UNIVERSITY OF CALIFORNIA
RIVERSIDE

Constraining the Intergalactic Medium During the Later Stages of Reionization Using QSO Spectra

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Physics

by

Yongda Zhu

March 2024

Dissertation Committee:

Dr. George Becker, Chairperson
Dr. Simeon Bird
Dr. Anson D'Aloisio

The Dissertation of Yongda Zhu is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to express my deepest gratitude to my advisor, George Becker, for his generous help and invaluable guidance over the past five years. He has taught me to think like an astrophysicist: keeping the big pictures in mind, focusing on the most important questions in the field, and always considering the physics behind the data. George has provided a wealth of insightful advice on project details and writing, while also granting me the freedom and space to develop the skills crucial for independent research. Beyond research, George has been an important support in my career development and day-to-day life as an international student. In my opinion, he is the best mentor one graduate student can dream for. Thank you, George!

I extend my gratitude to my committee members, Anson D'Aloisio and Simeon Bird, for their invaluable help throughout my Ph.D. journey. Special thanks to Elisa Boera and Fahad Nasir for their mentorship when I started exploring reionization. I am grateful to my collaborators for their support and contributions to the research presented in this dissertation, including Sarah Bosman, Laura Keating, Valentina D'Odorico, Rebecca Davies, Holly Christenson, Eduardo Bañados, Fuyan Bian, Manuela Bischetti, Huanqing Chen, Frederick Davies, Anna-Christina Eilers, Xiaohui Fan, Prakash Gaikwad, Bradley Greig, Martin Haehnelt, Girish Kulkarni, Samuel Lai, Andrea Pallottini, Yuxiang Qin, Emma Ryan-Weber, Fabian Walter, Feige Wang, Jinyi Yang, Anson D'Aloisio, Tom Bakx, Christopher Cain, Hai-Xia Ma, Andrei Mesinger, Sindhu Satyavolu, Tsutomu Takeuchi, and Hideki Umehata. Also, thanks to my teammates, Holly Christenson, Azeem Bari, and Azim Hashemi, for their camaraderie and support.

I want to thank my undergraduate advisor, Xiao-Bo Dong, who has provided great encouragement for me to continue pursuing the career as a researcher. I am also grateful to other professors who have offered valuable advice during my Ph.D. journey, including Yongtao Cui, Gabriela Canalizo, Brant Robertson, and Matt Malkan. My appreciation extends to my peer mentors and friends, including Krista Lepiane, Ming-Feng Ho, and Mahdi Qezlou. Finally, I want to thank my family for their love and support.

Support for this work was provided by the NSF through award SOSPADA-029 from the NRAO, and through grants AST-1615814 and AST-1751404. This work is based on observations collected at the European Southern Observatory under ESO program 1103.A-0817. Some of the data presented herein were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation. I wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. I am most fortunate to have the opportunity to conduct observations from this mountain. Finally, this research has made use of the Keck Observatory Archive (KOA), which is operated by the W.M. Keck Observatory and the NASA Exoplanet Science Institute (NExScI), under contract with the National Aeronautics and Space Administration.

This work makes use of the following ALMA data: ADS/JAO.ALMA # 2022.1.00662.S, # 2021.1.01018.S, and # 2019.1.00111.S. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada),

MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

This dissertation makes use of the following papers: 2021ApJ...923..223Z, 2022ApJ ...932...76Z, 2023ApJ...955..115Z, and Zhu et al. (to be submitted).

To my wife, Weiyi:

In the vastness of space and the immensity of time,

it is my joy to share a planet and an epoch with you.

ABSTRACT OF THE DISSERTATION

Constraining the Intergalactic Medium During the Later Stages of Reionization Using QSO Spectra

by

Yongda Zhu

Doctor of Philosophy, Graduate Program in Physics
University of California, Riverside, March 2024
Dr. George Becker, Chairperson

The reionization of neutral hydrogen in the intergalactic medium (IGM) represents the last major phase transition in our universe's history. The timing and sources responsible for reionization carry important implications for the formation and evolution of early galaxies. This dissertation focuses on the IGM during the later stages of reionization using hydrogen absorption lines in spectra of QSOs at redshift $5.0 \lesssim z \lesssim 6.5$. In our search of 55 high-quality QSO spectra from Keck/ESI and VLT/X-Shooter for dark gaps (where >95% of Lyα photons are absorbed), we discovered that long dark gaps (length $L \geq 30h^{-1}$Mpc) extend down to $z \simeq 5.4$. Our comparison of the dark gap distribution in observed spectra with mock data suggests that the scenario where reionization completes by $z = 6$ with a homogeneous ultraviolet background (UVB) is disfavored. However, models featuring neutral islands and/or large-scale UVB fluctuations persisting down to $z \lesssim 5.5$ align with our findings. We also report the discovery of an extremely long dark gap in the Lyβ forest, persisting down to $z \sim 5.5$. We derive the upper limits on IGM neutral fraction from dark gap length distributions based on a conservative model. The high fraction of QSO spectra

exhibiting long Lyβ dark gaps further suggests that the IGM could remain considerably neutral (approximately 15%) as late as $z = 5.8$. We also directly measure the mean free path of ionizing photons ($\lambda_{\rm mfp}$) over $5 < z < 6$. We fit the Lyman continuum ($\lambda_{\rm rest} < 912$Å) transmission in composite QSO spectra with the proximity zone effect included. The results convincingly demonstrate a rapid evolution of $\lambda_{\rm mfp}$ over $5 < z < 6$ and provided the strongest evidence to date that reionization ends significantly below $z = 6$. Finally, we present the discovery of a damping wing-like absorption profile in the stacked Lyα forest at the red edge of Lyβ dark gaps. Such an extended absorption profile likely originates from genuine neutral hydrogen islands, and suggests that the IGM is at least $6.1 \pm 3.9\%$ neutral at $z \approx 5.8$. Our results offer robust evidence supporting the extension of reionization below $z = 6$.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background: cosmic reionization

Reionization is the last major phase transition in the Universe (see e.g., Fan et al. 2023 for a recent review). During the epoch of reionization (EoR), neutral hydrogen atoms in the intergalactic medium (IGM) were ionized by ultraviolet (UV) photons from the first luminous objects (e.g., Bromm & Larson 2004; McQuinn 2016). Understading the EoR, therefore, is key to our understanding of the formation and evolution of these luminous objects, including first stars, supermassive blackholes (SMBHs), and galaxies. Under the ΛCDM framework (e.g., Ostriker & Steinhardt 2003), these luminous objects are futher related to the properties of dark matter and dark energy.

