Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **512,** 5390–5403 (2022) https://doi.org/10.1093/mnras/stac825
Advance Access publication 2022 March 25

IGM damping wing constraints on reionization from covariance reconstruction of two $z \gtrsim 7$ QSOs

Bradley Greig [1,2★] Andrei Mesinger [3] Frederick B. Davies,[4] Feige Wang,[5†] Jinyi Yang[5‡] and Joseph F. Hennawi [6,7]

[1]*School of Physics, University of Melbourne, Parkville, VIC 3010, Australia*
[2]*ARC Centre of Excellence for All-Sky Astrophysics in 3 Dimensions (ASTRO 3D) , Australia*
[3]*Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*
[4]*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*
[5]*Steward Observatory, University of Arizona, 933 N Cherry Ave, Tucson, AZ 85721, USA*
[6]*Department of Physics, Broida Hall, University of California, Santa Barbara, CA 93106-9530, USA*
[7]*Leiden Observatory, Leiden University, P.O. Box 9513, NL-2300 RA Leiden, the Netherlands*

Accepted 2022 March 18. Received 2022 February 22; in original form 2021 December 7

ABSTRACT

Bright, high-redshift ($z > 6$) QSOs are powerful probes of the ionization state of the intervening intergalactic medium (IGM). The detection of Ly α damping wing absorption imprinted in the spectrum of high-z QSOs can provide strong constraints on the epoch of reionization (EoR). In this work, we perform an independent Ly α damping wing analysis of two known $z > 7$ QSOs; DESJ0252−0503 at $z = 7.00$ (Wang et al.) and J1007+2115 at $z = 7.51$ (Yang et al.). For this, we utilize our existing Bayesian framework which simultaneously accounts for uncertainties in: (i) the intrinsic Ly α emission profile (reconstructed from a covariance matrix of measured emission lines; extended in this work to include N V) and (ii) the distribution of ionized (H II) regions within the IGM using a 1.6^3 Gpc^3 reionization simulation. This approach is complementary to that used in the aforementioned works as it focuses solely redward of Ly α ($1218 < \lambda < 1230$ Å) making it more robust to modelling uncertainties while also using a different methodology for (i) and (ii). We find, for an EoR morphology driven by galaxies within $M_{\rm h} \gtrsim 10^9$ M$_\odot$ haloes, $\bar{x}_{\rm H\,I} = 0.64^{+0.19}_{-0.23}$ (68 per cent) at $z = 7$ and $\bar{x}_{\rm H\,I} = 0.27^{+0.21}_{-0.17}$ at $z = 7.51$ consistent within 1σ to the previous works above, though both are slightly lower in amplitude. Following the inclusion of N V into our reconstruction pipeline, we perform a reanalysis of ULASJ1120+0641 at $z = 7.09$ (Mortlock et al.) and ULASJ1342+0928 at $z = 7.54$ (Bañados et al.) finding $\bar{x}_{\rm H\,I} = 0.44^{+0.23}_{-0.24}$ at $z = 7.09$ and $\bar{x}_{\rm H\,I} = 0.31^{+0.18}_{-0.19}$ at $z = 7.54$. Finally, we combine the QSO damping wing constraints for all four $z \gtrsim 7$ QSOs to obtain a single, unified constraint of $\bar{x}_{\rm H\,I} = 0.49^{+0.11}_{-0.11}$ at $z = 7.29$.

Key words: galaxies: high-redshift – intergalactic medium – cosmology: theory – dark ages, reionization, first stars – diffuse radiation – early Universe.

1 INTRODUCTION

Following the recombination of protons and electrons at the surface of last scattering, the Universe is permeated by a dense, cold fog of neutral hydrogen. This fog is only lifted when the cumulative ionizing radiation from stars, galaxies, and QSOs exceeds the number of neutral hydrogen atoms plus its recombinations rate. This baryonic phase change is referred to as the epoch of reionization (EoR). Reionization proceeds in a patchy manner: individual ionized (H II) regions around the first structures merge with their neighbours before percolating and eventually ionizing the entire intergalactic medium (IGM). The timing and duration of the EoR can loosely be inferred through integral measurements such as the Thomson scattering of cosmic microwave background (CMB) photons off the liberated electrons (measured as an optical depth, $\tau_{\rm e}$; e.g. Planck Collaboration 2020).

Recent observational evidence appears to point towards the tail-end of reionization occurring relatively late. For example, the observed 110 h^{-1} Mpc Gunn–Peterson down to $z = 5.5$ (Becker et al. 2015), the large scatter in the Ly α effective optical depth at $z \sim 5.5$ (Eilers, Davies & Hennawi 2018; Yang et al. 2020b; Bosman et al. 2021a) and the Bosman et al. (2018) Ly α forest data that when interpreted in a forward-modelled Bayesian framework implies reionization completed at $z \lesssim 5.6$ (Qin et al. 2021, see also Kulkarni et al. 2019; Nasir & D'Aloisio 2020; Keating et al. 2020a,b; Choudhury, Paranjape & Bosman 2021).

Direct constraints on reionization via the absorption of Ly α photons by lingering neutral regions at $z \gtrsim 5$ are problematic due to the amplitude of the scattering cross-section of Ly α photons in the resonant core. Here, the IGM is in photoionization equilibrium whereby the density of the residual neutral hydrogen

★ E-mail: greigb@unimelb.edu.au
† NHFP Hubble Fellow.
‡ Strittmatter Fellow.

($x_{\rm H\,I} \gtrsim 10^{-4}$–$10^{-5}$) is sufficient to saturate transmission (Fan et al. 2006). Thus order unity fluctuations in the neutral fraction during reionization are difficult to distinguish from fluctuations in the ultraviolet background, density, or temperature, post reionization. Alternatively, a more robust probe of the IGM neutral fraction is available through the Ly α damping wing (e.g. Rybicki & Lightman 1979; Miralda-Escudé 1998). The absorption cross-section of the extended Lorentzian wings in the Ly α line profile are several orders of magnitude below that of the line centre, making them ideal for studying the order unity fluctuations in neutral fraction during reionization.

Constraints on the IGM neutral fraction from the damping wing have been successfully employed using both galaxies and QSOs. For galaxies, the imprint is only measurable after averaging over large statistical samples (e.g. Mesinger et al. 2015; Mason et al. 2018; Hoag et al. 2019; Mason et al. 2019). QSOs on the other hand, being orders of magnitude brighter, enable the damping wing to be detectable from a single spectrum (e.g. Mesinger & Haiman 2007; Bolton et al. 2011; Mortlock et al. 2011; Bosman & Becker 2015; Greig et al. 2017b; Bañados et al. 2018; Davies et al. 2018a; Greig, Mesinger & Bañados 2019; Ďurovčíková et al. 2020; Wang et al. 2020; Yang et al. 2020a). However, care must be taken to properly characterize the individual QSO intrinsic emission, together with the corresponding uncertainties.

There are several approaches in the literature specifically tailored towards reconstructing the intrinsic QSO emission profile. One of the most well-studied approaches is through principal component analysis (PCA). Here, spectral features of the QSO continuum are deconstructed into principal component vectors. These principal components are separated redward and blueward of $\lambda = 1280$ Å with either a linear (Davies et al. 2018b; Bosman et al. 2021b) or neural network based (Ďurovčíková et al. 2020) mapping used to reconstruct the blue-side from the measured red-side components. These mappings are learned from ∼13 000 moderate-z QSOs. Alternatively, both Fathivavsari (2020) and Liu & Bordoloi (2021) have employed deep neural networks (i.e. deep learning). In the former, a network was trained on ∼18 000 high signal to noise (S/N) QSOs using spectral information from the C IV, Si IV, and C III] broad emission lines to predict Ly α. In the latter, the network was trained on high-resolution *Hubble Space Telescope* spectra at $z \sim 0.2$ to predict the entire 1020 Å; $\leq \lambda \leq$ 1600 Å; range from spectral information provided between 1216 Å; $\leq \lambda \leq$ 1600 Å;. Finally, Reiman et al. (2020) employ conditional neural spline flows to provide a fully probabilistic reconstruction of the 1190 Å; $\leq \lambda \leq$ 1290 Å; region from redward information over 1290 Å; $\leq \lambda \leq$ 2900 Å.

In this work, we use the Bayesian reconstruction pipeline developed in Greig et al. (2017a). Here, the Ly α profile is reconstructed using a covariance matrix of correlations between the Ly α line and other high ionization emission lines (e.g. C IV, Si IV, and C III]). This assumes that all emission line profiles are modelled by a single, or two component Gaussian profile characterized entirely by its width, height, and velocity offset. Further, that the correlations themselves are fully encapsulated by an N-dimensional normal distribution. This Gaussian assumption appears justified based on its performance relative to our training set data.

Importantly, our full damping wing analysis pipeline combines the reconstructed profile with a distribution of synthetic IGM damping wings extracted from a large-scale EoR simulation (Mesinger, Greig & Sobacchi 2016). This step accounts for the statistical uncertainties that arise from the fact we are only considering single sightlines (QSO spectra) to place constraints on the EoR. A similar approach is adopted by Davies et al. (2018a) whose method is applied in Wang et al. (2020) and Yang et al. (2020a). However, unlike these works we consider only the damping wing imprint redward of Ly α, avoiding the uncertainties associated with modelling the QSO host environment. For all other works, damping wing constraints are extracted using the simplistic Miralda-Escudé (1998) damping wing model.

Specifically, we apply our analysis framework to the two recently discovered $z \gtrsim 7$ QSOs: DESJ0252−0503 at $z = 7.00$ (hereafter J0252; Yang et al. 2019; Wang et al. 2020) and J1007 + 2115 at $z = 7.51$ (hereafter J1007; Yang et al. 2020a). Our analysis (discussed in detail below), notably differs from the previous ones, thus this work serves as an independent and complementary verification of the inferred EoR constraints. For reference, these authors recover constraints on the IGM neutral fraction of $\bar{x}_{\rm H\,I} = 0.70^{+0.20}_{-0.23}$ at $z = 7.0$ from J0252 and $\bar{x}_{\rm H\,I} = 0.39^{+0.22}_{-0.13}$ at $z = 7.51$ from J1007 (both at 68 per cent confidence).

This work is structure as follows. In Section 2, we provide a brief description of the observational data, before outlining our analysis pipeline. In Section 3, we provide the main results and in Section 4 our discussion. In Section 5, we conclude with our closing remarks. Unless stated otherwise, we quote all quantities in co-moving units and adopt the cosmological parameters: (Ω_Λ, $\Omega_{\rm M}$, $\Omega_{\rm b}$, n, σ_8, H_0) = (0.69, 0.31, 0.048, 0.97, 0.81, 68 km s^{-1} Mpc^{-1}), consistent with recent results from the Planck mission (Planck Collaboration XIII 2016).

