., Mateo M., Bailey J. I., 2017, ApJ, 843, 133
Parsa S., Dunlop J. S., McLure R. J., 2018, MNRAS, 474, 2904
Parzen E., 1962, Ann. Math. Stat., 33, 1065
Pentericci L. et al., 2011, ApJ, 743, 132
Pentericci L. et al., 2014, ApJ, 793, 113
Pentericci L. et al., 2018, A&A, 619, A147
Pérez F., Granger B. E., 2007, Comput. Sci. Eng., 9, 21
Planck Collaboration VIII, 2015, A&A, 594, A8
Planck Collaboration XLVII, 2016, A&A, 596, A108
Postman M. et al., 2012, ApJS, 199, 25
Ren K., Trenti M., Mutch S. J., 2018, ApJ, 856, 81
Roberts-Borsani G. W. et al., 2016, ApJ, 823, 143
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
Robitaille T. P. et al., 2013, A&A, 558, A33
Rosenblatt M., 1956, Ann. Math. Stat., 27, 832
Rutkowski M. J. et al., 2017, ApJ, 841, L27
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
Schenker M. A., Ellis R. S., Konidaris N. P., Stark D. P., 2014, ApJ, 795, 20
Schmidt K. B. et al., 2014a, ApJ, 782, L36
Schmidt K. B. et al., 2014b, ApJ, 786, 57
Schmidt K. B. et al., 2016, ApJ, 818, 38
Schmidt K. B. et al., 2017, ApJ, 839, 17
Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., 2003, ApJ, 588, 65
Sharples R. et al., 2013, The Messenger, 151, 21
Shibuya T., Kashikawa N., Ota K., Iye M., Ouchi M., Furusawa H., Shimasaku K., Hattori T., 2012, ApJ, 752, 114
Shibuya T., Ouchi M., Harikane Y., 2015, ApJS, 219, 15
Shull J. M., Harness A., Trenti M., Smith B. D., 2012, ApJ, 747, 100
Simons R. C. et al., 2017, ApJ, 843, 46
Smit R. et al., 2018, Nature, 553, 178
Sobacchi E., Mesinger A., 2015, MNRAS, 453, 1843
Song M. et al., 2016a, ApJ, 825, 5
Song M., Finkelstein S. L., Livermore R. C., Capak P. L., Dickinson M., Fontana A., 2016b, ApJ, 826, 113
Stanway E. R., McMahon R. G., Bunker A. J., 2005, MNRAS, 359, 1184
Stark D. P. et al., 2014, MNRAS, 445, 3200
Stark D. P. et al., 2015, MNRAS, 454, 1393
Stark D. P. et al., 2017, MNRAS, 464, 469
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
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
Stott J. P. et al., 2016, MNRAS, 457, 1888
Tamura Y. et al., 2018, preprint (arXiv:1806.04132)
Tanvir N. R. et al., 2012, ApJ, 754, 46
Tilvi V. et al., 2014, ApJ, 794, 5
Tilvi V. et al., 2016, ApJ, 827, L14
Trainor R. F., Steidel C. C., Strom A. L., Rudie G. C., 2015, ApJ, 809, 89
Trebitsch M., Blaizot J., Rosdahl J., Devriendt J., Slyz A., 2017, MNRAS, 470, 224
Trenti M. et al., 2011, ApJ, 727, L39
Trenti M., Perna R., Levesque E. M., Shull J. M., Stocke J. T., 2012, ApJ, 749, L38
Treu T. et al., 2015, ApJ, 812, 114
Treu T. et al., 2017, JWST Proposal ID 1324. Cycle 0 Early Release Science, STScI, JWST proposal
Treu T., Trenti M., Stiavelli M., Auger M. W., Bradley L. D., 2012, ApJ, 747, 27
Treu T., Schmidt K. B., Trenti M., Bradley L. D., Stiavelli M., 2013, ApJ, 775, L29
Van Der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22
Vanzella E. et al., 2011, ApJ, 730, L35
Vanzella E. et al., 2016, ApJ, 821, L27
Verhamme A. et al., 2018, MNRAS, 478, L60
Verhamme A., Schaerer D., Maselli A., 2006, A&A, 460, 397
Verhamme A., Orlitová I., Schaerer D., Hayes M., 2015, A&A, 578, A7
Vulcani B., Trenti M., Calvi V., Bouwens R., Oesch P., Stiavelli M., Franx M., 2017, ApJ, 836, 239

Wang X. et al., 2016, ApJ, 837, 89
Weisz D. R., Boylan-Kolchin M., 2017, MNRAS, 469, L83
Wilkins S. M., Bunker A. J., Stanway E., Lorenzoni S., Caruana J., 2011, MNRAS, 417, 717
Wisnioski E. et al., 2015, ApJ, 799, 209
Wisotzki L. et al., 2016, A&A, 587, A98
Yan H., Windhorst R. A., 2004, ApJ, 600, L1
Yang H., Malhotra S., Gronke M., Rhoads J. E., Dijkstra M., Jaskot A., Zheng Z., Wang J., 2016, ApJ, 820, 130
Zheng W. et al., 2012, Nature, 489, 406
Zheng Z.-Y. et al., 2017, ApJ, 842, L22
Zitrin A. et al., 2015, ApJ, 810, L12