We need some quantities to describe reionization, and the connection between sources and the IGM. The ionizing emissivity for a source (e.g., a galaxy) is given by

$$\dot{n}_{\rm ion} \propto L_{\rm UV} \xi_{\rm ion} f_{\rm esc} \,, \tag{1.1}$$

where $L_{\rm UV}$ denotes the UV luminosity, $\xi_{\rm ion}$ is the ionizing efficiency, a factor that converts the UV luminosity into the number of ionizing photons, and $f_{\rm esc}$ is the escape fraction, which represents the fraction of ionizing photons that escape to the IGM from the galaxy. The ionization rate per hydrogen atom (Γ), can be determined by the flux of ionizing photons and the cross-section for photoionization. Using a local source approximation, Γ is given by

$$\Gamma \propto \dot{n}_{\rm ion} \lambda_{\rm mfp} \,, \tag{1.2}$$

where $\lambda_{\rm mfp}$ is the mean free path of ionizing photons. On large scales, the mean free path also sets the strength of fluctuations in Γ: a small $\lambda_{\rm mfp}$ creates large fluctuations in the ionizing UV background (UVB; see e.g., Davies & Furlanetto 2016; Nasir & D'Aloisio 2020), while a infinitely large $\lambda_{\rm mfp}$ yields a homogeneous UVB. Then, assuming photoionization equilibrium, the properties of the IGM follow

$$n_{\rm H} X_{\rm HI} \Gamma = n_{\rm H} (1 - X_{\rm HI}) \alpha(T) n_e \,, \tag{1.3}$$

where $n_{\rm H}$ is the total number density of hydrogen atoms, $X_{\rm HI}$ is the fraction of hydrogen atoms that are neutral, $\alpha(T)$ is the recombination coefficient, which is a function of the temperature T, and n_e is the electron number density.

The two most important questions about reionization are 1) what source(s) drive reionization and 2) what is the timeline. For the first question, determining whether randomly distributed faint sources (a "democratic" model) or rare and bright sources (an "oligarchic" model) dominate reionization is of interest (e.g., Cain et al. 2023). At high redshifts, however, observationally determining $\xi_{\rm ion}$ and $f_{\rm esc}$ can be challenging due to the faintness of the objects, the lack of robust constraints on the spectral energy distributions

(SEDs), and the high opacity of the IGM, which strongly absorbs the ionizing part of the spectrum. Thus, measuring ionizing emissivity "directly" and precisely can be difficult (Equation 1.1). In terms of determining which type of sources that drive reionization, measurements of the QSO luminosity functions near $z \sim 6$ suggest that QSOs are less dominant in producing ionizing photons for cosmic reionization (e.g. Parsa et al. 2018). On the galaxy side, Sharma et al. (2016) and Naidu et al. (2020) suggest that brighter, more massive galaxies with higher Lyman continuum (LyC) escape fractions played a significant role in reionizing the Universe (see also Bouwens et al. 2015). However, Rutkowski et al. (2016) found that the LyC escape fraction of faint, low-mass galaxies at $z \sim 1$ are weakly constrained, making it difficult to reconcile their role in reionization with observations. We expect observations from James Webb Space Telescope (JWST) will provide more clues in the near future.

In contrast, significant progress has been made toward answering the second question. About two decades ago, initial observations from Wilkinson Microwave Anisotropy Probe (WMAP) suggested that reionization took place from $11 < z < 30$ (Kogut et al. 2003). Later on, improved Cosmic microwave background (CMB) measurements from Planck indicate a mid-point of reionization at redshift $z_{\rm re} = 7.7 \pm 0.7$ (Planck Collaboration et al. 2020) — reionization occurred at a later time. Such a mid-point of reionization is consistent with other observations, including the the redshift evolution in the fraction of UV-selected galaxies detected in Lyα emission (e.g., Mason et al. 2018, 2019; Hoag et al. 2019; Umeda et al. 2023), the IGM thermal history at $z > 4$ inferred from the Lyα flux power spectrum (e.g., Boera et al. 2019, see also Walther et al. 2019; Gaikwad et al. 2021), and Lyα damping

wing measurements of $z > 7$ QSOs (e.g., Greig et al. 2017; Bañados et al. 2018; Davies et al. 2018b; Yang et al. 2020a). Figure 1.1 shows that a mid-point at $z \sim 7-8$ is also consistent with the measurements of dark pixels in the Lyα and Lyβ forests (Jin et al. 2023).

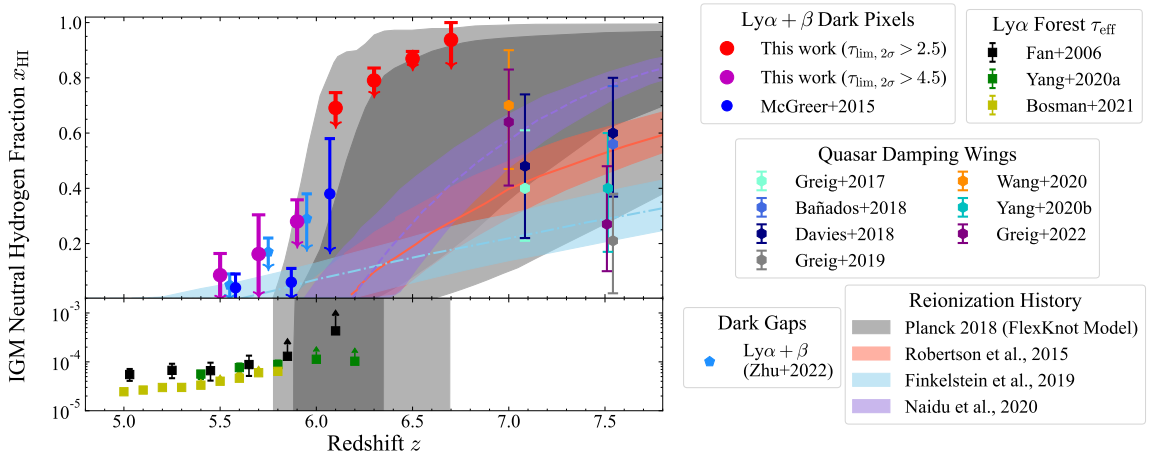


Figure 1.1: Constraints on the IGM neutral hydrogen fraction from high-z QSO studies and Planck 2018 results. Figure is from Jin et al. (2023) — *(Nearly) Model-independent Constraints on the Neutral Hydrogen Fraction in the Intergalactic Medium at $z \sim 5-7$ Using Dark Pixel Fractions in Lyα and Lyβ Forests, The Astrophysical Journal, 942:59*, and is used under the terms of the Creative Commons Attribution 4.0 licence.