2 METHOD

In Section 2.1, we first provide a brief summary of the observational data for J0252 and J1007 used in this work. In Section 2.2, we summarize the covariance matrix reconstruction pipeline introduced by Greig et al. (2017a) including the modifications for this work and in Section 2.3 we outline the simulation procedure for extracting the synthetic damping wing profiles. Finally, in Section 2.4 we summarize the joint fitting procedure used to obtain constraints on the IGM neutral fraction.

2.1 Observational data

2.1.1 J1007+2115

In this work, we make use of the high-quality combined spectrum outlined in Yang et al. (2020a). This combines a 5.5 h (on-source) Gemini Near-Infrared Spectrograph (GNIRS) observation ($R \sim 620$) with a 2.2 h (on-source) observation with Keck/Near-Infrared Echellette Spectromenter (NIRES; $R \sim 2700$) using inverse-variance weighting. The systemic redshift for this QSO, $z = 7.5149 \pm 0.0004$, is obtained from [C II] emission using the Atacama Large Millimeter/submillimeter Array (ALMA).

2.1.2 DESJ0252−0503

For J0252, we use the combined spectrum outlined in Wang et al. (2020). This combines a high quality 4.8 h on-source near-infrared observation ($R \sim 2700$) from Keck/NIRES with earlier Gemini Multi-Object Spectrogram observations (combined $R \sim 1300$). The systemic redshift for this QSO, $z = 7.000 \pm 0.001$, is obtained from [C II] emission using the IRAM Northern Extended Millimeter Array (NOEMA).

2.2 Covariance reconstruction of the intrinsic Ly α profile

Previously, in Greig et al. (2017a) we developed a method to reconstruct the intrinsic Ly α line profile using a measured covariance matrix of correlations amongst emission line parameters describing Ly α and other readily measurable high ionization lines (namely, C IV, Si IV + O IV] and C III]). This covariance matrix was obtained from a carefully cultivated training set[1] of 1673 moderate-z ($2.08 < z < 2.5$), high signal to noise (S/N >15) QSOs from SDSS-III (BOSS) DR12 (Dawson et al. 2013; Alam et al. 2015).

Each line profile was modelled as a Gaussian, fully described by its peak height, width, and velocity offset from systemic. For Ly α and C IV we found a strong preference for a two-component Gaussian, containing both a broad and narrow line profile. In total, this amounts to an 18×18 covariance matrix describing all emission line correlations.

Importantly, in this work, we extend this model to additionally include the correlations amongst the N V line, resulting in an expanded 21×21 covariance matrix of emission line parameters. Previously, we neglected the N V line as it did not strongly correlate with Ly α and because the rest-frame line centre ($\lambda_{\rm N\,V} = 1240.81$ Å) fell outside the region over which we fit for the damping wing imprint ($\lambda = 1218$–1230 Å). However, the N V line profile can be blueshifted closer to this region while also be being sufficiently broad to contribute flux into this wavelength window. Thus, in this work we include N V to account for this. In Appendix A, we present our updated correlation matrix including N V, while in Appendix B we reanalyse the previous $z \gtrsim 7$ QSOs for which we have performed a damping wing analysis for, following the inclusion of N V into our updated reconstruction pipeline.

Following the extension of our covariance matrix to include N V, our reconstruction of the intrinsic Ly α line profile proceeds as follows:

(i) We fit the observed rest-frame QSO spectrum over the range $\lambda = 1275$–2300 Å using the systemic redshifts[2] outlined in Section 2.1.

(ii) We perform a joint fit of a power-law continuum component plus the Gaussian profiles for the high-ionization lines Si IV + O IV], C IV (2-component), and C III] emission line profiles. To improve the overall fit, we also fit for other known emission line profiles as well as simultaneously fitting a variable number of absorption lines each modelled as a single Gaussian profile. This is shown for J1007 and J0252 in Figs C1 and C2, respectively.

(iii) Using the fits to the C IV, Si IV + O IV] and C III] emission line components (amplitude, width, and velocity offset), we collapse the 21D (Gaussian distributed) covariance matrix into a 9D estimate of the joint Ly α and N V intrinsic emission line profile (two Gaussian components for Ly α and one for N V).

(iv) With this 9D covariance matrix, we then draw joint intrinsic Ly α and N V profiles from this distribution. We additionally apply a flux prior over the range, $1250 < \lambda < 1275$ Å, to ensure our reconstructed profiles match the flux level of the observed spectrum over this range.[3]

2.3 Modelling the IGM damping wing during reionization

In order to quantify a putative imprint of the neutral IGM in the QSO spectrum we must create synthetic IGM damping wing profiles. Specifically, our synthetic profiles are computed using the Evolution of 21-cm Structure (EOS; Mesinger et al. 2016)[4] 2016 simulations. These are large volume (side length of 1.6 Gpc over 1024 voxels) seminumerical reionization simulations that include state-of-the-art sub-grid prescriptions for inhomogeneous recombinations and photoheating suppression of star formation. In this simulation suite there are three different EoR morphologies, characterized by different star formation efficiencies for the typical haloes hosting star-forming galaxies.

In Greig et al. (2019), we explored the impact of these different EoR morphologies on the recovered constraints on the IGM neutral fraction from the damping wing, finding only weak evidence for a morphology dependence.[5] As such, in this work, we shall only consider one EoR morphology, the INTERMEDIATE H II model, which is the middle of the three EoR simulations. This model is characterized by an EoR driven by galaxies residing in $M_{\rm h} \gtrsim 10^9\,{\rm M}_\odot$ haloes, consistent with recent results from forward-modelling the Ly α forest opacity distributions and ultraviolet (UV) galaxy luminosity functions (LFs; Qin et al. 2021).

Out of this simulation we extract a total of 10^5 synthetic IGM damping wing profiles. For the 10^4 largest identified haloes (corresponding to roughly $6 \times 10^{11} < M_{\rm h} < 3 \times 10^{12}\,{\rm M}_\odot$) we draw 10 randomly oriented sightlines originating from the central host halo. When constructing the synthetic damping wing profiles, we exclude the cumulative contribution from all encountered H I patches inside the measured QSO near-zone for either object. Implicitly, we are assuming that most of the neutral gas inside the proximity zone (at least down to Lyman-limit systems) will have been ionized by the host QSO. This is motivated by the fact we do not observe any evidence for high-column density systems within the proximity zone. For J0252, the estimated near-zone size is ∼13.6 comoving Mpc (1.7 physical Mpc) while for J1007 it is ∼12.6 comoving Mpc (1.5 physical Mpc). Then, to have these synthetic damping wing profiles vary with IGM neutral fraction, we sample the ionization fields obtained from different redshift snapshots from the simulation.[6] In total, we have 24 unique values of the IGM neutral fraction.

2.4 Joint fitting to obtain IGM neutral fraction constraints

Having outlined the individual procedures to obtain both the reconstructed intrinsic Ly α profile and the synthetic damping wing profiles, we now summarize the joint pipeline used to extract

[1] QSOs were removed in cases where the Ly α line profile could not be well characterized by our fitting procedure, most commonly caused by excessive absorption at or near Ly α, see appendix C in Greig et al. (2017a) for further details.

[2] Note, our training set of BOSS QSOs uses the SDSS-III pipeline redshift to convert to rest frame rather than a [C II] redshift as used for the observed $z \gtrsim 7$ QSOs in this work. While this can lead to biases in the recovered line profile blueshifts, this is likely sub-dominant relative to the scatter in Ly α reconstruction profiles and synthetic damping wings following our analysis pipeline, see Greig et al. (2019) for further details.

[3] In practice, this flux prior will only be relevant if the observed spectrum exhibits a large/broad N V emission line profile extending redward of 1250 Å, where its shape information could be used to further constrain the reconstructed N V emission profile.

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

[5] Quantitatively, the constraints on the IGM neutral fraction owing to different EoR morphologies was found to be ±0.05 relative to the INTERMEDIATE H II model. In contrast, the typical errors on our damping wing constraints are ±0.15–0.20.

[6] This assumes that the halo locations do not significantly change due to bulk motions across the different redshift snapshots. Relative to the resolution of the simulation (cell size ∼1.6 Mpc) this is a reasonable assumption.

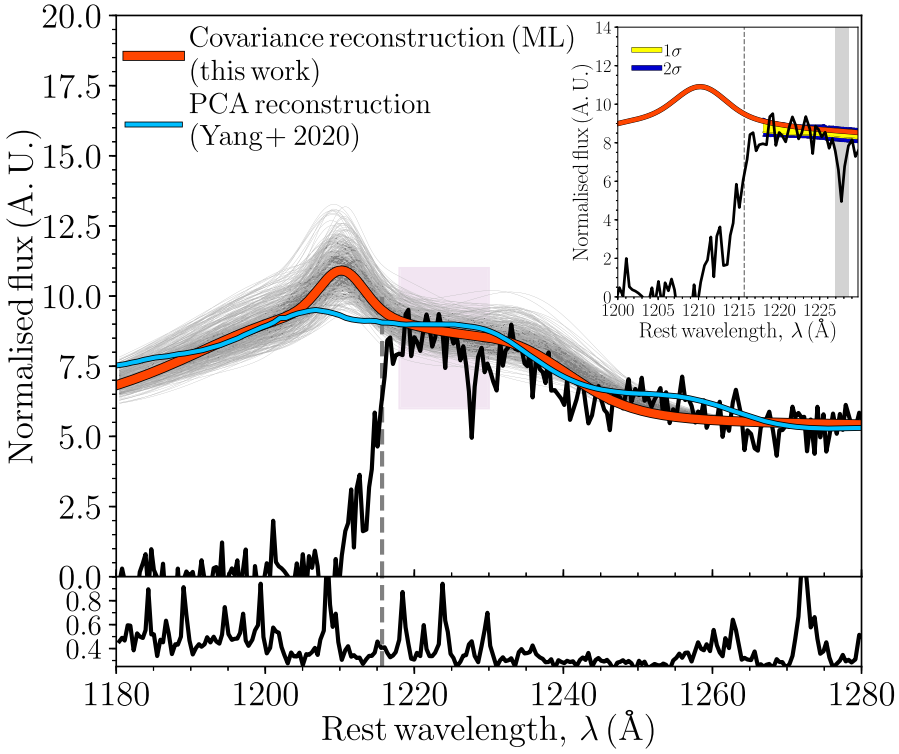

Figure 1. The reconstructed maximum likelihood Ly α emission line profile for J1007 at $z = 7.51$ (red curve; shown for visualization purposes only) and a subsample of 300 Ly α line profiles (thin grey curves) randomly drawn from the full posterior distribution of reconstructed profiles (see Section 2.2; i.e. the analysis pipeline samples the full posterior rather than just the maximum likelihood profile). The black curve (and associated error below) corresponds to the combined GNIRS and NIRES spectrum and the blue line corresponds to the PCA reconstructed profile from Yang et al. (2020a) and the grey dashed line denotes rest-frame Ly α. The purple shaded region corresponds to the region over which we fit for the IGM damping wing imprint, corresponding to the range $\lambda = 1218$–1230 Å.

constraints on the IGM neutral fraction from the damping wing imprint. In summary the steps are as follows:

(i) We draw $\sim 10^5$ estimates of the intrinsic QSO profile from the combined reconstruction of the Ly α and N V line profiles as described in Section 2.2.