APPENDIX A: INDEPENDENT CONFIRMATION OF CIV EMISSION AT $z = 6.11$

As well as Lyα candidates we also targeted three images of a multiply imaged $z = 6.11$ galaxy in RXJ2248 to observe C IVλ1548,1551 emission. This multiple-image system has been detected with Lyα emission by Boone et al. (2013), Balestra et al. (2013), Monna et al. (2014), Karman et al. (2015), Schmidt et al. (2016), and Mainali et al. (2017). Detections of C IV and O III] λ1666 emission in one of the images were also presented by Mainali et al. (2017), and C IV detections and limits in all of the images by Schmidt et al. (2017). The presence of these highly ionized lines and lack of observed He IIλ1640 indicates this this object has a hard ionizing spectrum, but unlikely to be dominated by an AGN.

In Fig. A1 we show our GLASS and KMOS spectra for the brightest image E. An emission line is seen in the KMOS data at the same wavelength as Mainali et al. (2017), and appears to be spatially compact with size approximately equal to our seeing (∼0.″6). However, due to the lower spectral resolution of our data ($R \sim 3400$ compared to $R \sim 6000$ in Mainali et al. 2017), the emission line is overlapping with an adjacent skyline, adding some uncertainty to the extracted parameters.

We fit a model with 2 Gaussian emission lines to an extracted 1D spectrum, weighted by the inverse variance, fixing $\mathrm{FWHM_{spec}} = 5$ Å (close to the instrumental resolution, 4 Å) and allowing flux and the wavelength position of the doublet centre to vary. We note that the spatial aperture we use to extract the spectrum (1.2 sq arcsec) is slightly larger than the one used to plot Fig. A1 (0.8 sq arcsec), to ensure we recover the full flux. We use the smaller aperture in the plot to maximize the plotted S/N and for comparison with the line identification technique described in Section 3.1, which also uses 0.8 sq arcsec apertures.

The strongest peak is fit at 11023.7 Å, which Mainali et al. (2017) assigned to the 1551 Å emitted peak. There is a weaker peak at 11002.5 Å, which would correspond to the 1548 Å emitted peak. We note that in the 1D spectrum there does appear to be a peak redward of the strong peak, around 11040 Å, which could alternatively be the 1551 Å emitted peak. However, the peak separation to the 11040 Å line is too low for this to be part of the doublet and by inspection of the 2D emission postage stamps in the KMOS cubes we see the 11002.5 Å flux is spatially coincident with the 11023.7 Å emission, whilst the 11040 Å flux is more spatially uniform and thus likely