On the other hand, the appearance of transmitted flux in the Lyα forest at $z \sim 6$ in Fan et al. (2006) has long been interpreted as the signature of the end of reionization. Such an early end of reionization (although it is "late" compared to WMAP 2003 results), however, is difficult to reconcile with the mid-point constrained above because star-forming galaxies would have to contribute ionizing photons extremely efficiently to finish half of reionization during a very short timeframe while not over-ionize the Universe at lower redshifts. An early end of reionization is also difficult to account for the observed large

scale fluctuations in the effective optical depth, $\tau_{\rm eff} = -\ln\langle F\rangle$, where F is the continuum-normalized transmission flux, at $z < 6$ in e.g., Fan et al. (2006).

Recently, the possibility of "ultalate" reionization has gained support from observations. The observations include improved and updated Lyα forest measurements that show large-scale $\tau_{\rm eff}$ fluctuations down to $z \lesssim 5.5$ (e.g., Becker et al. 2015; Eilers et al. 2018; Yang et al. 2020b; Bosman et al. 2018, 2022). Becker et al. (2015) also report the discovery of an extremely long dark gap (or Gunn-Peterson trough, Gunn & Peterson 1965) spanning $\sim$ 110 comoving-h^{-1}Mpc and extending down to $z \sim 5.5$ toward QSO ULAS J0148+0600. Such an extreme trough supports the picture that reionization is a patchy process and might end late (e.g., Keating et al. 2020b). The rapid counter evolution of the comoving number density of O I absorbers (Becker et al. 2019) and C IV absorbers (Davies et al. 2023a) over $5 \lesssim z \lesssim 6$ may also indicate a change in the ionzing UVB during or shortly after reionization.

The fact that the most extreme Lyα troughs are found in underdense regions (Becker et al. 2018; Kashino et al. 2020; Christenson et al. 2021) is also consistent with late-ending reionization. On the other hand, these observations are broadly consistent with a scenario wherein the IGM is fully ionized by $z = 6$ but retains large-scale fluctuations in the meta-galactic ionizing UVB (e.g., Davies & Furlanetto 2016; D'Aloisio et al. 2018). We note that, however, Christenson et al. 2023 report that two transmissive QSO sightlines are also surrounded by underdense regions traced by Lyman-alpha emitting galaxies (LAEs). At this point it is unclear whether any of the existing models can explain the IGM opacity-density relation naturally.

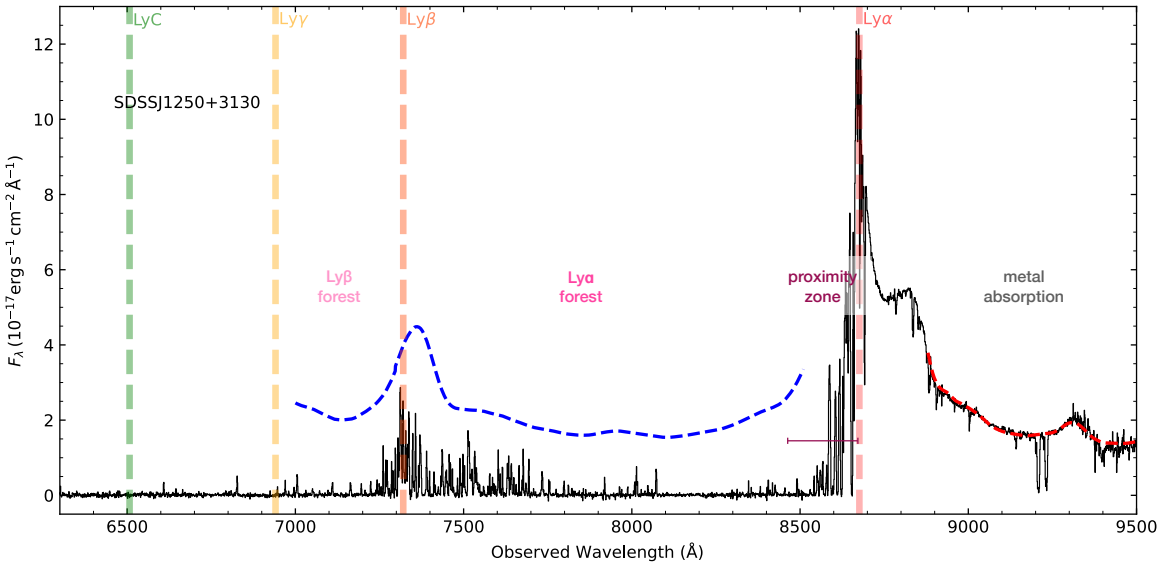


Figure 1.2: A Keck/ESI spectrum of QSO SDSS J1250+3130 at $z = 6.15$ used in Zhu et al. (2021). Red and blue dashed curves show the PCA fitting and prediction of the intrinsic transmission, respectively. Important features are labeled accordingly for reference.

1.2 Hydrogen absorption lines in QSO spectra

QSOs, also known as quasars, are extremely luminous sources powered by accreting supermassive blackholes. Figure 1.2 shows an example optical spectrum of a high-redshift QSO at $z \sim 6$. The intrinsic QSO transmission is mainly composed of emission lines including hydrogen recombination lines (Lyα, Lyβ, etc.), collisionally excited metal lines (e.g., Si IV, C IV, etc.), and continuum emission mainly contributed by thermal radiation from the accretion disk. Jets and outflows can cause absorption features near or redward of the systemic redshift. The observed wavelength λ is redshifted from the restframe wavelength λ_0 by

$$\lambda = (1 + z_{\rm QSO})\lambda_0 \,. \tag{1.4}$$

The IGM is primarily composed of hydrogen. Hydrogen has the simplest atomic structure and the absorption features are well understood. Therefore, we can use hydrogen absorption lines in QSO spectra to learn about the IGM and constrain reionization process along the line of sight. This following subsections briefly introduce the absorption features blueward the Lyα peak in high-z QSO spectra.

1.2.1 The Lyα forest

The Lyα transition corresponds to an electron transitioning between the ground state and the first excited state, and has a rest-frame wavelength of $\lambda_\alpha = 1215.67$ Å. In the region known as the "Lyα forest," found between the Lyβ emission and the Lyα peak in QSO spectra, a variety of absorption features are observed. This region primarily includes numerous Lyα absorption lines due to neutral hydrogen along the line of sight. Due to the Universe expansion, each line represents light that has been absorbed at a different redshift. This region also contains absorption lines from other elements and phenomena, such as metal absorption lines from intervening galaxies or gas cloud, broad absorption lines (BALs) attributed to high-velocity QSO outflows, etc.