(ii) For each intrinsic profile, we multiply it by the 10^5 synthetic damping wing opacities obtained from Section 2.3. This results in $\sim 10^{10}$ mock spectra for each $\bar{x}_{\rm H\,I}$ snapshot from the EoR simulation.

(iii) Each of these $\sim 10^{10}$ mock spectra are then compared to the observed QSO spectrum over 1218 Å $< \lambda <$ 1230 Å (consistent with Greig et al. 2017b, 2019). For each mock spectra we calculate its χ^2 relative to observed flux and the error spectrum. Additionally, where appropriate, we mask out any features (e.g. absorption) over the 1218 Å $< \lambda <$ 1230 Å range that may erroneously bias the resultant fit.

(iv) The resulting likelihood, averaged (i.e. marginalized) over all $\sim 10^{10}$ mock spectra, is then assigned to the $\bar{x}_{\rm H\,I}$ of the current simulation snapshot.

(v) We then repeat steps (ii)–(iv) for all available $\bar{x}_{\rm H\,I}$ snapshots (of which there are 24) to obtain a final 1D probability distribution function (PDF) of $\bar{x}_{\rm H\,I}$ for the observed QSO.

3 RESULTS

3.1 Reconstruction of J1007

In Fig. 1, we provide our reconstructed intrinsic QSO profile for J1007. The best-fitting (maximum likelihood; ML) reconstruction profile is given by the red curve while the thin grey curves highlight a random subsample of 300 reconstruction profiles drawn from the full posterior distribution. In comparison, the blue curve corresponds to the PCA reconstructed profile from Yang et al. (2020a) using the method outlined in Davies et al. (2018b). Note, our reconstruction pipeline does not predict Si II ($\lambda \sim 1263$), which is included in the PCA pipeline, explaining the different flux levels observed just blueward of 1260 Å. Additionally, for reference in Fig. C1, we provide the MCMC fit of J1007 that is used to obtain the high ionization line parameters used in our reconstruction pipeline.

Qualitatively speaking, from Fig. 1 we recover noticeably different reconstruction profiles across the two different pipelines; covariance matrix in this work and the PCA approach from Davies et al. (2018b). We find a clear preference for a narrow component, peaked blueward of Ly α at $\lambda \sim 1210$ Å along with a considerable broad component. In the PCA approach, no clear peak in Ly α is found, though there is a small peak closer to $\lambda \sim 1205$ Å. Both pipelines however do recover similarly broad N V line components. Interestingly though, the velocity offset of the small amplitude Ly α peak and the N V peak from the PCA approach are similarly blueshifted relative to the Ly α and N V peaks extracted from our covariance matrix approach indicating that the correlations in the velocity offsets between Ly α and N V are consistent between the two pipelines. The similarities and differences between the two pipelines (covariance matrix and PCA) have been discussed in detail in Greig et al. (2019), however we will return to this discussion in Section 4. Nevertheless, despite the different observable profile features, statistically speak-

ing the two different reconstruction methods are consistent with one another when comparing the full distribution of reconstructed profiles.[7]

Across the region for which we fit the damping wing (1218 Å $< \lambda <$ 1230 Å), the amplitudes of the reconstructed profiles from the two pipelines are relatively similar, with our covariance matrix approach preferring a slightly lower amplitude flux. However, our reconstructed narrow Ly α line profile does enter at the edge of our fitting region (near $\lambda \sim$ 1218 Å), where our reconstructed profile becomes larger in amplitude.

In the inset of Fig. 1, we present the confidence intervals following our joint fitting procedure (Section 2.4) with the yellow (blue) shaded regions corresponding to the 68 (95) percentiles. The vertical grey band demarcates a notable absorption feature which we mask from our damping wing fitting interval. As the ML (red) curve is only marginally offset (larger) than the yellow/blue regions this suggests that there is evidence of only a weak IGM damping wing imprint. That is, the intrinsic Ly α line profile does not require significant attenuation by a neutral (or partially) neutral IGM to match the observed QSO spectrum.

3.2 Reconstruction of J0252

Next, in Fig. 2, we provide our reconstructed intrinsic QSO profile for J0252. Again, the best-fitting (ML) reconstruction profile is given by the red curve while the thin grey curves highlight a random subsample of 300 reconstruction profiles drawn from the full posterior distribution. In comparison, the blue curve corresponds to the PCA reconstructed profile from Wang et al. (2020) using the method outlined in Davies et al. (2018b) while the orange curve is a composite spectrum constructed from 83 low-z (SDSS/BOSS) QSOs selected for their large C IV blueshift (3000–5000 km s^{-1}), analogous to J0252.[8] Again, in Fig. C2, we provide the full MCMC fit of J0252 used in our reconstruction pipeline.

Once again, qualitatively speaking, we observe notably different reconstructed profiles from the two approaches. Again, our covariance matrix approach prefers the existence of a strong narrow Ly α emission line component, peaked around $\lambda \sim$ 1210 Å. Unlike for J1007, we do not find any preference for any notable N V line component. This differs to the PCA reconstruction in Wang et al. (2020), where significant emission is observed between $\lambda \sim$ 1220 and 1235 Å. This could be either Ly α and N V or simply a large amplitude, though strongly blueshifted, N V component. It is harder to ascertain the origins of these features from PCA approaches as they do not predict single peaked emission features. As mentioned earlier, we will discuss differences in the two pipelines in Section 4. Importantly though, as was the case for J1007 previously, within the error bounds of both pipelines the recovered profiles appear to be statistically consistent.

Across the region for which we fit the damping wing (1218 Å $< \lambda <$ 1230 Å), the shape and amplitude of the reconstructed profiles differ considerably. Our reconstructed profiles are systematically lower than those from the PCA reconstruction of Wang et al. (2020). In the inset of Fig. 2, we again show the confidence intervals following our joint fitting procedure (Section 2.4) with the yellow (blue) shaded regions corresponding to the 68 (95) percentiles. Here, the ML (red) curve notably deviates from these shaded regions indicative clear evidence of notable attenuation of the intrinsic flux by a neutral IGM.

3.3 Recovered IGM damping wing constraints

In the left-hand panel of Fig. 3, we present the 1D PDFs of our IGM neutral fraction constraints from our damping wing fits to J1007 (blue curve) and J0252 (red curve) using the INTERMEDIATE H II EoR morphology. For comparison, we additionally show the 1D PDFs recovered by the PCA approach of Davies et al. (2018b) presented by Yang et al. (2020a) (blue dashed curve) and Wang et al. (2020) (red dashed curve).

Quantitatively, we recover IGM neutral fraction constraints (with 68 per cent confidence limits) of:

(i) $\bar{x}_{\rm H\,I} = 0.27^{+0.21}_{-0.17}$ at $z = 7.51$ for J1007
(ii) $\bar{x}_{\rm H\,I} = 0.64^{+0.19}_{-0.23}$ at $z = 7.00$ for J0252.

In comparison, for J1007 Yang et al. (2020a) found $\bar{x}_{\rm H\,I} = 0.39^{+0.22}_{-0.13}$ and for J0252 Wang et al. (2020) recovered $\bar{x}_{\rm H\,I} = 0.70^{+0.20}_{-0.23}$. For both QSOs, the results in this work recover neutral fraction systematically lower than those recovered from the PCA approach. However, these differences are relatively minor, given the relatively large confidence limits. Quantitatively, our results are $\lesssim 0.5\sigma$ below the previously reported constraints. In the next section, we will discuss in detail the similarities and differences between the two approaches.

Finally, we also update our constraints on the IGM neutral fraction from the two previously known $z \gtrsim 7$ QSOs following the extension of our covariance matrix reconstruction pipeline to include the N V emission line. Specifically, these are ULASJ1120+0641 (hereafter J1120; Mortlock et al. 2011) at $z = 7.08$ and ULASJ1342+0928 (hereafter J1342; Bañados et al. 2018) at $z = 7.54$. Previously, using the same INTERMEDIATE H II EoR morphology, we recovered $\bar{x}_{\rm H\,I} = 0.40^{+0.21}_{-0.19}$ for J1120 and $\bar{x}_{\rm H\,I} = 0.21^{+0.17}_{-0.19}$ for J1342.

In the right-hand panel of Fig. 3, we present a compilation of the IGM neutral fraction constraints for of all four known $z \gtrsim 7$ QSOs using our covariance matrix approach (with N V) assuming the INTERMEDIATE H II EoR morphology. The red and blue curves correspond to J0252 and J1007 as per the right-hand panel of Fig. 3, whereas the black dotted and dashed curves correspond to the new constraints on the IGM neutral fraction from J1120 and J1342, respectively. Quantitatively, following the inclusion of N V we now update our constraints to the following:

(i) $\bar{x}_{\rm H\,I} = 0.44^{+0.23}_{-0.24}$ at $z = 7.08$ for J1120
(ii) $\bar{x}_{\rm H\,I} = 0.31^{+0.18}_{-0.19}$ at $z = 7.54$ for J1342.

For both, we find a higher IGM neutral fraction owing to an overall increase in the predicted intrinsic flux following the inclusion of the N V line. Further, we also note an increase to the 68 per cent confidence intervals, owing to the increased scatter in the reconstructed profiles going from a 6D covariance matrix for the two component of Ly α (broad and narrow) to the new, 9D covariance matrix jointly reconstructing N V. In Appendix B, we

[7] A more rigorous quantitative analysis between the two approaches is beyond the scope of this paper. While for both methods it is trivial to produce uncertainties for the reconstruction profiles as a function of wavelength, this is not an accurate comparison as the reconstructed profiles themselves are correlated across wavelength bins. Thus, what is actually required is a comparison between the covariance matrices of reconstructed profiles from the two pipelines. As such, we defer such a comparison to future work.