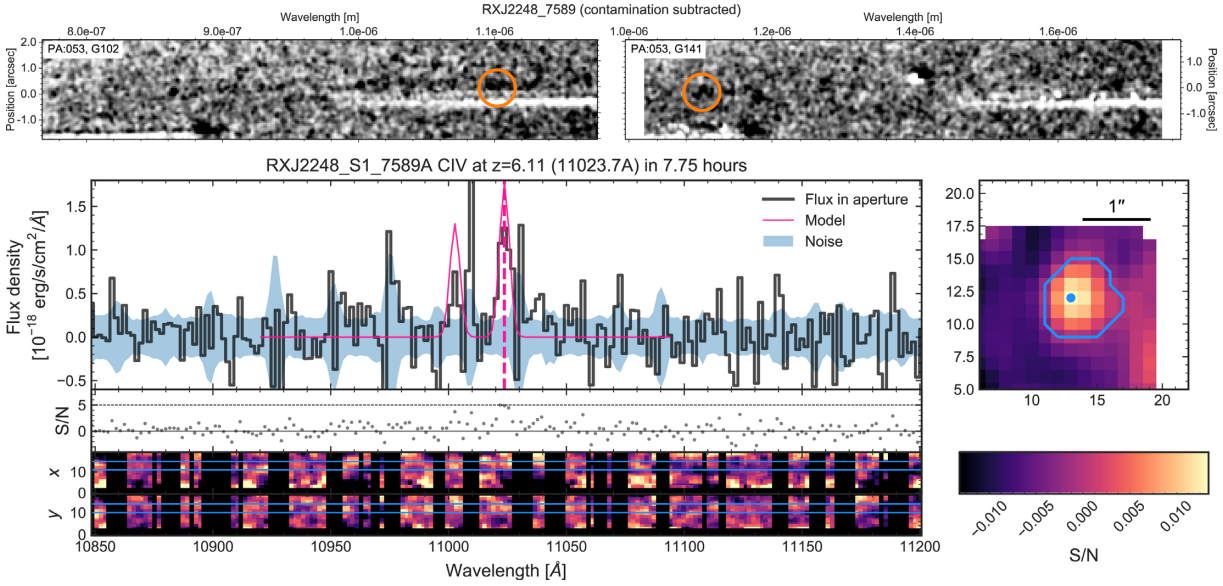

Figure A1. *Upper panels:* GLASS *HST* grism spectra taken at two position angles (PA) in G102 and G141. Positive flux is shown in black, and white is negative flux. The candidate emission line is within the orange circle. *Lower left-hand panel*: KLASS KMOS spectra of the same object, centred at the wavelength of the GLASS candidate emission. The top left-hand panel shows a 1D spectrum (flux – black line, noise – blue-shaded region) extracted in an aperture shown by the blue lines in the 2D postage stamp image on the right (an aperture containing the brightest 10 per cent of pixels, with area 0.8 sq arcsec). The wavelength of the 1551 Å emission line is shown with the pink-dashed vertical line. The model for the doublet using 2 Gaussians, extracting voxels in a 1.2 sq arcsec aperture (shown as the blue contour on the 2D image – lower right panel), is shown as a pink solid line. The middle panel shows the S/N at each wavelength pixel. The lower spectra show simulated slit spectra along the x and y directions, with the same aperture plotted as blue horizontal lines. Regions with strong sky emission are masked. The spectra are smoothed with a 3D Gaussian kernel with line spread function $\mathrm{FWHM_{spec}}$ equal to the instrumental resolution (4 Å) and point spread function $\mathrm{FWHM_{spat}}$ equal to the seeing (0.″6). *Lower right panel:* 2D flux postage stamp image of the emission summed over a 10 Å wavelength range centred at the wavelength shown by the pink dashed vertical line in the left-hand panel. The emission is clearly spatially compact. The blue contour shows the spatial aperture used to extract the 1D spectra. The S/N colourbar refers to the 2D slit spectra. We have performed an additional residual sky subtraction to the KLASS cube by subtracting a median 1D spectrum obtained in spaxels away from the emission line.

to be spurious noise peak. Thus our assignment of the doublet wavelengths is consistent with the observations of Mainali et al. (2017).

Using our model, we find a total line flux of $(1.6 \pm 0.3) \times 10^{-17}$ erg s^{-1} cm^{-2}. We find a flux ratio of C IVλ1548:C IVλ1551 = 0.7:1, which is much lower than the theoretically motivated expected value of 2:1 (e.g. Flower, Nussbaumer & Schild 1979). Low-mass metal-poor galaxies at $z \sim 2-3$ have been observed with flux ratios both comparable to the theoretical value (Caminha et al. 2016; Vanzella et al. 2016) and closer to a 1:1 ratio (Christensen et al. 2012; Stark et al. 2014).

However, the confusion with the adjacent skylines makes it difficult to measure an accurate flux. Our measured total flux is lower than those measured by Mainali et al. (2017) and Schmidt et al. (2017), but consistent with the value measured by Schmidt et al. (2017) within 2σ. We also find a tentative (S/N ~3) detection at the wavelength where O III] λ1666 was identified by Mainali et al. (2017) in our data (11837.1 Å).

APPENDIX B: TESTING THE SELECTION FUNCTION OF OUR FINAL SUB-SAMPLE

As described in Section 4.2 we define a selection function based on the photometric redshift distributions to select a sub-sample of KLASS targets to use in the reionization inference. In this Appendix we demonstrate that the sub-sample was not a biased selection from the final parent catalogues.

Objects in the sub-sample must have a match in our final deep photometric catalogues: for A2744, M0416, and M1149, we used the ASTRODEEP catalogues (Castellano et al. 2016; Merlin et al. 2016; Di Criscienzo et al. 2017). For M2129, RXJ1347, and RXJ2248, we created our own catalogues based on the ASTRODEEP methodology. And the objects must have $P(7.2 \geq z_{\rm phot} \geq 8.8) > 0.01$ based on photometric redshift probability distributions calculated using EAzY (Brammer et al. 2008). We use the full photometric redshift distribution in our inference to robustly weight objects based on their probability of being in our redshift range of interest (Section 4.1 and Appendix C).

However, given that we construct this sub-sample *after* the observations were taken, we must check that the objects we observed were not a biased sample from the final catalogues. There are many more objects in the final catalogues that were not observed in KMOS so it could be that the KLASS targets are a biased sample of the final catalogue.

In Fig. B1 we show the distribution of $P(7.2 \geq z_{\rm phot} \geq 8.8)$ values of our KLASS sub-sample and the parent catalogues. The parent catalogues have many more objects with low $P(7.2 \geq z_{\rm phot} \geq 8.8)$, showing that our sub-sample is skewed towards objects that are most likely to be within that redshift range, demonstrating that our initial $z_{\rm phot} > 7.2$ target selection was good. In the right-hand panel we show the probability distribution of objects in the parent catalogues and our sub-sample, using only objects with $P(7.2 \geq z_{\rm phot} \geq 8.8) > 0.6$. As noted previously, we do not expect all the LBGs to have $P(7.2 \geq z_{\rm phot} \geq 8.8) = 1$ as there are often degeneracies in the photometry that make $z_{\rm phot} \sim 1-2$ solutions possible. Here we see that the distributions are consistent and demonstrate that the best targets in KLASS sub-sample (which weight most in the reionization inference) are drawn randomly from the best targets in the parent catalogues.