We use optical depth, τ, to describe how opaque the IGM is to Lyα transmission. The observed flux is given by

$$F_\lambda = F_{\lambda,0} e^{-\tau} \,, \tag{1.5}$$

where $F_{\lambda,0}$ is the intrinsic continuum. The optical depth for an ensemble of neutral hydrogen atoms along the line of sight can be described by the summation of a series of Voigt profiles. Given the Lyα cross section $\sigma_\alpha = 4.48 \times 10^{-18}$ cm^2, if we know the column density of

neutral hydrogen $N_{\rm HI}$ and the Doppler parameter $b_{\rm HI} = \sqrt{(2k_{\rm B}T/m_{\rm H})^2 + b_{\rm turb.}^2}$, the optical depth and can be calculated as

$$\tau(\nu) = N_{\rm HI}\frac{c\sigma_\alpha}{b_{\rm HI}\pi^{1/2}}H(a,x)\,, \tag{1.6}$$

following Bolton (2007). Here, $a = \Lambda_\alpha\lambda_\alpha/4\pi b_{\rm HI}$ with $\Lambda_\alpha = 6.265\times10^8$ s^{-1} being the damping constant, and $x = [c(\nu-\nu_\alpha)]/\nu_\alpha b_{\rm HI}$. $H(a,x)$ is the Hjerting function and is defined as

$$H(a,x) = \frac{a}{\pi}\int_{-\infty}^{\infty}\frac{e^{-y^2}}{a^2+(x-y)^2}dy\,. \tag{1.7}$$

To test models or simulations, we often build mock spectra and compare them with observations. To construct synthetic Lyα forest from simulated skewers, for each pixel with a Hubble velocity $v_{\rm H}$ and proper width dr, we need the corresponding neutral hydrogen number density $n_{\rm HI}$, overdensity δ, IGM temperature T, and peculiar velocity $v_{\rm pec}$. Following Bolton (2007), the optical depth to Lyα in pixel i is

$$\tau(i) = \frac{c\sigma_\alpha\delta dr}{\pi^{1/2}}\sum_{j=1}^{N}\frac{n_{\rm HI}(j)}{b_{\rm HI}(j)}H(a,x)\,, \tag{1.8}$$

where $x = [v_{\rm H}(i)-u(j)]/b_{\rm HI}(j)$, $u(j) = v_{\rm H}(j)+v_{\rm pec}(j)$, and $a = \Lambda_\alpha\lambda_\alpha/4\pi b_{\rm HI}(j)$. In Python, the Hjerting function can be conveniently implemented with the `wofz` function provided by `SciPy` (Virtanen et al. 2020):

```
from scipy.special import wofz
def Hjerting(a, x):
    z = x + 1j*a
    return wofz(z).real
```

This implementation will provide a better computational performace compared to numerical integrals. After we acquire the optical depth arrays, we can convert them to Lyα transmission arrays, and add noise to build more realistic mock spectra (see Chapter 2).

One important thing to consider is that we can only observe the *flux* in the Lyα forest. To acquire the Lyα *transmission* (i.e., F/F_0), we need to know the intrinsic QSO continuum, F_0. At very low redshifts, the QSO continuum can be inferred from the upper envelope of the Lyα forest. For QSOs at the EoR redshifts, however, we need to predict the continuum based on the flux redward the Lyα peak because the forest itself has been heavily absorbed. One approach is to predict the continuum over the Lyα forest by fitting a powerlaw based on the continuum flux over 1270 Å$\lesssim \lambda \lesssim$ 1450 Å in the rest frame (e.g., Fan et al. 2006; Eilers et al. 2018). The power-law extrapolations can have a large scatter of $\sim 20\%$ over the Lyα forest. New fitting techniques based on principal component analysis (PCA), machine learning, etc., have been developed to provide more robust QSO continuum prediction (see Bosman et al. 2021, and references therein). Figure 1.2 shows an example of continuum fitting and prediction based on PCA.

Another important feature to consider is the QSO proximity zone in the observed QSO spectra (see Figure 1.2). In regions close to the QSO redshift, the IGM opacity is decreased due to the enhanced ionization dominated by the QSO itself. The proximity effect affects both Lyman-series and Lyman-continuum absorption See Section 4.4 for a detailed introduction and modeling of the enhanced ionization from the QSO. The proximity zone includes a wealth of information. For example, the size of proximity zone may indicate the

QSO lifetime (e.g., Eilers et al. 2017, 2020). We can also infer the density field around QSOs through the proximity zone (e.g., Chen et al. 2022).

In Figure 1.2, we can see pixels and continuous regions with zero flux in the Lyα forest. These are dark pixels (e.g., McGreer et al. 2011, 2015; Jin et al. 2023) and dark gaps (e.g., Fan et al. 2006; Becker et al. 2015; Zhu et al. 2021). Lyα absorption is saturated when the IGM is $\gtrsim 10^{-4} - 10^{-3}$ neutral. Dark gaps contain multi-scale information of the IGM opacity. See Chapter 2 for a detailed discussion. If the column density of neutral hydrogen of an absorber is large enough ($N_{\rm HI} > 2 \times 10^{20}$ cm^{-2}), we see extended absorption wings extending beyond the scale of this absorber. This kind of absorbers are Damped Lyman-α absorbers (DLAs), and the extended absorption wings are quantum-mechanical damping wings. Damping wings are very useful as indicators of neutral gas. For example, some QSOs at $z > 7$ show significant damping wing features over their Lyα emission lines. This indicates that the IGM likely to be considerably neutral at the QSO redshifts, unless the damping wing is caused by the dense circum-galactic medium (CGM) of an intervening system or the QSO hosting galaxy (e.g., Greig et al. 2017, 2022; Bañados et al. 2018; Davies et al. 2018b; Wang et al. 2020; Yang et al. 2020a). If we can find damping wing features in the Lyα forest, they would provide robust constraints on the IGM neutral fraction in un-biased locations (see Chapter 5).