[8] Note that our training sample does not contain QSOs with such extreme C IV blueshifts. In appendix A of Greig et al. (2017b) we explored the reconstruction profiles of the most extremely blueshifted QSOs within our sample, finding no clear degradation in the quality in the reconstructed profiles. Thus, we assume it is safe to extrapolate our model to these more extreme C IV blueshifts.

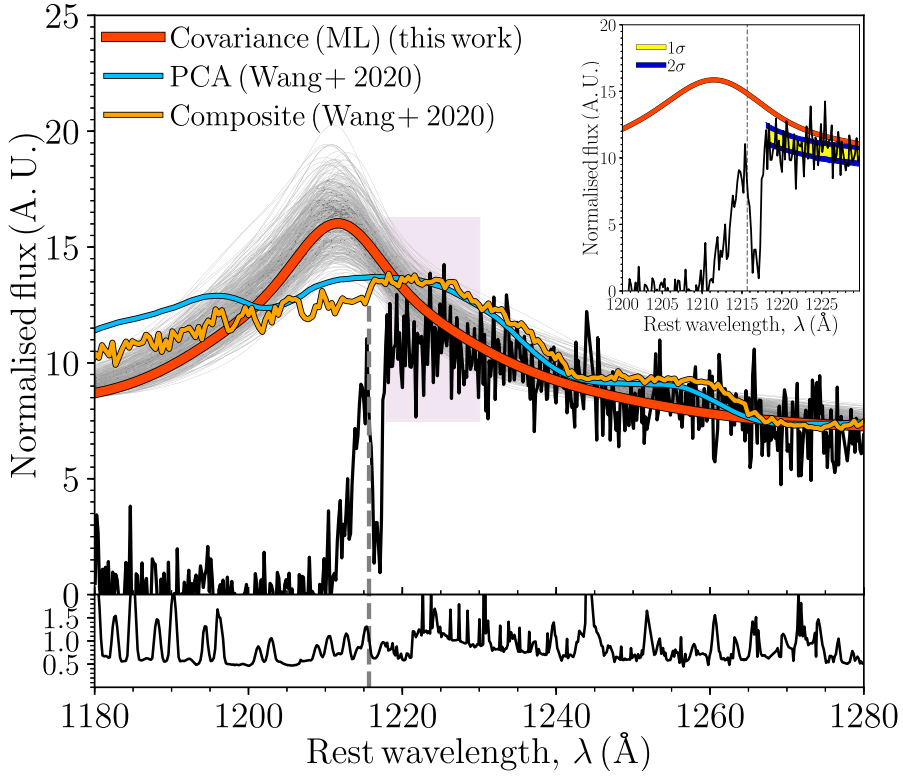

Figure 2. The same as Fig. 1, except for J0252 at $z = 7.0$. The black curve (and associated error below) corresponds to the Gemini/GMOS + Keck/NIRES spectrum, the blue line corresponds to the PCA reconstructed profile and the orange line corresponds to a composite spectrum of 83 low-z analogue QSOs (Wang et al. 2020) with large C IV blueshifts (3000–5000 km s^{-1}) comparable to J0252.

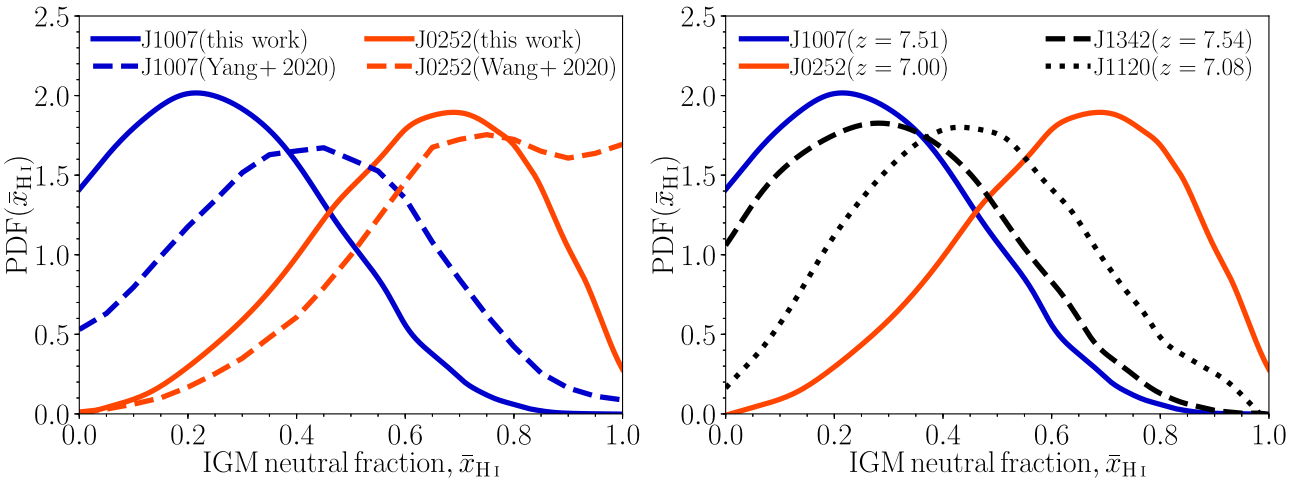

Figure 3. The marginalized 1D PDFs of the IGM neutral fraction. *Left-hand panel*: Comparison of the recovered constraints on the IGM neutral fraction for J0252 at $z = 7.00$ (red curve) and J1007 at $z = 7.51$ (blue curve). Solid curves denote constraints from this work (covariance matrix reconstruction) whereas dashed curves correspond to previous results using PCA reconstruction, Wang et al. (2020) and Yang et al. (2020a). *Right-hand panel*: Compilation of the IGM neutral fraction constraints from all $z \gtrsim 7$ QSOs using covariance matrix reconstruction (including N V) and the INTERMEDIATE H II EoR morphology.

provide the updated reconstructed profiles for J1120 (Fig. B1) and J1342 (Fig. B2) along with a discussion of the new profiles following the inclusion of N V into our analysis pipeline.

3.4 Compilation of reionization constraints

We now have IGM damping wing constraints on four $z \gtrsim 7$ QSOs obtained from two distinctly different reconstruction methods and damping wing analyses, as discussed in earlier sections. Here, we combine all of these constraints to obtain one, unified constraint on the IGM neutral fraction from the IGM damping wing imprint.[9] For this, we first average the two individual neutral

[9]Note here that we specifically focus on damping wing analyses that consider an inhomogeneous IGM. That is we do not include the constraints on J1120

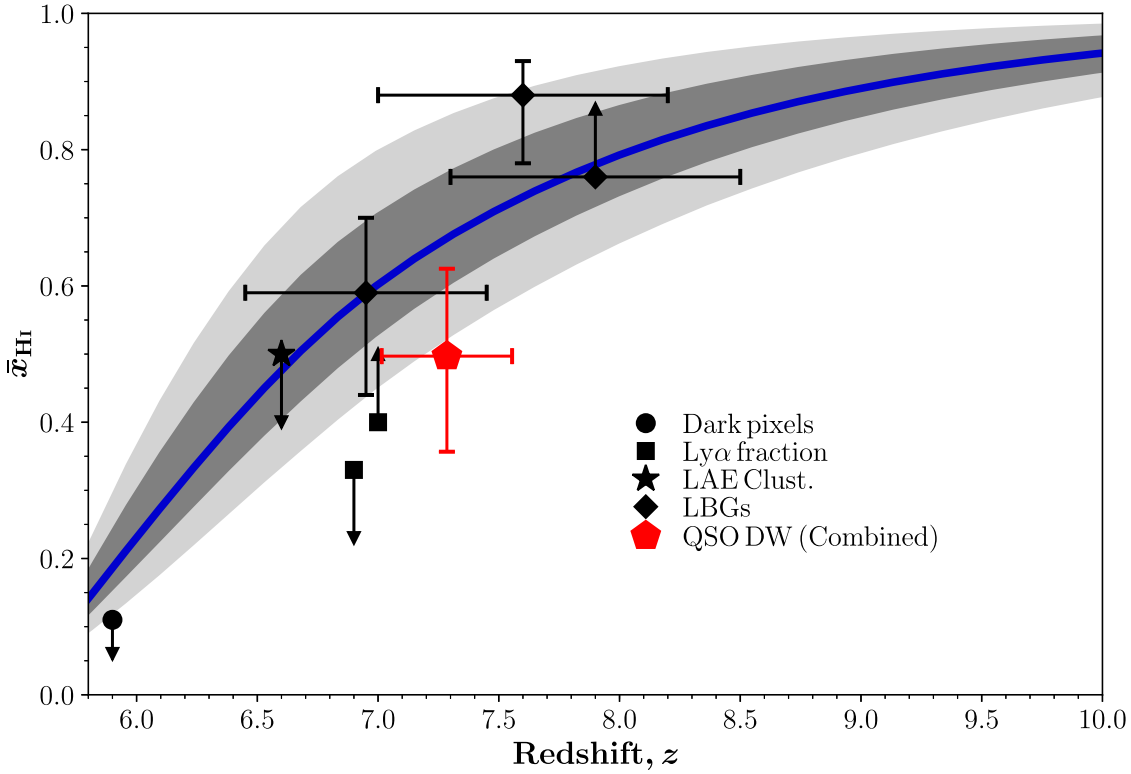

Figure 4. A compilation of existing constraints on the IGM neutral fraction as a function of redshift. *Circles*: Dark pixels at $z = 5.9$ (McGreer et al. 2015), *Squares*: the Ly α fraction at $z = 6.9$ (Wold et al. 2022) and $z = 7$ (Mesinger et al. 2015), *Stars*: LAE clustering at $z = 6.6$ (Sobacchi & Mesinger 2015), *Diamonds*: LBGs at $z = 7$ (Mason et al. 2018), $z = 7.6$ (Hoag et al. 2019), and $z = 8$ (Mason et al. 2019). The red pentagon corresponds to the combined constraints of all four $z \sim 7$ QSOs considered in this work. The blue curve and the dark and light shaded regions corresponds to the median, 1σ and 2σ constraints from forward-modelling: (i) large-scale Ly α forest opacity fluctuations; (ii) the CMB optical depth, τ_e; (iii) the fraction of dark pixels in the Ly α and Ly β forests and (iv) observed UV LFs at $z = 6$–10 (Qin et al. 2021).

fraction PDFs (corresponding to the two separate reconstruction pipelines; covariance matrix Greig et al. 2017a and PCA Davies et al. 2018b) for each of the four QSOs. In doing so we in effect average out the inherent systematics from the two different approaches. We then treat these four averaged results as independent measurements on the IGM neutral fraction, allowing us to multiply each individual likelihood to obtain a single, joint posterior for the IGM neutral fraction. Following this procedure, we obtain:

(i) $\bar{x}_{\rm HI} = 0.49^{+0.13}_{-0.14}$ at $z = 7.29 \pm 0.27$.