We also want to be sure our sub-sample is drawn randomly from the parent sample in terms of photometry. In particular, if we had selected only the reddest objects for our KMOS targets, our sub-sample could be biased towards dustier galaxies. As dust can significantly attenuate Lyα emission (Hayes et al. 2011), this would mean our non-detections of Lyα could be due to stronger dust attenuation as well as reionization. However, we demonstrate in Fig. B2 that the F125W−F160W colour distribution and the UV slope (β, where flux $f_\lambda \propto \lambda^\beta$) distribution of our sub-sample is consistent with that of the parent catalogues. We measure UV slopes for the GLASS and KLASS objects by fitting to the F125W and

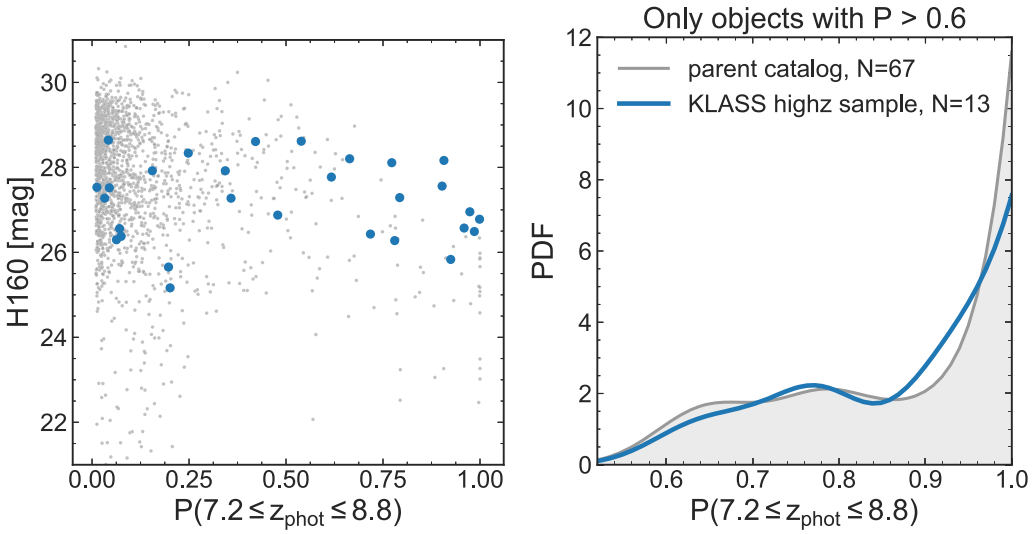

Figure B1. *Left*: Total probability (from the EAzY photometric redshift distributions) of objects being within our redshift range of interest ($P(7.2 \geq z_{\rm phot} \geq 8.8) > 0.01$) versus their F160W apparent magnitude. The KLASS sub-sample used for the inference is shown as large blue circles, and objects from the full parent photometric catalogues are shown as small grey dots. Our sample is skewed towards higher probability of $7.2 \geq z_{\rm phot} \geq 8.8$ compared to the parent sample, with a smaller range in F160W magnitude. *Right*: Probability distribution functions of $P(7.2 \geq z_{\rm phot} \geq 8.8)$ values of all objects with $P(7.2 \geq z_{\rm phot} \geq 8.8) > 0.6$ from our KLASS sub-sample (blue) and the parent catalogues (grey). The distributions are plotted using a Gaussian kernel density estimator (Rosenblatt 1956; Parzen 1962). The two distributions are very similar, demonstrating our KLASS sub-sample is drawn randomly from the parent catalogue sample, and thus is not a biased sample in terms of photometric redshift distribution, despite being constructed after the observations were taken.