1.2.2 The Lyβ forest

The Lyβ transition corresponds to an electron transitioning between the ground state and the second excited state, and has a rest-frame wavelength of $\lambda_\beta = 1025.72$ Å. Similar to the Lyα forest, when neutral hydrogen atoms absorb Lyβ photons, we can see

absorption lines blueward the Lyβ emission in the redshifted QSO spectra. These absorption lines form the Lyβ forest. The observed Lyβ forest at $z_{\mathrm{abs},\beta}$ is also attenuated by the Lyα forest in the foreground at a lower redshift, $z_{\mathrm{f},\alpha} = (1 + z_{\mathrm{abs},\beta})\lambda_\beta/\lambda_\alpha - 1$.

We can use Equation 1.6 to calculate the optical depth to Lyβ photons after replacing Lyα parameters to those of Lyβ. As mentioned in Section 1.2.1, Lyα is easily saturated even when the neutral fraction is low. Lyβ absorption, however, has a shorter wavelength and a lower oscillator strength ($f_\beta = 0.0791$) compared to those of Lyα absorption. The ratio of optical depth is given by $\tau_{\mathrm{Ly}\beta}/\tau_{\mathrm{Ly}\alpha} = (f_{\mathrm{Ly}\beta}\lambda_{\mathrm{Ly}\beta})/(f_{\mathrm{Ly}\alpha}\lambda_{\mathrm{Ly}\alpha}) \simeq 0.16$. Given its lower optical depth, Lyβ should be a more sensitive probe of neutral gas in the $z \lesssim 6$ IGM and we can use absorption troughs in the Lyβ forest to set more robust constraints on the IGM. For example, dark gaps in the Lyβ forest tend to indicate regions with a higher neutral hydrogen fraction than Lyα dark gaps. The fraction of QSO spectra showing Lyβ dark gaps or dark pixels can provide conservative upper-limit constraints on the volume-averaged neutral fraction $\langle x_{\mathrm{HI}} \rangle$ (e.g., Zhu et al. 2022; Jin et al. 2023; Chapter 3). In addition, the stacked Lyα forest at the redshift of red edges of long Lyβ dark gaps show a damping wing-like profile, suggesting that these dark gaps may trace genuine neutral hydrogen islands (Chapter 5).

1.2.3 Higher-order Lyman series and Lyman continuum transmission

Similar to the Lyα and Lyβ forests, neutral hydrogen creates absorption corresponding to the energies of higher-order Lyman transitions (between ground state and n-th excited state). The observed Lyn forest includes the attenuation form all foreground Lyα though Ly$(n-1)$ forest at the corresponding lower redshifts, respectively. The effective

optical depth of the foreground is calculated by

$$\tau_{\rm eff}^{\rm foreground}(\lambda_{\rm obs}) = \sum_{j=1}^{n-1} \tau_{\rm eff}^{{\rm Ly}j}(z_j)\,, \tag{1.9}$$

where $\tau_{\rm eff}^{j}(z_j)$ is the effective opacity of transition j at redshift z_j such that $(1+z_j)\lambda_j = \lambda_{\rm obs}$, and λ_j is the wavelength of transition j in the rest frame.

The ionization energy of hydrogen is 13.6 eV, corresponding to a wavelength of 912 Å. [1] The Lyman continuum (LyC) part in high-z QSO spectra corresponds to transmission with a shorter wavelength than 912 Å. Following (Becker et al. 2021), the observed flux, $f_\lambda^{\rm obs}$ in the LyC part, will be the mean intrinsic QSO continuum, $f_\lambda^{\rm cont}$, attenuated by the effective Lyman series opacity of the foreground IGM, $\tau_{\rm eff}^{\rm Lyman}$, and the Lyman continuum effective optical depth, $\tau_{\rm eff}^{\rm LyC}$,

$$f_\lambda^{\rm obs} = f_\lambda^{\rm cont} \exp\left(-\tau_{\rm eff}^{\rm Lyman}\right) \exp\left(-\tau_{\rm eff}^{\rm LyC}\right). \tag{1.10}$$

This region is crucial for studying the opacity to ionizing photons during the epoch of reionization. One important application is to directly measure the mean free path ($\lambda_{\rm mfp}$), the evolution of which can put robust constraints on when and how reionization ends (Chapter 4; D'Aloisio et al. 2020; Becker et al. 2021; Zhu et al. 2023 and references therein).

1.3 This work

As mentioned in Section 1.1, an end of reionization at $z \geq 6$ is challenging to reconcile with a midpoint of $z \sim 7-8$ in terms of the ionizing photon budget. In particular, star-forming galaxies at $z > 6$ would have to emit ionizing photons extremely efficiently in

[1] We use 912 Å to represent the Lyman limit wavelength of 911.76 Å.

order to complete reionization within such a short interval. Such a tension, however, would be relaxed with reionization ending by $z \sim 5.5$. A key question, therefore, is **whether and for how long reionization and/or its impacts on the IGM continued below $z = 6$.**

This dissertation focuses on constraining the IGM during the later stages of reionization using hydrogen absorption lines in high-z QSO spectra. Chapter 2 presents the search of long dark gaps in the Lyα forest toward high-z QSOs. Our sample includes 55 high-quality spectra of QSOs and the goal is to search for dark gaps ($> 95\%$ Lyα photons absorbed) at $z < 6$. Dark gap measurements can complement $\tau_{\rm eff}$ measurements by providing multi-scale information of the IGM. The chapter will show our discovery of long ($L \geq 30\ h^{-1}\,$Mpc) dark gaps extending down to $z \simeq 5.4$, and discuss the implications for reionization models by comparing the dark gap statistics.

Chapter 3 includes the first search of Lyβ dark gaps at $z > 5$. While our observed long Lyα dark gaps disfavor a fully ionized IGM with a homogeneous UVB by $z = 6$, it is still unclear whether the observed distribution of Lyα dark gaps requires large neutral regions at $z < 6$ (i.e., late reionization), or if they can be produced purely by UV background fluctuations. Because of its smaller oscillator strength and shorter wavelength, Lyβ is not as easily saturated as Lyα. In the Lyβ forest, pure UVB fluctuations are then less efficient in creating long absorption troughs (Nasir & D'Aloisio 2020). Therefore, pushing the dark gap statistics to the Lyβ forest in QSO spectra can help probe genuine neutral regions.

Chapter 4 presents the direct measurements of the ionizing photon mean free path ($\lambda_{\rm mfp}$) over $5 < z < 6$. The mean free path measures how far on average ionizing photons

can travel freely before being absorbed. Near the end of reionization, $\lambda_{\rm mfp}$ is expected to evolve rapidly due to i) the overlapping of ionized bubbles, and ii) the photoevaporation of ionization sinks (Furlanetto & Oh 2005; Sobacchi & Mesinger 2014; D'Aloisio et al. 2020). Tracing the evolution of $\lambda_{\rm mfp}$ can therefore give strong insights into the end of reionization. This chapter will detail how we model the LyC transmission, and show our measured $\lambda_{\rm mfp}$ evolution based on stacked QSO spectra.