Note, in collapsing these constraints into a single data point we are in effect conservatively averaging over all modelling differences and systematics. Further, as these four QSOs span a redshift range of $\Delta z \sim 0.5$, cosmic evolution across all these QSOs should be fairly modest.

In Fig. 4, we place this unified QSO damping wing constraint (red pentagon) in context with other constraints on the IGM neutral fraction during reionization. Here, we consider constraints and limits obtained from: (i) dark pixels (McGreer, Mesinger & D'Odorico 2015), (ii) Ly α fraction at $z = 6.9$ (Wold et al. 2022) and at $z = 7$ (Mesinger et al. 2015), (iii) the clustering of Ly α emitters (LAEs) at $z = 6.6$ (Sobacchi & Mesinger 2015) and (iv) Lyman-break galaxies (LBGs) at $z = 7$ (Mason et al. 2018), $z = 7.6$ (Hoag et al. 2019) and at $z = 8$ (Mason et al. 2019). Additionally, we provide constraints on the reionization history obtained from a Monte Carlo Markov Chain (MCMC) analysis of the simulated 21-cm signal constrained by existing observations of the reionization epoch (Qin et al. 2021). Specifically, these models are constrained by observed UV galaxy LFs at $z = 6$–10, the electron scattering optical depth, τ_e, measured by Planck (Planck Collaboration et al. 2020), the dark pixel limits on the IGM neutral fraction (McGreer et al. 2015) and PDFs of the Ly α effective optical depth from the Ly α forest at $z = 5$–6 (Bosman et al. 2018). The median reionization history is represented by the blue line, whereas the dark and light grey shaded regions correspond to the 1σ and 2σ confidence intervals.

This unified QSO data point implies a mid-point of reionization at $z \sim 7.3$, slightly below similar limits and constraints from LBGs. However, within the appreciable 1σ uncertainties they are consistent. With respect to the observationally constrained reionization histories extracted by forward modelling Ly α forest data in a Bayesian framework by Qin et al. (2021), the median QSO damping wing constraint is 2–3σ below the median reionization history, however, again owing to the relatively large uncertainties in averaging across all QSOs, it is still consistent within error. This lower amplitude constraint from the combined QSO damping wing is driven by both $z \sim 7.5$ QSOs, which all individually sit below these 2σ reionization histories. Interestingly, the posteriors from Qin et al. (2021) did not consider any of the QSO data points shown in Fig. 4 (except for the upper limits from dark pixels), thus it is reassuring that these constraints are consistent.

or J1342 from Bañados et al. (2018), Ďurovčíková et al. (2020), or Reiman et al. (2020).

4 DISCUSSION

In this work, we have updated the covariance matrix reconstruction pipeline originally outlined in Greig et al. (2017a) to include a N v emission line component and to extract independent constraints on the IGM neutral fraction using our own methodology (Greig et al. 2017b, 2019). This serves as a complimentary analysis to that performed in Yang et al. (2020a) for J1007 and Wang et al. (2020) for J0252 using the PCA-based reconstruction and damping wing fitting method of Davies et al. (2018a,b). Although we have recovered consistent constraints on the IGM neutral fraction between the two methods, these have been achieved following various different assumptions for which we outline below.

At their core, the covariance matrix reconstruction pipeline (Greig et al. 2017a) used in this work and the PCA approach from Davies et al. (2018b) are very similar. Both, are based on similar[10] training sets of BOSS QSOs spanning $z \sim 2.1$–2.5 and use learnt correlations in the properties of the QSO spectrum redward of Ly α to reconstruct the intrinsic Ly α profile. In Greig et al. (2017a), it is assumed that all the correlations between the measured Gaussian emission line parameters can be fully described by its covariance matrix (i.e. that it is normally distributed). In Davies et al. (2018b), the correlations are instead measured amongst the decomposed principal components of the observed spectra. In utilizing principal components, this reduces the dimensionality of the reconstruction pipeline. For example, in this work we have a 21D covariance matrix that collapses into a 9D matrix in order to perform the reconstruction, whereas in Davies et al. (2018b) correlations are linearized between 10 principal components on the red-side and 6 on the blue-side. Add to this the fact that our training set is an order of magnitude smaller to that of Davies et al. (2018b) and that not all emission line parameters strongly covary amongst one another, we expect the uncertainties (scatter) in reconstructed profiles to be larger in our reconstruction pipeline.

Further, this choice to reconstruct on a series of emission line parameters defined by Gaussian profiles or principal components results in notably different shapes for the predicted intrinsic profiles. Our method, based on the three Gaussian emission line components (two for Ly α and one for N v) will have a relatively simple and smooth shape. On the other hand, for the PCA approach the shape is not tied to predicting individual components, rather it is the weighted sum of the several predicted blueward principal components. As such the predicted profile following the PCA approach can result in additional features not reproducible by our Gaussian profiles. Thus the overall shape and amplitude of the two pipelines will always differ making direct comparisons less trivial. This difference additionally extends into the distribution of posteriors drawn around the maximum likelihood profile across the two pipelines. While the qualitative trends will be similar across both pipelines, our covariance matrix approach can trivially extract meaningful posterior profiles representing realistic spectra, whereas the PCA approach cannot as its posteriors are generated via perturbations around the maximum likelihood principal components. Thus, the posterior profiles extracted from the PCA approach have less physical meaning than those from the covariance matrix approach. However, while the PCA posteriors may not have physical shapes, one advantage of this approach are that they are drawn from errors from actual spectra rather than from errors in fitted Gaussian profiles.

[10]In Greig et al. (2017a) the training set was limited to QSOs with S/N >15, whereas in Davies et al. (2018b) it was S/N >7. As a result, after the removal of contaminants the training set sizes were 1653 and 12 764, respectively.

Equally, the methodology for extracting the IGM damping wing constraints is distinctly different between the two approaches. Our pipeline specifically focuses on avoiding the detailed modelling of the QSO host environment by only fitting for the damping wing imprint sufficiently redward of Ly α to avoid signatures of inflowing gas (motivating our fitting range of $\lambda = 1218$–1230 Å). This approach considerably minimizes modelling uncertainties. Whereas in Davies et al. (2018a), they use a hybrid scheme of numerical simulations to account for uncertainties within the QSO environment to allow them to utilize a much broader fitting region ($\lambda \sim 1190$–1230 Å, with the exact range differing on a per QSO basis).

Both approaches use large-scale semi-numerical simulations (i.e. 21CMFAST Mesinger, Furlanetto & Cen 2011; though with different physical prescriptions) to extract the smooth damping wing profiles from interactions with neutral hydrogen along the line of sight. However, by additionally modelling blueward of Ly α Davies et al. (2018a) must also use a hybrid scheme of hydrodynamical simulations to model the density environment of the QSO host coupled with 1D radiative transfer simulations to model the impact of the QSO flux on its nearby environment. As such, this adds in the additional complexity of the QSO ionizing contribution to the environment, which can be somewhat degenerate with the IGM neutral fraction and must be marginalized over when obtaining constraints from the IGM damping wing imprint. Therefore, while this approach has the advantage of fitting for the damping wing imprint over a broader spatial range, there are many additional complexities and uncertainties embedded in the models. On the other hand, by restricting ourselves to only considering redward of Ly α our constraints can be considered more model independent. However, our approach does assume a prior on the minimum H II region size surrounding the QSO, which effectively translates into a non-trivial prior on the QSO's ionizing contribution to it's environment. Such an ad hoc prior could bias results towards larger QSO lifetimes when the neutral fraction approaches unity and when the likelihood is not strongly constraining (see Greig et al. 2019 for further details).

Importantly, despite the notable differences in the modelling and methodology of the two pipelines, in this work we have found similar constraints on the IGM neutral fraction from both J1007 and J0252. However, this is not always the case. For example, for J1342, even despite the inclusion of N v increasing our recovered constraints on the IGM neutral fraction from this QSO, we still differ by $\sim 1\sigma$. In order to get a true handle on the similarities, differences and any potential biases inherent in either pipeline when applied to these high-z QSOs, it will be useful to perform a detailed comparison on a unified set of QSOs. However, we leave such an exploration to future work.

5 CONCLUSION

In this work, we perform an independent and complimentary IGM damping wing analysis on two recently discovered $z \gtrsim 7$ QSOs; J1007 (Yang et al. 2020a) at $z = 7.51$ and J0252 (Wang et al. 2020) at $z = 7.00$. Additionally, we update the existing Gaussian covariance matrix approach to predict the intrinsic QSO profile developed by Greig et al. (2017b) to additionally reconstruct the N v emission line. Using this modified reconstruction pipeline we then extract the damping wing imprint of the neutral IGM by jointly fitting the predicted intrinsic QSO flux with synthetic damping wing profiles drawn from realistic EoR simulations to the observed spectra of the $z \gtrsim 7$QSOs.

Specifically, we fit the emission lines of the observed QSO spectrum over $\lambda = 1275$–2300 Å and evaluate our measured covariance

matrix of various known emission lines. In doing so, we obtain a reduced covariance matrix (containing Ly α and N V emission line parameters) from which we extract 10^5 reconstructed intrinsic QSO profiles. We then combine these intrinsic profiles with 10^5 synthetic damping wing profiles drawn from large-volume EoR simulation (1.6 Gpc on a side with an EoR morphology driven by galaxies residing in $M_{\rm h} \gtrsim 10^9\,{\rm M}_\odot$ haloes, in our nomenclature referred to as an INTERMEDIATE H II model). We then jointly fit these 10^{10} profiles to the observed spectrum of J1007 and J0252 redward of Ly α ($\lambda = 1218$–1230 Å) in a Bayesian framework. Following this procedure, we recover the following constraints on the IGM neutral fraction (with 68 per cent confidence intervals):

(i) $\bar{x}_{\rm H\,I} = 0.27^{+0.21}_{-0.17}$ at $z = 7.51$ for J1007
(ii) $\bar{x}_{\rm H\,I} = 0.64^{+0.19}_{-0.23}$ at $z = 7.00$ for J0252.

These constraints are consistent (within error), albeit slightly lower than those previously reported in the literature; $\bar{x}_{\rm H\,I} = 0.39^{+0.22}_{-0.13}$ for J1007 (Yang et al. 2020a) and $\bar{x}_{\rm H\,I} = 0.70^{+0.20}_{-0.23}$ for J0252 (Wang et al. 2020). These can be attributed to the subtleties of the two methodologies: (i) our covariance matrix approach reconstructs multiple Gaussian emission profiles whereas the other works reconstruct on principal components (PCA) resulting in different profile shapes and (ii) we fit the damping wing only redward of Ly α ($\lambda = 1218$–1230 Å) whereas the other works additionally fit blueward of Ly α requiring additional modelling of the QSO host environment.