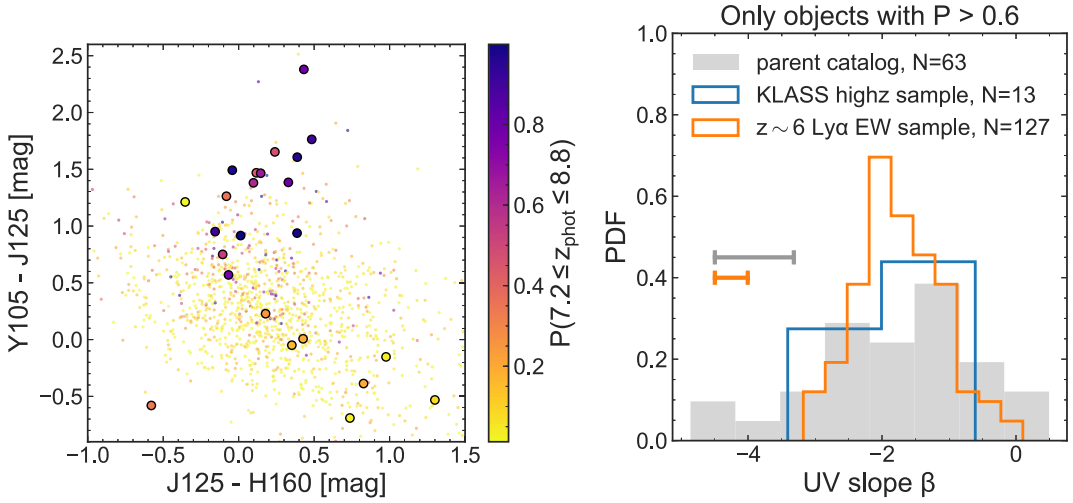

Figure B2. *Left*: F125W–F160W colour versus F105W–F125W colour for our KLASS inference sub-sample (large circles with black edges) and the parent catalogues (small points). The markers are colour-coded by their $P(7.2 \geq z_{\rm phot} \geq 8.8)$ so that the darkest points are the most likely to be within that redshift range. The most probable objects in the KLASS sub-sample appear to be in the same region of colour–colour space as the full parent catalogue. *Right*: Probability distribution functions of UV β slope values of all objects with $P(7.2 \geq z_{\rm phot} \geq 8.8) > 0.6$ from our KLASS sub-sample (blue) and the parent catalogues (grey). As in Fig. B1 the two distributions are very similar, demonstrating our KLASS sub-sample is not a strongly biased sample in terms of colour, despite being constructed after the observations were taken. We also plot the β distribution for the $z \sim 6$ (De Barros et al. 2017) sample used to create the intrinsic Lyα EW distribution for our inference (orange). Median errorbars for the GLASS parent sample and De Barros et al. (2017) sample are shown in grey and orange, respectively. Using a Kolmogorov–Smirnov test, we find the De Barros et al. (2017) sample, KLASS high-z sample, and GLASS parent catalogue β distributions are all consistent with being drawn from the same population.

F160W magnitudes (Castellano et al. 2012). The parent catalogues contain a small fraction (< 10 per cent) of redder objects that are not present in our KLASS sub-sample, but given the small number of objects in our sub-sample missing this small fraction of redder objects is expected.

In addition, in the right-hand panel of Fig. B2 we compare the UV slope distribution of the $z \sim 6$ EW calibration sources used for the inference EW models (De Barros et al. 2017) to our GLASS and KLASS catalogues. Again, we want to be sure that the (De Barros et al. 2017) sample and KLASS sample are similar as we assume that these $z \sim 6$ sources are a good proxy for the $z \sim 8$ galaxies in our inference. We plot the UV slope distributions and the median errors on the β measurements. For the KLASS $z \sim 8$ objects we can only fit the slopes using F125W and F160W magnitudes, so these measurements have large uncertainties. Using a Kolmogorov–Smirnov test, we find that the De Barros et al. (2017) sample, KLASS high-z sample, and GLASS parent catalogue β distributions are all consistent with being drawn from the same population. Therefore, differences in the Lyα transmission from $z \sim 6$ to $z \sim 8$ due to dust absorption are likely to be negligible. Spectral coverage at $> 2\,\mu$m with JWST will hugely improve constraints on UV slopes for $z \gtrsim 6$ galaxies and enable a better understanding of how dust mediates Lyα escape at high redshifts.

APPENDIX C: REIONIZATION INFERENCE

This appendix derives the posterior distribution for the IGM neutral fraction, $\overline{x}_{\rm H\,I}$, extending the framework of M18a from single EW measurements to an input flux density spectrum as a function of wavelength, and galaxy UV apparent magnitude and gravitational lensing magnification.

C1 Likelihood at one spectral pixel

We want to obtain the likelihood of observing a flux density spectrum $\{f\}$ with flux f_i at wavelength pixel i given our the neutral fraction $\overline{x}_{\rm HI}$ and the properties of that observed galaxy.

The likelihood of measuring flux density f_i at wavelength λ_i given that the photons originate at z_d from a galaxy with apparent magnitude m and travel through an IGM with neutral fraction $\overline{x}_{\rm H\,I}$ is

$$p(f_i \mid \overline{x}_{\rm HI}, m, \mu, z_d, {\rm FWHM}) = \int_0^\infty {\rm d}W\; p(f_i \mid W, m, z_d, {\rm FWHM})\; p(W \mid \overline{x}_{\rm HI}, m, \mu, z_d) \tag{C1}$$

Including Gaussian errors in the spectra (with error σ_i at spectral pixel i), the probability of measuring flux density f_i at spectral pixel i is given by a Gaussian distribution at each spectral pixel, with mean given by the model flux density for a given equivalent width and standard deviation σ_i:

$$p(f_i \mid W, m, z_d, {\rm FWHM}) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(f_i - f_{\rm mod}(\lambda_i, W, m, z_d, {\rm FWHM}))^2}{2\sigma_i^2}\right] \tag{C2}$$

The total flux of the model emission line is given by $F_{\rm tot} = W f_{\rm cont}(1 + z)$. For simplicity, we model emission lines as Gaussians, so that the flux density produced at a single spectral pixel i by an emission line at pixel d (at wavelength $\lambda_d = \lambda_\alpha(1 + z_d)$) is

$$f_{\rm mod}(\lambda_i, W, m, z_d, {\rm FWHM}) = \frac{W f_{cont}(m, z_d)(1 + z_d)}{\sqrt{2\pi}\sigma_\lambda} \exp\left[-\frac{(\lambda_i - \lambda_d)^2}{2\sigma_\lambda^2}\right] \tag{C3}$$

where m is the observed apparent magnitude of the source, and σ_λ = FWHM/2.355 is the spectral linewidth.