In Chapter 5, we revisit Lyβ dark gaps to search for signature of remaining neutral hydrogen islands at $z < 6$, given the evidence of ultra-late reionization in Chapters 2-4. We report a damping wing-like absorption profile in the stacked Lyα forest at the redshift of the red edge of Lyβ dark gaps with $L \geq 7\ h^{-1}\,$Mpc. After compared with mock spectra from simulations, such an extended absorption profile provides a robust lower-limit constraint on $\langle x_{\rm HI} \rangle$ at $z \simeq 5.8$. Finally, Chapter 6 summarizes the implications for reionization.

Chapter 2

Chasing the Tail of Cosmic Reionization with Dark Gap Statistics in the Lyα Forest over $5 < z < 6$

2.1 Abstract

We present a new investigation of the intergalactic medium (IGM) near the end of reionization using "dark gaps" in the Lyman-alpha (Lyα) forest. Using spectra of 55 QSOs at $z_{\rm em} > 5.5$, including new data from the XQR-30 VLT Large Programme, we identify gaps in the Lyα forest where the transmission averaged over 1 comoving h^{-1} Mpc bins falls below 5%. Nine ultra-long ($L > 80\ h^{-1}$ Mpc) dark gaps are identified at $z < 6$. In

addition, we quantify the fraction of QSO spectra exhibiting gaps longer than 30 h^{-1} Mpc, F_{30}, as a function of redshift. We measure $F_{30} \simeq 0.9$, 0.6, and 0.15 at $z = 6.0$, 5.8, and 5.6, respectively, with the last of these long dark gaps persisting down to $z \simeq 5.3$. Comparing our results with predictions from hydrodynamical simulations, we find that the data are consistent with models wherein reionization extends significantly below redshift six. Models wherein the IGM is essentially fully reionized that retain large-scale fluctuations in the ionizing UV background at $z \lesssim 6$ are also potentially consistent with the data. Overall, our results suggest that signature of reionization in the form of islands of neutral hydrogen and/or large-scale fluctuations in the ionizing background remain present in the IGM until at least $z \simeq 5.3$. [1]

2.2 Introduction

The reionization of the intergalactic medium (IGM) is the last major phase transition in the history of the Universe. In the widely accepted picture, neutral hydrogen in the IGM was reionized by ultraviolet photons emitted by the first luminous sources (e.g., Bromm & Larson 2004; McQuinn 2016; Dayal & Ferrara 2018). Determining when reionization occurred as well as what sources were responsible is therefore important for the understanding of formation and evolution of the first stars, galaxies, and black holes.

Multiple observations now constrain the timing of reionization. Cosmic microwave background (CMB) measurements suggest a midpoint at redshift $z_{\rm re} = 7.7 \pm 0.7$ (Planck Collaboration et al. 2020; see also de Belsunce et al. 2021). The redshift evolution in the

[1] A version of this chapter is published as Zhu et al. (2021) — *Chasing the Tail of Cosmic Reionization with Dark Gap Statistics in the Lyα Forest over $5 < z < 6$, The Astrophysical Journal, 923:223.*

fraction of UV-selected galaxies detected in Lyman-alpha (Lyα) emission also suggests that the IGM was significantly neutral near $z \sim 7$–8 (e.g., Mason et al. 2018, 2019; Hoag et al. 2019; Hu et al. 2019, and references therein, but see Wold et al. 2021). These results are broadly consistent with multiple probes of the IGM using QSO spectra. For example, the IGM thermal history at $z > 4$ inferred from the Lyα flux power spectrum suggests a mean redshift of reionization near $z_{\rm re} \simeq 8.5^{+1.1}_{-0.8}$ (Boera et al. 2019, see also Gaikwad et al. 2021; Walther et al. 2019). Similarly, Lyα damping wing measurements of $z_{\rm em} > 7$ QSOs indicate that the IGM was significantly neutral at $z \sim 7$–7.5 (e.g., Bañados et al. 2018; Davies et al. 2018b; Greig et al. 2017, 2019; Wang et al. 2020; Yang et al. 2020a). The appearance of transmitted flux in the Lyα and Lyβ forests suggests that the reionization mostly completed by $z \simeq 6$ (e.g., McGreer et al. 2015). On the other hand, large fluctuations in the observed IGM effective opacity ($\tau_{\rm eff} = -\ln \langle F \rangle$, where F is the continuum-normalized flux) in the Lyα forest at $z < 6$ suggest that signatures of reionization may persist in the IGM down to even lower redshifts (Fan et al. 2006; Becker et al. 2015; Bosman et al. 2018; Eilers et al. 2018; Yang et al. 2020b; Bosman et al. 2022).

Multiple models have been proposed to explain the large-scale fluctuations in IGM Lyα opacity at $z < 6$. If the IGM is mostly ionized at $z \geq 6$, then large variations in Lyα opacity may persist to lower redshifts due to either lingering temperature fluctuations produced by inhomogeneous reionization (D'Aloisio et al. 2015), or fluctuations in the ionizing UV background produced by a short and spatially variable mean free path (Davies & Furlanetto 2016; Nasir & D'Aloisio 2020), or rare sources such as QSOs (Chardin et al. 2017; see also Meiksin 2020). Alternatively, if reionization continues substantially to $z < 6$ then

the observed scatter in $\tau_{\rm eff}$ could be due to the presence of large patches of neutral gas coupled with UVB fluctuations (e.g., Kulkarni et al. 2019a; Keating et al. 2020a; Nasir & D'Aloisio 2020; Qin et al. 2021). The combination of neutral patches and UVB fluctuations may naturally explain the presence of giant Lyα troughs such as the 110 h^{-1} Mpc trough towards ULAS J0148+0600 identified by Becker et al. (2015) (e.g., Keating et al. 2020b). A late-ending reionization[2] scenario is also consistent with the evolution of O I absorbers at $z \simeq 6$ (Becker et al. 2019), and is preferred by recent Bayesian inference results that simultaneously match Lyα forest, CMB, and galaxy data (Choudhury et al. 2021; Qin et al. 2021).