Additionally, since we have updated the covariance matrix reconstruction approach to include the N V emission line, we perform a reanalysis of the IGM damping wing constraints from both J1120 (Mortlock et al. 2011) at $z = 7.08$ and J1342 (Bañados et al. 2018) at $z = 7.54$. After including N V, under the INTERMEDIATE H II EoR morphology we find:

(i) $\bar{x}_{\rm H\,I} = 0.44^{+0.23}_{-0.24}$ at $z = 7.08$ for J1120
(ii) $\bar{x}_{\rm H\,I} = 0.31^{+0.18}_{-0.19}$ at $z = 7.54$ for J1342.

These are consistent within error to our previously reported results of $\bar{x}_{\rm H\,I} = 0.40^{+0.21}_{-0.19}$ for J1120 (Greig et al. 2017b) and $\bar{x}_{\rm H\,I} = 0.21^{+0.17}_{-0.19}$ for J1342 (Greig et al. 2019). Note that for J1120, the differences are minor as in that work we applied a prior on the reconstructed flux down to $\lambda = 1230$ Å allowing for the fitting of N V. For J1342, the increase in the IGM neutral fraction arises from the N V line component, and slightly diminishes the tension with the reported results from Davies et al. (2018a) from $\sim 1.5\sigma$ to $\sim 1\sigma$.

Finally, we combined the constraints on the IGM neutral fraction from all four known $z \gtrsim 7$ QSOs across the two analysis pipelines (covariance matrix, used in this work and the PCA approach in the literature) to obtain:

(i) $\bar{x}_{\rm H\,I} = 0.49^{+0.11}_{-0.11}$ at $z = 7.29$.

We found this IGM neutral fraction, with the relatively broad uncertainties obtained from averaging across all four $z \gtrsim 7$ QSOs, to be roughly consistent to within $\sim 1\sigma$ of the median constraints on the IGM neutral fraction obtained by forward-modelling recent Ly α forest data by Qin et al. (2021).

ACKNOWLEDGEMENTS

We thank the anonymous referee for their comments, Aaron Barth for providing the Keck NIRES data and both Aaron Barth and Xiaohiu Fan for reading an early version of this manuscript. We thank Yuxiang Qin for providing the reionization history posteriors. Parts of this research were supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. AM acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 638809 – AIDA). The results presented here reflect the authors' views; the ERC is not responsible for their use. FW acknowledge the support provided by NASA through the NASA Hubble Fellowship grant #*HST*-HF2- 51448.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Alam S. et al., 2015, ApJS, 219, 12
Bañados E. et al., 2018, Nature, 553, 473
Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, MNRAS, 447, 3402
Bolton J. S., Haehnelt M. G., Warren S. J., Hewett P. C., Mortlock D. J., Venemans B. P., McMahon R. G., Simpson C., 2011, MNRAS, 416, L70
Bosman S. E. I., Becker G. D., 2015, MNRAS, 452, 1105
Bosman S. E. I., Fan X., Jiang L., Reed S., Matsuoka Y., Becker G., Haehnelt M., 2018, MNRAS, 479, 1055
Bosman S. E. I. et al., 2021a, preprint(arXiv:2108.03699)
Bosman S. E. I., Ďurovčíková D., Davies F. B., Eilers A.-C., 2021b, MNRAS, 503, 2077
Choudhury T. R., Paranjape A., Bosman S. E. I., 2021, MNRAS, 501, 5782
Davies F. B. et al., 2018a, ApJ, 864, 142
Davies F. B. et al., 2018b, ApJ, 864, 143
Dawson K. S. et al., 2013, AJ, 145, 10
Ďurovčíková D., Katz H., Bosman S. E. I., Davies F. B., Devriendt J., Slyz A., 2020, MNRAS, 493, 4256
Eilers A.-C., Davies F. B., Hennawi J. F., 2018, ApJ, 864, 53
Fan X. et al., 2006, AJ, 132, 117
Fathivavsari H., 2020, ApJ, 898, 114
Greig B., Mesinger A., McGreer I. D., Gallerani S., Haiman Z., 2017a, MNRAS, 466, 1814
Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017b, MNRAS, 466, 4239
Greig B., Mesinger A., Bañados E., 2019, MNRAS, 484, 5094
Hoag A. et al., 2019, ApJ, 878, 12
Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020a, MNRAS, 491, 1736
Keating L. C., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020b, MNRAS, 497, 906
Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, MNRAS, 485, L24
Liu B., Bordoloi R., 2021, MNRAS, 502, 3510
Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018, ApJ, 856, 2
Mason C. A. et al., 2019, MNRAS, 485, 3947
McGreer I. D., Mesinger A., D'Odorico V., 2015, MNRAS, 447, 499
Mesinger A., Haiman Z., 2007, ApJ, 660, 923
Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955
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
Miralda-Escudé J., 1998, ApJ, 501, 15
Mortlock D. J. et al., 2011, Nature, 474, 616
Nasir F., D'Aloisio A., 2020, MNRAS, 494, 3080
Planck Collaboration XIII, 2016, A&A, 594, A13
Planck Collaboration VI et al., 2020, A&A, 641, A6
Qin Y., Mesinger A., Bosman S. E. I., Viel M., 2021, MNRAS, 506, 2390

Reiman D. M., Tamanas J., Prochaska J. X., Ďurovčíková D., 2020, preprint(arXiv:2006.00615)
Rybicki G. B., Lightman A. P., 1979, Radiative Processes in Astrophysics. Wiley-Interscience, New York
Simcoe R. A., Sullivan P. W., Cooksey K. L., Kao M. M., Matejek M. S., Burgasser A. J., 2012, Nature, 492, 79
Sobacchi E., Mesinger A., 2015, MNRAS, 453, 1843
Wang F. et al., 2020, ApJ, 896, 23
Wold I. G. B. et al., 2022, ApJ, 927, 36
Yang J. et al., 2019, AJ, 157, 236
Yang J. et al., 2020a, ApJ, 897, L14
Yang J. et al., 2020b, ApJ, 904, 26

APPENDIX A: COVARIANCE MATRIX WITH N V COMPONENT

In this work, we have added the N V emission line to the original covariance matrix reconstruction pipeline from Greig et al. (2017a). Previously, when performing the Monte Carlo Markov Chain (MCMC) fits to our full QSO training set, we fit for the N V emission line. However, we decided to ignore N V from the covariance matrix as the N V line profile parameters (peak, width, and velocity offset) did not strongly correlate with Ly α or the other primary high ionization lines (C IV, Si IV + O IV] and C III]) used for our Ly α reconstruction pipeline. Therefore, to include N V into our analysis we simply updated our covariance matrix using the existing data at hand.

To explore the strength of correlations between the N V line parameters and the other lines, in Fig. A1 we present the updated correlation matrix of the 1673 QSOs in the training sample. This correlation coefficient matrix, $\boldsymbol{R}_{ij}$, is defined as

$$\boldsymbol{R}_{ij} = \frac{\boldsymbol{C}_{ij}}{\sqrt{\boldsymbol{C}_{ii}\boldsymbol{C}_{jj}}}, \tag{A1}$$

where each diagonal entry, $\boldsymbol{R}_{ij}$, represents the correlation coefficient between the ith and jth emission line parameter and $\boldsymbol{C}_{ij}$ is our full

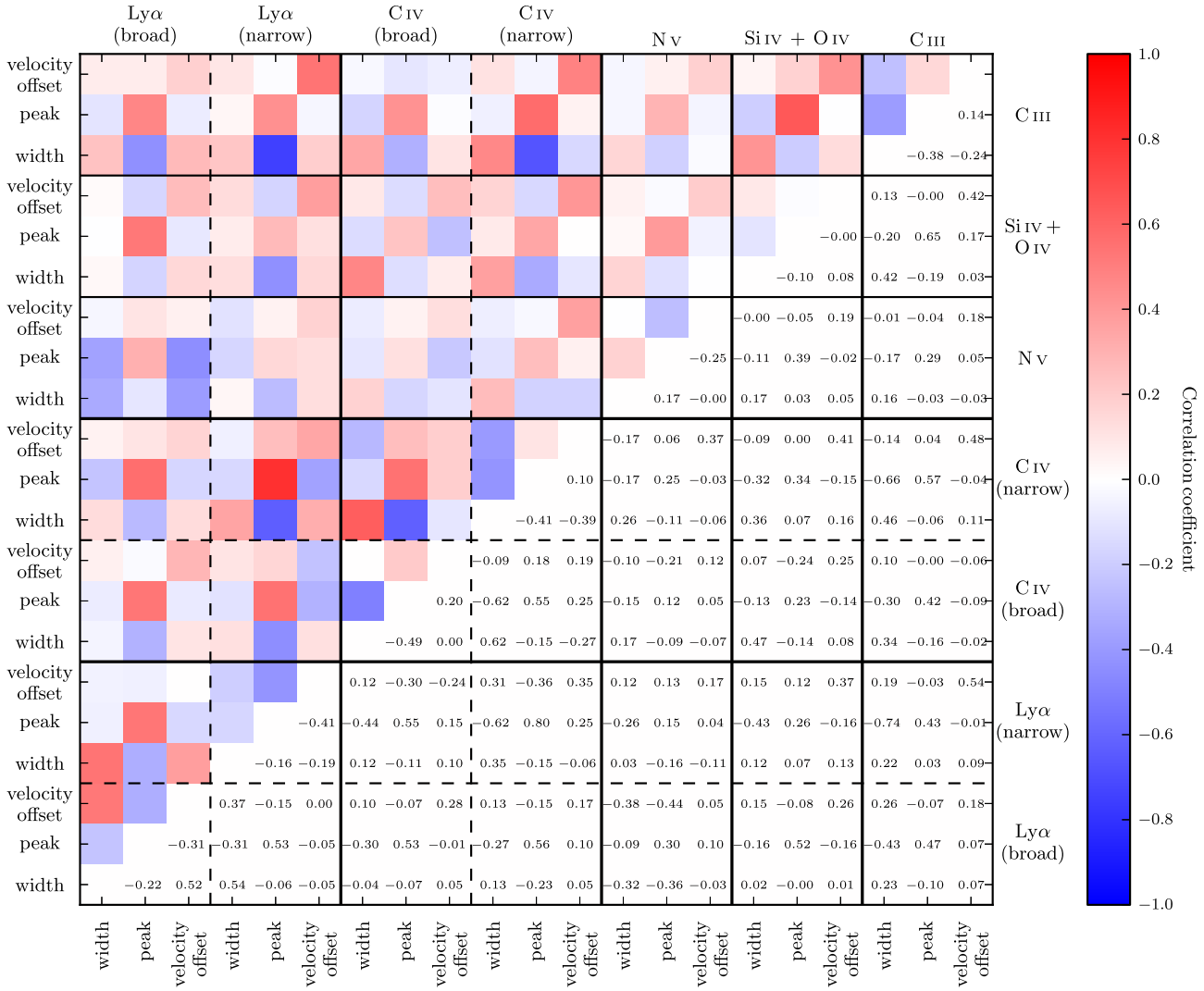

Figure A1. The updated correlation coefficient matrix (correlation coefficients listed in the lower half) including the N V component constructed from the 'good' sample of 1673 QSOs from Greig et al. (2017a). This 21 dimensional matrix contains double component Gaussians for both Ly α and C IV and single component Gaussians for N V, Si IV + O IV] and C III]. For each Gaussian line component we have three parameters, the peak width, peak height, and velocity offset from systemic.

data covariance matrix defined as

$$\boldsymbol{C}_{ij} = \frac{1}{N-1}\sum_{i}^{N}(\boldsymbol{X}_i - \boldsymbol{\mu}_i)(\boldsymbol{X}_j - \boldsymbol{\mu}_j). \quad \text{(A2)}$$

Here, $\boldsymbol{X}_i$ is the data vector containing all ith emission line parameters from the full QSO sample and $\boldsymbol{\mu}$ is its mean.