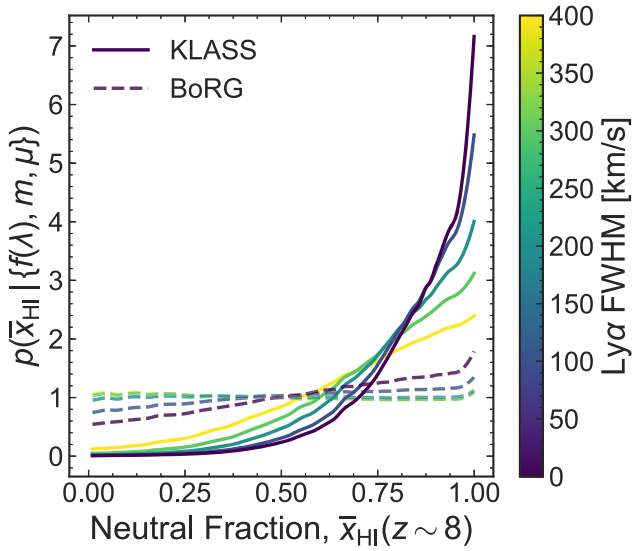

Figure C1. Posterior probability distribution for the IGM neutral fraction derived using the KLASS (solid lines) and BORG (dashed lines) observations, as a function of the assumed FWHM (shown by the different colours) for the samples. As the assumed Lyα FWHM increases, the inferred posterior flattens. This is because with increasing FWHM, our EW sensitivity decreases.

The strength of our inferred limit on the neutral fraction will depend on the choice of linewidth, as that determines the EW sensitivity. We discuss our choice of these values below in Appendix C4.

The second term on the right-hand side of equation (C1) can be expanded as

$$p(W\,|\,\overline{x}_{\rm H\,I}, m, \mu, z_d) = \int_{-\infty}^{\infty} {\rm d}M_{\rm UV}\; p(W\,|\,\overline{x}_{\rm H\,I}, M_{\rm UV}) p(M_{\rm UV}\,|\,m, \mu, z_d) \qquad \text{(C4)}$$

where the integral convolves the simulated $p(W\,|\,\overline{x}_{\rm H\ I}, M_{\rm UV})$ from M18a with the probability distribution of the absolute UV magnitude, $M_{\rm UV}$, given our observed data:

$$p(M_{\rm UV}\,|\,m, \mu, z_d) = \frac{1}{\sqrt{2\pi\sigma_M^2}} \exp\left[-\frac{\left(M_{\rm UV} - M_{\rm UV,mod}(\mu, z_d)\right)^2}{2\sigma_M^2}\right] \qquad \text{(C5)}$$

where $M_{\rm UV,mod}(\mu, z_d) = m - 5\log_{10}(D_L/10\,{\rm pc}) + 2.5\log_{10}(1+z_d) + 2.5\log_{10}\mu$ converts observed magnitudes to rest-frame UV magnitudes, assuming the UV spectral slope $\beta = -2$ to calculate the K-correction (e.g. Blanton & Roweis 2007). We assume that the magnification distribution is log-normally distributed such that we can easily add the uncertainties in magnification and apparent magnitude: $\sigma_M^2 = \sigma_m^2 + (2.5\sigma_{\log\mu})^2$, where σ_m is the error on the observed apparent magnitude and $\sigma_{\log\mu}$ is the uncertainty in the logarithmic magnifications.

We note that the dependence on $M_{\rm UV}$ in $p(W\,|\,\overline{x}_{\rm H\,I}, M_{\rm UV})$ is weak compared to the dependence on $\overline{x}_{\rm H\ I}$, and was parametrized with a smooth transition between two EW distributions for $M_{\rm UV} > -20$ and $M_{\rm UV} < -21$. Thus, only the parameters thta dominate changes in $M_{\rm UV}$ are important. We note that the distance modulus term changes the magnitude in our redshift range of interest, $z = 7.2-8.8$ ($\Delta M_{\rm UV} \sim 0.5$) less than magnification ($\Delta M_{\rm UV} \sim 0.75$ for $\mu = 2$), so for ease of computation we compute $p(W\,|\,\overline{x}_{\rm H\,I}, m, \mu, z_d)$ ahead of time for each galaxy, setting $z_d = 8$, rather than having to redo this integral at every spectral pixel.