Some progress has been made towards distinguishing between these models observationally. Measurements of Lyα emitting galaxies (LAEs, Becker et al. 2018, Christenson et al., in prep) and Lyman break galaxies (LBGs, Kashino et al. 2020) in the field of ULAS J0148+0600 have demonstrated that the Lyα trough along this line of sight is associated with a large-scale underdensity. This result disfavors the temperature fluctuation model, but is consistent with either the pure UVB fluctuation or late-reionization model. The QSO UVB model is potentially also consistent with this result, though it is disfavored by measurements of the QSO luminosity function near $z \sim 6$ (e.g., Parsa et al. 2018; Kulkarni et al. 2019b). Recent measurements of the mean free path of ionizing photons over $5 < z < 6$ are consistent with late reionization models wherein the IGM is still significantly neutral at $z = 6$, and disfavor models in which reionization ends early enough that the IGM relaxes hydrodynamically by $z = 6$ (Becker et al. 2021). Simultaneously matching the mean free

[2]Throughout this paper, for convenience, we refer to "the end of reionization" as when the volume filling factor of ionized gas in the IGM reaches 99%. We use "early" for scenarios wherein reionization ends at $z \geq 6$, and "late" for cases where reionization ends below $z = 6$.

path measurements and other IGM constraints further favors a late and rapid reionization scenario driven by galaxies that are efficient at producing and/or emitting ionizing photons (Cain et al. 2021; Davies et al. 2021).

A key question, therefore, is whether and for how long the impacts of reionization continued below $z = 6$. It is also of interest to determine whether a late reionization scenario with islands of neutral gas and UVB fluctuations at $z < 6$ can be distinguished observationally from an early reionization scenario with UVB fluctuations alone. Better measurements of the spatial scale of the Lyα opacity fluctuations may provide some insights. Determining how long these fluctuations persist and how they evolve with redshift may also be helpful. One way to do this is by identifying individual "dark gaps" in the Lyα forest (e.g. Songaila & Cowie 2002; Furlanetto et al. 2004; Paschos & Norman 2005; Fan et al. 2006; Gallerani et al. 2008), which could be created by regions of neutral IGM and/or low UV background (e.g., Nasir & D'Aloisio 2020). Because dark gaps contain multi-scale spatial information, they provide complementary information to $\tau_{\rm eff}$ measurements averaged over intervals of fixed length, and may therefore be useful for distinguishing between models of the IGM at $z < 6$.

In this paper, we use dark gap statistics to characterize the opacity of the IGM over $5 \lesssim z \lesssim 6$. In particular, we use these statistics to determine how long large Lyα-opaque regions persist in the IGM, and whether the data are consistent with existing late reionization models and/or early reionization models that retain a fluctuating UVB. We use a sample of 55 high signal-to-noise (S/N) spectra of QSOs at $5.5 \lesssim z_{\rm em} \lesssim 6.5$, including 23 new X-Shooter spectra from the XQR-30 VLT Large Programme (D'Odorico et al., in

prep). In addition to the distribution of dark gap lengths, we measure the fraction of QSO spectra exhibiting long ($L \geq 30\ h^{-1}\,\mathrm{Mpc}$) dark gaps as a function of redshift for the first time.

We present our data in Section 2.3. In Section 2.4 we describe our methods for measuring dark gaps and the results of dark gap statistics. Section 2.5 introduces the models to which we compare our measurements. We then discuss the implications for the models in Section 2.6. Finally, we summarize our conclusions in Section 2.7. Throughout this paper we quote distances in comoving units unless otherwise noted, and assume a ΛCDM cosmology with $h = 0.678$, $\Omega_m = 0.308$ and $\Omega_\Lambda = 0.692$.

2.3 The data

2.3.1 QSO spectra

This study is based on spectra of 55 QSOs at $5.5 \lesssim z_{\rm em} \lesssim 6.5$ taken with the X-Shooter spectrograph on the Very Large Telescope (VLT; Vernet et al. 2011) and the Echellette Spectrograph and Imager (ESI) on Keck (Sheinis et al. 2002). Of these, 23 X-Shooter spectra are from the XQR-30 VLT Large Programme. The XQR-30 program is targeting 30 bright QSOs at $5.8 \lesssim z \lesssim 6.6$ for the study of reionization and other aspects of the early Universe. The full data set will be described in D'Odorico et al., in prep. The 23 objects out of the XQR-30 sample selected for this project are those that meet our S/N threshold and do not contain strong BAL features. In addition, we use 30 spectra reduced from archival X-Shooter and ESI data, of which 27 are from the sample of Becker et al. (2019). Recent deep (20 hour) X-Shooter observations (PI: Fuyan Bian) of the lensed

$z = 6.5$ QSO J0439+1634 are also included in the dark gaps statistics. Finally, we acquired a deep (7 hour) ESI spectrum of SDSS J1250+3130. Observations for all objects except SDSS J1250+3130 were taken without any foreknowledge of dark gaps in the Lyα forest. In the case of SDSS J1250+3130, we targeted the QSO based on indications from a shallower (1 hour) ESI spectrum that its spectrum contained a long dark gap in the Lyα forest. We discuss the impact of including this object on our results in Section 2.4.4.

Details of the data reduction are given in Becker et al. (2019). Briefly, we used a custom pipeline that includes optimal techniques for sky subtraction (Kelson 2003) and one-dimensional spectral extraction (Horne 1986). Telluric absorption corrections were computed for individual exposures using models based on the Cerro Paranal Advanced Sky Model (Noll et al. 2012; Jones et al. 2013). The spectra were extracted using 10 km s^{-1} pixels for the VIS arm of X-Shooter and 15 km s^{-1} pixels for ESI. Typical resolutions for the X-Shooter and ESI are FWHM $\approx$ 25 km s^{-1} and 45 km s^{-1}, respectively. In addition, for J0439+1634, to reduce the contamination from continuum emission of a foreground galaxy, we fit a power law of the flux zero point over the Lyα forest and subtract it from the flux. The spectra are plotted in Figure Set 2.1.