Relative to the other emission lines in our sample, the correlations with any N v emission line parameters are moderate to weak. Primarily, N v covaries strongest with the broad Ly α component which makes sense under the assumptions of our model (fitting multiple Gaussian profiles). The extended wings of the broad Ly α component can overlap with the N v line component owing to their reasonably close separation ($\sim$20 Å). This is evident through the measured anticorrelation between the width of the broad Ly α component and the width of the N v line ($\rho = -0.32$). Note, there are a couple of other interesting correlations between the N v line and individual line properties (e.g. amplitude or velocity offset). For example, between the velocity offsets of N v and the narrow component of C IV and between the peak amplitude of the N v and the Si IV + O IV] line profile. The former highlights the importance of including N v into the damping wing procedure, as it implies that a strongly blue-shifted C IV narrow line component (as appears to be the case with these $z \gtrsim 7$ QSOs) would result in a strongly blueshifted N v line either being close to or entering into our damping wing fitting region.

This implies that the reconstruction of the N v emission is primarily dependent on the reconstruction of the Ly α emission and is essentially independent on all other considered emission lines in our covariance matrix. Overall, this will result in larger uncertainties in the reconstructed profiles due to: (i) an increased number of parameters in the model and (ii) the weak dependency of N v to all other lines producing a larger range of plausible profiles.

APPENDIX B: REANALYSIS OF PREVIOUS QSOS WITH N v COMPONENT

As we have modified our reconstruction pipeline in this work to simultaneously predict the N v emission line, it is prudent we perform a reanalysis of our previous work. In Sections B1 and B2, we provide the updated reconstructed intrinsic QSO profiles along with the new constraints on the IGM neutral fraction for J1120 and J1342, respectively.

B1 ULASJ1120+0641

In Fig. B1, we present the updated maximum likelihood reconstruction for J1120 (blue curve) including the N v line component relative to the original reconstruction (Greig et al. 2017b) without N v (red curve). The thin grey curves correspond to 300 randomly drawn profiles from the full posterior distribution of reconstructed profiles (including N v). First, it is important to note that in Greig et al. (2017b), the reconstruction of the intrinsic profile used a flux prior over λ = 1230–1275Å. However, since Greig et al. (2019) a flux prior has only been applied over λ = 1250–1275 Å. As such, the original reconstruction for J1120 includes a N v emission line as we were able to fit for it in combination with the Ly α reconstruction.

Interestingly, the reconstructed N v emission line closely resembles that of the fitted N v emission line from Greig et al. (2017b). This gives us confidence that the reconstruction pipeline works sufficiently well following the inclusion of N v into the covariance matrix. The primary difference between the two reconstruction methods is the decrease in amplitude of the narrow Ly α line component, along with a marginal broadening of the line. Further, the broad Ly α line component is reduced in amplitude as indicated by the systematically lower flux amplitude for the full profile across λ = 1190–1230 Å. This lower amplitude Ly α profile somewhat reduces the difference in profile shapes between the covariance matrix and PCA pipelines.

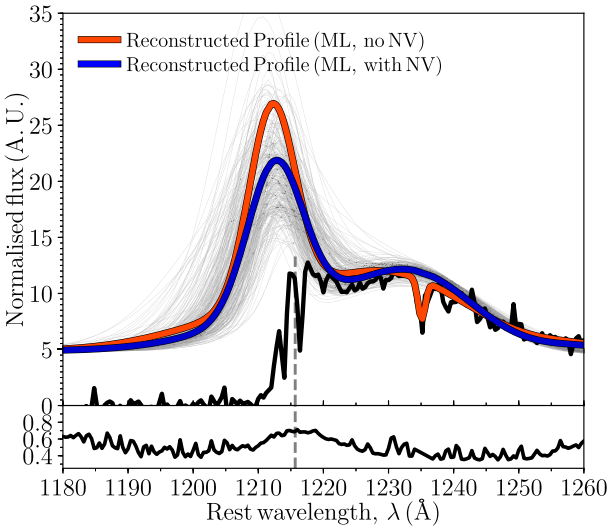

Figure B1. A reanalysis of the reconstructed maximum likelihood Ly α emission line profile including a N v line component for ULASJ1120+0641 ($z = 7.08$). The red curve corresponds to the previous maximum likelihood reconstruction from Greig et al. (2017a) whereas the blue curve is the new maximum likelihood reconstruction including the N v component. Note previously, in Greig et al. (2017a) features in the QSO spectrum were fit up to 1230 Å, thus N V was fit rather than reconstructed. We also present a subsample of 300 Ly α line profiles (thin grey curves) randomly drawn from the full posterior distribution of reconstructed profiles (including N v). The black curve corresponds to the FIRE spectrum (Simcoe et al. 2012), with the corresponding error spectrum shown in the sub-panel below. The vertical grey dashed line denotes rest-frame Ly α.

For the updated reconstruction of J1120 we recover IGM neutral fraction constraints of $\bar{x}_{\rm HI} = 0.44^{+0.23}_{-0.24}$ for the INTERMEDIATE H II EoR morphology. In Greig et al. (2017b), we had previously recovered constraints of $\bar{x}_{\rm HI} = 0.46^{+0.21}_{-0.21}$. Clearly, owing to the relatively similar amplitude and shape between the two reconstructions over the damping wing fitting region (λ = 1218–1230 Å) we recover essentially the same constraints on the IGM neutral fraction. The main difference to note is that following the inclusion of N v into the pipeline, the relative amplitude of the confidence intervals have increased owing to the increased scatter in reconstructed profiles due to the expanded covariance matrix. The slight decrease in the recovered IGM neutral fraction arises due to the decreased flux amplitude of the reconstructed profile between λ = 1220–1227 Å. The decrease in the IGM neutral fraction however is minimized by the slight increase in predicted flux in the λ = 1218–1221 Å region. The preference for a slightly wider narrow Ly α line component will add some weight to damping wing profiles from a more neutral IGM.

B2 ULASJ1342+0928

In Fig. B2, we present the updated maximum likelihood reconstruction for J1342 (blue curve) including the N v line component relative to the original reconstruction (Greig et al. 2019) without N v (red curve). Once again, the thin grey curves represent 300 randomly

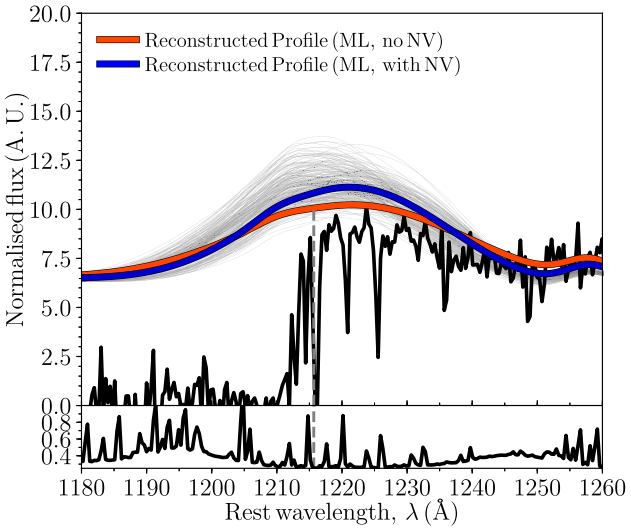

Figure B2. The same as Fig. B1, except showing the reanalysis of ULASJ1342+0928 ($z = 7.54$) originally presented in Greig et al. (2019). The black curve corresponds to the combined Magellan/FIRE and Gemini/GNIRS spectrum (Bañados et al. 2018), with the corresponding error spectrum shown in the sub-panel below. The vertical grey dashed line denotes rest-frame Ly α.

drawn profiles from the full posterior distribution of reconstructed profiles (including N v).

Immediately evident is the increased amplitude of the predicted flux between $\lambda = 1210$–1230 Å. This increase is primarily driven by a blueshifted, broad N v emission line. Equally, this broader N v component results in a slightly narrower broad Ly α component (due to the anticorrelation between the line widths discussed in Section A) which can then lead to an increased peak amplitude for the broad Ly α component. Again, this reduces the difference between the maximum likelihood reconstruction profiles from the covariance matrix and PCA reconstruction pipeline highlighting the importance of including N v into the reconstruction pipeline.

Performing our joint analysis for the IGM damping wing (Intermediate H II EoR morphology), we find $\bar{x}_{\rm HI} = 0.31^{+0.18}_{-0.19}$ at 68 per cent confidence. In contrast, in Greig et al. (2019) we recovered an IGM neutral fraction of $\bar{x}_{\rm HI} = 0.21^{+0.17}_{-0.19}$. This increase in the recovered IGM neutral fraction is driven by the increased amplitude across the entire damping wing fitting region used in our analysis due to the addition of the N v emission line component. Previously, for J1342 we had noted that the results from our covariance matrix approach was mildly in tension with the equivalent PCA analysis from Davies et al. (2018a), who recovered constraints of $\bar{x}_{\rm HI} = 0.60^{+0.20}_{-0.23}$. While the results from the two pipelines are still notably different, the difference is now closer to $\sim 1\sigma$ compared to the $\sim 1.5\sigma$ reported previously.