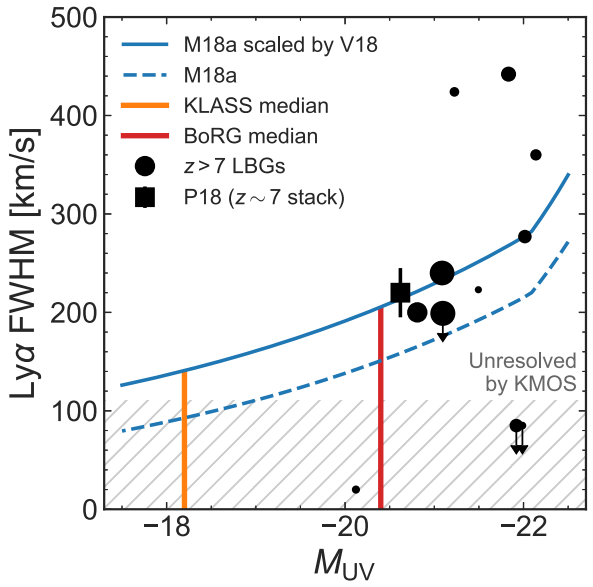

Figure C2. Lyα spectral FWHM as a function of UV magnitude. Black circles are measurements from LBGs with confirmed $z > 7$ Lyα (Vanzella et al. 2011; Ono et al. 2012; Schenker et al. 2012; Shibuya et al. 2012; Finkelstein et al. 2013; Oesch et al. 2015; Zitrin et al. 2015; Roberts-Borsani et al. 2016; Song et al. 2016a; Stark et al. 2017; Laporte et al. 2017a). The black square shows median values derived from a $z \sim 7$ sample by Pentericci et al. (2018). The blue dashed line shows the M18a, model for velocity offset (Δv) as a function of UV magnitude and redshift (via halo mass). The blue solid line shows the M18a model scaled by the FWHM–$\Delta v_{\rm Ly\alpha}$ relation presented by Verhamme et al. (2018). The grey-hatched region shows FWHM unresolved by KMOS ($\lesssim 4$ Å or $\lesssim 110\,{\rm km\,s^{-1}}$). The orange (red) vertical line shows the median UV magnitude of the KLASS (BoRG) samples.

C2 Likelihood for a full spectrum

For a full spectrum $\{f\} = f(\lambda_i)$ the likelihood is just the product of the likelihoods at each wavelength pixel:

$$p(\{f\}\,|\,\overline{x}_{\rm H\,I}, m, \mu, z_d, {\rm FWHM}) = \prod_i^N \int_0^\infty {\rm d}W\; \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\left(\frac{f_i - f_{\rm mod,\,i}}{\sigma_i}\right)^2} p(W\,|\,\overline{x}_{\rm H\,I}, m, \mu, z_d) \qquad \text{(C6)}$$

where d is the index of the emission line, and $z_d = \lambda_d/\lambda_\alpha - 1$, and $f_{\rm mod,\,i}$ is given by equation (C3).

C3 Posteriors

Using Bayes' Theorem the posterior distribution for $\overline{x}_{\rm H\,I}$, FWHM and z_d is

$$p(\overline{x}_{\rm H\,I}, z_d, {\rm FWHM}\,|\,\{f\}, m, \mu) \propto p(\{f\}\,|\,\overline{x}_{\rm H\,I}, m, \mu, z_d, {\rm FWHM}) \times p(z_d)\, p(\overline{x}_{\rm H\,I})\, p({\rm FWHM}) \qquad \text{(C7)}$$

We use a uniform prior on $\overline{x}_{\rm H\,I}$ between 0 and 1, we use the photometric redshift for the prior $p(z)$, and we use a lognormal prior on FWHM (discussed in Section C4). As we are interested only in the posterior probability of $\overline{x}_{\rm H\,I}$ we can marginalize over z_d and FWHM:

$$p(\overline{x}_{\rm H\,I}\,|\,\{f\}, m, \mu) = \int {\rm dFWHM} \int {\rm d}z_d\; p(\overline{x}_{\rm H\,I}, z_d, {\rm FWHM}\,|\,\{f\}, m, \mu) \qquad \text{(C8)}$$

To account for the incomplete wavelength coverage, we make use of the fact if the object has Lyα outside of the wavelength range