In this work, we explored this difference in the recovered IGM neutral fraction a little further given that both pipelines now predict relatively similar reconstructed Ly α profiles. To do this, we perform the same analysis as in Section 2.4, except we do not average over the full distribution of reconstructed profiles from our covariance matrix pipeline. That is, we only consider the maximum likelihood reconstruction profile while averaging over all synthetic damping wing profiles. In doing so, we recover $\bar{x}_{\rm HI} = 0.48^{+0.16}_{-0.14}$. This implies that the full distribution of reconstruction profiles tends to sit on average lower in amplitude than the maximum likelihood profile or have reconstructed Ly α profile shapes which prefer slightly more gradual damping wing profiles (i.e. lower amplitude neutral fractions) to match the observed spectrum. This goes some way towards explaining the difference between the two reconstruction pipelines. Further differences likely arise due to the shape of the reconstructed profiles blueward of Ly α. The reconstructed profiles of the Davies et al. (2018a) PCA approach tend to be larger in amplitude blueward of Ly α where they also fit for the damping wing profile (unlike only the redside in our work). Thus, having higher amplitude reconstructed profiles at these blueward wavelengths would prefer higher IGM neutral fraction values to be able to suppress the reconstructed profiles to match the observed flux.

APPENDIX C: MCMC FITTING OF THE QSOS

In order to perform our covariance matrix reconstruction pipeline, we perform an MCMC fit to the observed spectrum over the range $\lambda = 1275$–2300 Å (see Section 2.2 for further details). In Fig. C1, we provide the MCMC fit to J1007. In the top panel, we show the two component power-law QSO continuum (red-dashed curve), while in the remaining panels we provide zoomed-in panels centred on all the emission lines that we fit in our MCMC pipeline. In Fig. C2, we provide the equivalent MCMC fit to J0252.

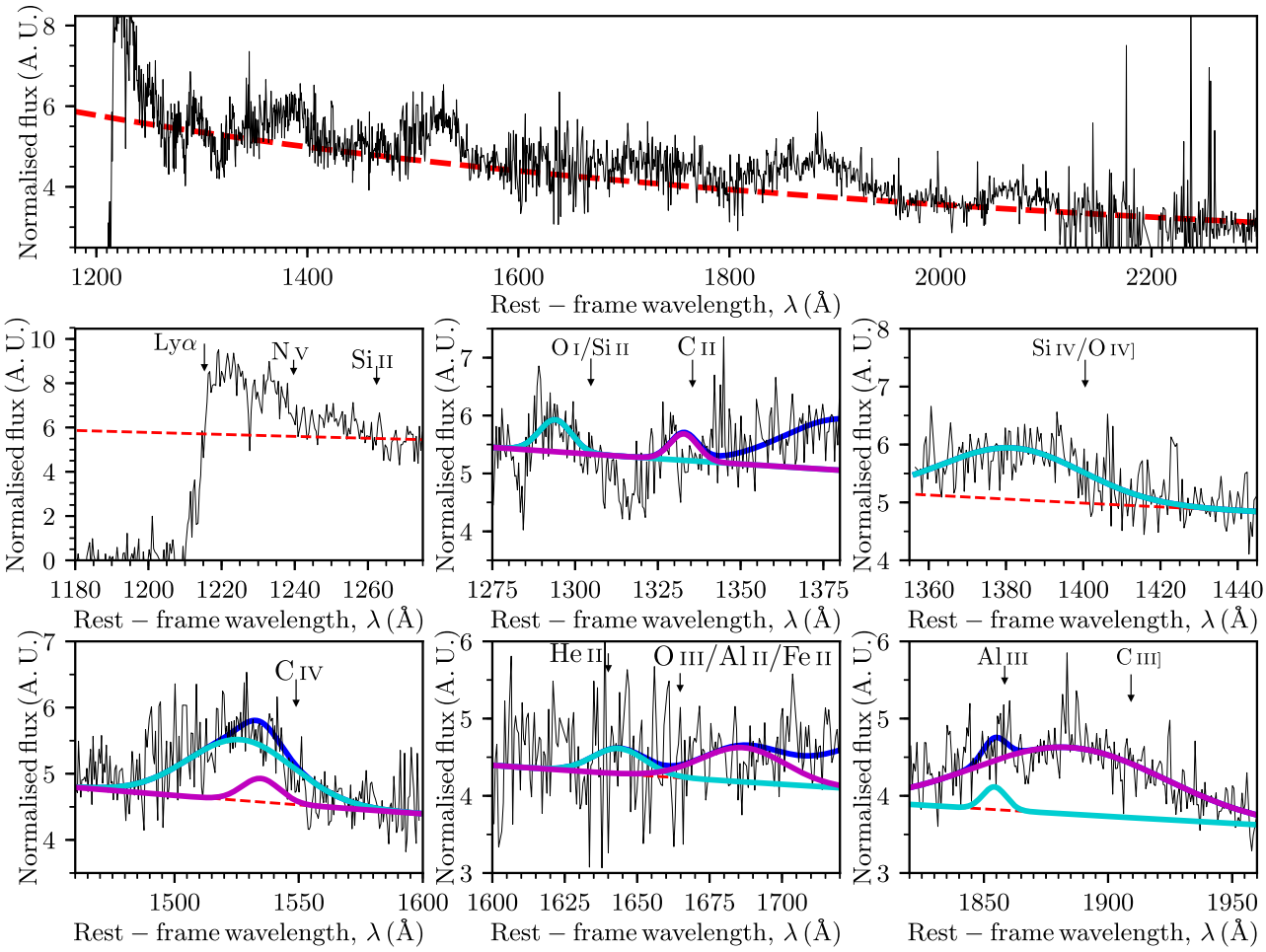

Figure C1. A zoom-in highlighting the MCMC fitting procedure from Greig et al. (2017a) applied to the combined GNIRS and NIRES spectrum of J1007 (Yang et al. 2020a). The QSO continuum flux is normalized at 1450 Å rest frame ($f_\lambda \propto \left(\frac{\lambda}{1450\text{Å}}\right)^\alpha$; 1 au $= 10^{-18}$ cm^{-2} s^{-1} Å^{-1}). Top panel: a single power-law continuum (red dashed curve) is fit to the QSO spectrum. Middle left: the attenuated Ly α profile. Middle centre: the low ionization lines, O I/Si II] (cyan), and C II (magenta). Middle right: the blended Si IV + O IV] line complex fit with a single component Gaussian. Bottom left: the C IV line fit with a double component Gaussian. Bottom centre: Low ionization lines, He II and O III/Al II/Fe II. Bottom right: single component Gaussians to describe the C III] line (magenta) and Al III (cyan) emission lines.

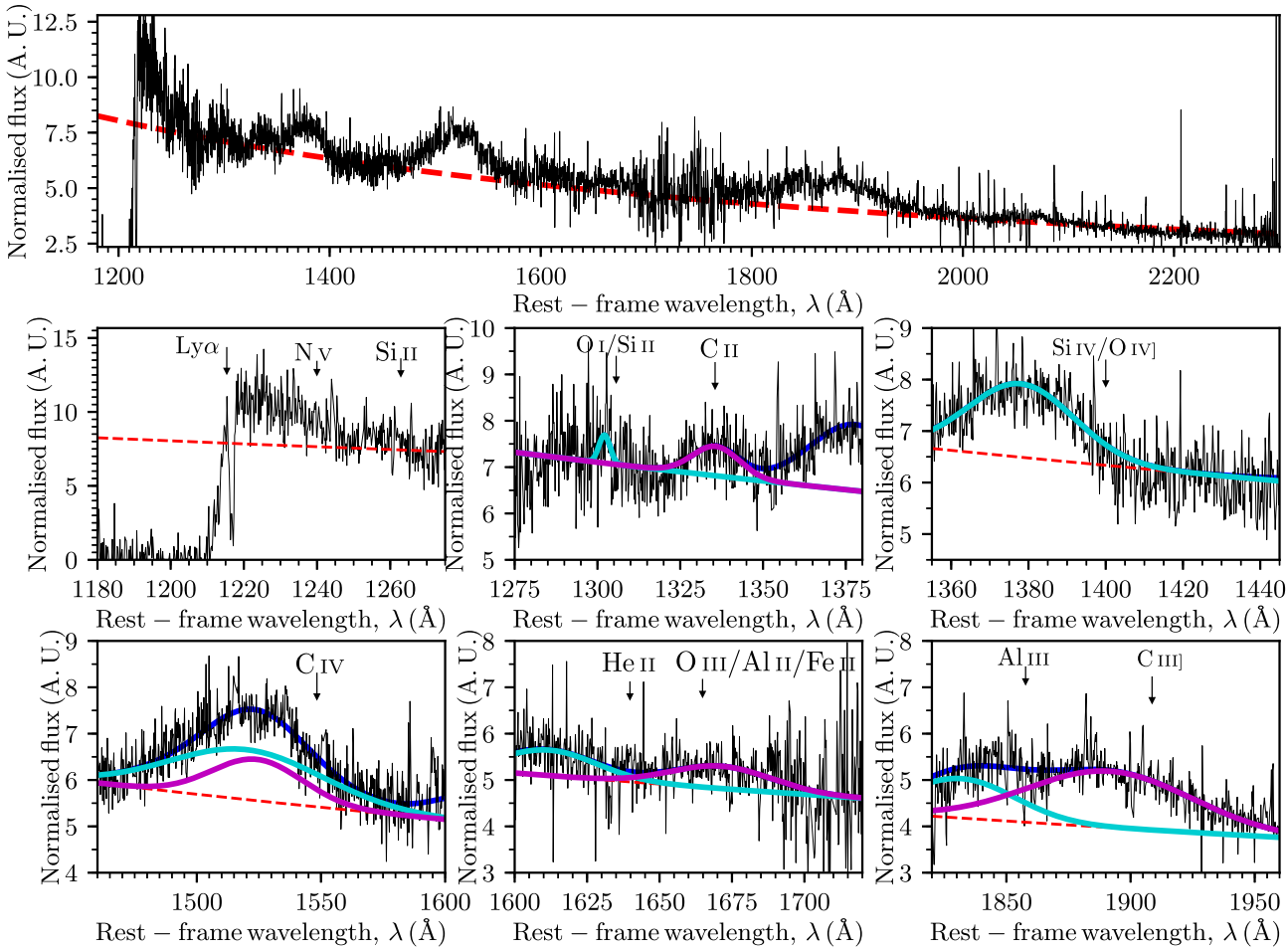

Figure C2. Same as Fig. C1, except applied to the Gemini/GMOS + Keck/NIRES spectrum of J0252 (Wang et al. 2020).

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