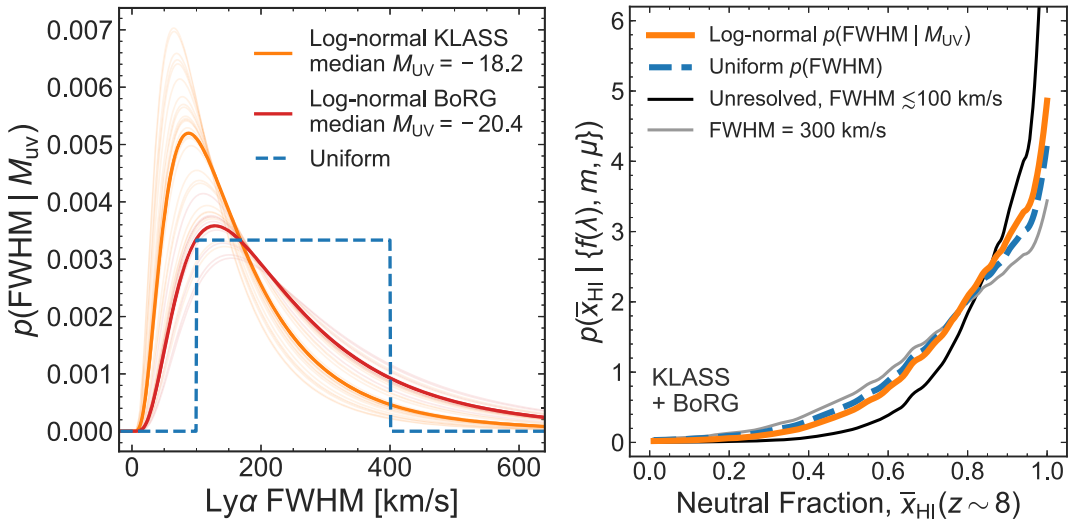

Figure C3. *Left:* Example prior probability distributions for Lyα FWHM used in our inferences. The orange curves shows the lognormal priors p(FWHM | M_{UV}) for the KLASS sample (median $M_{UV} = -18.2$ shown as thick orange line), while the red curves show the lognormal priors for the BoRG sample (prior for median $M_{UV} = -20.4$ is the thick red line). The blue dashed line shows a uniform prior (independent of M_{UV}) between 100 and 400 km s^{-1}. *Right:* Posterior probability distribution for the IGM neutral fraction derived from the KLASS and BORG samples using several different FWHM priors. The thin black line shows the posterior obtained assuming all lines are unresolved (FWHM $\lesssim$ 100 km s^{-1}), while the thin grey line is the posterior obtained assuming all emission lines are 300 km s^{-1}, demonstrating how the posterior flattens with increasing linewidth as the EW sensitivity decreases. For our UV faint targets we expect the lines to be relatively narrow, so the 300 km s^{-1} case is extreme. The orange solid line shows the posterior obtained using our lognormal FWHM prior p(FWHM | M_{UV}) calculated for each galaxy (described in Appendix C4), and the blue dashed line shows the posterior obtained using the uniform prior that is independent of M_{UV}. The difference in the 68 per cent limits obtained using these two posteriors is negligible: $\Delta\bar{x}_{HI} \sim 0.03$, Thus our results are robust for emission lines with widths ∼100−400 km s^{-1}.

(covering [z_{min}, z_{max}]) we would measure a non-detection in our data. Thus the integral over z_d becomes:

$$p(\bar{x}_{HI} \mid \{f\}, m, \mu) \propto \int_{z_{min}}^{z_{max}} dz_d \; p(\{f\}|\bar{x}_{HI}, m, \mu, z_d) p(z_d) + \prod_i p(\{f\} = 0) \left(1 - \int_{z_{min}}^{z_{max}} dz_d \; p(z_d)\right) \quad \text{(C9)}$$

We assume all galaxies observed are independent, so that the final posterior is the product of the normalized posteriors (equation C8) for each object.

C4 The impact of linewidth on the inference

The observed linewidth of Lyα emission will impact our inferences: broader lines will have lower significance and so reduce the strength of our inferences (see Fig. C1. As all of our observations are non-detections, we must be careful to account for this effect, which we do by marginalizing our posterior over FWHM using a empirically motivated FWHM prior.

In Fig. C2 we plot M_{UV} versus Lyα FWHM (deconvolved with instrumental resolution) for current detections of Lyα in $z > 7$ LB (Vanzella et al. 2011; Ono et al. 2012; Schenker et al. 2012; Shibuya et al. 2012; Finkelstein et al. 2013; Oesch et al. 2015; Zitrin et al. 2015; Roberts-Borsani et al. 2016; Song et al. 2016a; Laporte et al. 2017a; Stark et al. 2017). The observed FWHM span 20−450 km s^{-1}. However, linewidth measurements only exist for objects with $M_{UV} < -20$, significantly brighter than our KLASS sample, and there is a lot of scatter, motivating future observations to further explore the relationship.

Rather than fitting a relation to this limited current sample, we construct FWHM priors using other simple empirical relations. To extrapolate to lower UV luminosities, motivated by observed correlations between Lyα velocity offsets and the linewidths (Verhamme et al. 2018), we also plot the $\Delta v(M_{UV})$ model derived empirically by Mason et al. (2018a) (their equation 3). We then rescale this model using the relation between Lyα FWHM and Δv derived by (Verhamme et al. 2018): FWHM = $(34 + \Delta v)/0.9$. The Verhamme et al. (2018) relation is derived from a sample of $z \sim$ 0–8 Lyα emitters, and suggests that the Lyα photon scattering in ISM that produces this correlation between line offsets and linewidth is not entirely erased by the IGM.

We use two test FWHM priors: a uniform distribution between 100 and 400 km s^{-1}, spanning the currently observed range; and a lognormal distribution with mean given by the Mason et al. (2018a) $\Delta v(M_{UV})$ model scaled by the Verhamme et al. (2018) relation, with a 0.3 dex scatter. These priors are plotted in the left-hand panel of Fig. C3.

In the right-hand panel of Fig. C3 we show the posteriors for the IGM neutral fraction obtained from the full KLASS and BoRG samples by marginalizing over FWHM using these two priors, as well as the case of FWHM = 300 km s^{-1} and the case of unresolved lines. We see that except for the case of unresolved lines the difference between the posteriors is negligible. Thus our results are robust for lines with FWHM <400 km s^{-1}. We adopt the lognormal prior for our final inference as it is the most physically motivated.

If the lines are truly very narrow (<100 km s^{-1}), our constraint would be stronger. However, without high-resolution observations of $z > 5$ Lyα in UV faint galaxies to provide evidence of narrow linewidths, we have decided to be more conservative and use the lognormal prior in our final inferences.

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