Table 2.1: QSO spectra used in this work

No.	QSO	$z_{\rm em}^{\rm [Ref.]}$	Source	Instrument	S/N
1	J2207-0416	5.529[b]	archival (B19)	X-Shooter	42
2	J0108+0711	5.577[b]	archival (B19)	X-Shooter	29
3	J1335-0328	5.693[b]	archival (B19)	X-Shooter	30

Continued on next page

Table 2.1 – *Continued from previous page*

No.	QSO	$z_{\rm em}^{\rm [Ref.]}$	Source	Instrument	S/N
4	SDSSJ0927+2001	5.7722[c]	archival (B19)	X-Shooter	76
5	SDSSJ1044-0125	5.7847[o]	other archival	ESI	71
6	PSOJ065+01	5.790[q]	XQR-30	X-Shooter	47
7	PSOJ308-27	5.794[q]	XQR-30	X-Shooter	58
8	SDSSJ0836+0054	5.810[g]	other archival	ESI	152
9	PSOJ004+17	5.8165[e]	other archival	X-Shooter	21
10	SDSSJ0002+2550	5.820[b]	archival (B19)	ESI	93
11	PSOJ242-12	5.834[q]	XQR-30	X-Shooter	28
12	SDSSJ0840+5624	5.8441[n]	archival (B19)	ESI	41
13	SDSSJ0005-0006	5.847[b]	archival (B19)	ESI	24
14	PSOJ025-11	5.849[q]	XQR-30	X-Shooter	53
15	PSOJ183-12	5.857[q]	XQR-30	X-Shooter	66
16	SDSSJ1411+1217	5.904[g]	archival (B19)	ESI	46
17	PSOJ108+08	5.950[q]	XQR-30	X-Shooter	70
18	PSOJ056-16	5.9670[e]	archival (B19)	X-Shooter	35
19	PSOJ029-29	5.981[q]	XQR-30	X-Shooter	51
20	SDSSJ0818+1722	5.997[b]	archival (B19)	X-Shooter	108
21	ULASJ0148+0600	5.998[b]	archival (B19)	X-Shooter	126
22	PSOJ340-18	5.999[b]	archival (B19)	X-Shooter	32
23	PSOJ007+04	6.0008[d]	XQR-30	X-Shooter	53
24	SDSSJ2310+1855	6.0031[o]	XQR-30	X-Shooter	81
25	SDSSJ1137+3549	6.007[j]	archival (B19)	ESI	28

Continued on next page

Table 2.1 – *Continued from previous page*

No.	QSO	$z_{\rm em}^{\rm [Ref.]}$	Source	Instrument	S/N
26	ATLASJ029.9915-36.5658	6.021[b]	XQR-30	X-Shooter	48
27	SDSSJ1306+0356	6.0330[k]	archival (B19)	X-Shooter	71
28	J0408-5632	6.035[q]	XQR-30	X-Shooter	71
29	ULASJ1207+0630	6.0366[d]	archival (B19)	X-Shooter	25
30	SDSSJ2054-0005	6.0391[o]	archival (B19)	ESI	29
31	PSOJ158-14	6.0681[e]	XQR-30	X-Shooter	59
32	SDSSJ0842+1218	6.0763[d]	XQR-30	X-Shooter	71
33	SDSSJ1602+4228	6.079[j]	archival (B19)	ESI	34
34	PSOJ239-07	6.1098[e]	XQR-30	X-Shooter	65
35	CFHQSJ1509-1749	6.1225[d]	archival (B19)	X-Shooter	54
36	SDSSJ2315-0023	6.124[b]	archival (B19)	ESI	25
37	ULASJ1319+0950	6.1330[o]	archival (B19)	X-Shooter	86
38	SDSSJ1250+3130	6.137[j]	new observation	ESI	53
39	VIKJ2318-3029	6.1458[d]	archival (B19)	X-Shooter	21
40	PSOJ217-16	6.1498[d]	XQR-30	X-Shooter	68
41	PSOJ217-07	6.165[q]	XQR-30	X-Shooter	42
42	PSOJ359-06	6.1718[e]	XQR-30	X-Shooter	67
43	PSOJ060+24	6.177[q]	XQR-30	X-Shooter	53
44	PSOJ065-26	6.1877[d]	XQR-30	X-Shooter	73
45	PSOJ308-21	6.2341[d]	archival (B19)	X-Shooter	26
46	SDSSJ1030+0524	6.309[f]	archival (B19)	X-Shooter	35
47	SDSSJ0100+2802	6.3270[l]	archival (B19)	X-Shooter	212

Continued on next page

Table 2.1 – *Continued from previous page*

No.	QSO	$z_{\rm em}^{\rm [Ref.]}$	Source	Instrument	S/N
48	ATLASJ025.6821-33.4627	6.3373[k]	archival (B19)	X-Shooter	61
49	J1535+1943	6.381[q]	XQR-30	X-Shooter	30
50	SDSSJ1148+5251	6.4189[h]	archival (B19)	ESI	64
51	J1212+0505	6.4386[d]	XQR-30	X-Shooter	41
52	J0439+1634	6.5188[p]	new observation	X-Shooter	224
53	VDESJ0224-4711	6.5223[m]	XQR-30	X-Shooter	29
54	PSOJ036+03	6.541[a]	archival (B19)	X-Shooter	38
55	PSOJ323+12	6.5881[i]	XQR-30	X-Shooter	30

Table Comments: Columns: (1) QSO index number, (2) QSO name, (3) QSO redshift with reference, (4) source of the spectrum used for dark gap statistics, (5) instrument used for dark gap statistics, (6) continuum signal-to-noise ratio per 30 km s^{-1} near rest wavelength 1285 Å.
Sources of the spectra: XQR-30: spectra from the XQR-30 program; new observation: spectra from new observations; archival (B19): archival spectra used and reduced in Becker et al. (2019); other archival: spectra from the public archives but not included in Becker et al. (2019).
References: Redshift lines and references. a. [C II] 158μm: Bañados et al. (2015); b. apparent start of the Lyα forest: Becker et al. (2019); c. CO: Carilli et al. (2007); d. [C II] 158μm: Decarli et al. (2018); e. [C II] 158μm: Eilers et al. (2020); f. Mg II: Jiang et al. (2007); g. Mg II: Kurk et al. (2007); h. [C II] 158μm: Maiolino et al. (2005); i. [C II] 158μm: Mazzucchelli et al. (2017); j. Mg II: Shen et al. (2019); k. [C II] 158μm: Venemans et al. (2020); l. [C II] 158μm: Wang et al. (2019); m. [C II] 158μm: Wang et al. (2021); n. CO: Wang et al. (2010); o. [C II] 158μm: Wang et al. (2013); p. [C II] 158μm: Yang et al. (2019a); q. apparent start of the Lyα forest: this work.

We adopt QSO redshifts measured from CO, [C II] 158μm or Mg II lines if available. Otherwise we use redshifts inferred from the apparent start of the Lyα forest, following Becker et al. (2019). Table 2.1 summarizes QSO spectra used in this work with the QSO redshifts, instruments, and estimated signal-to-noise ratios, which is calculated as the me-

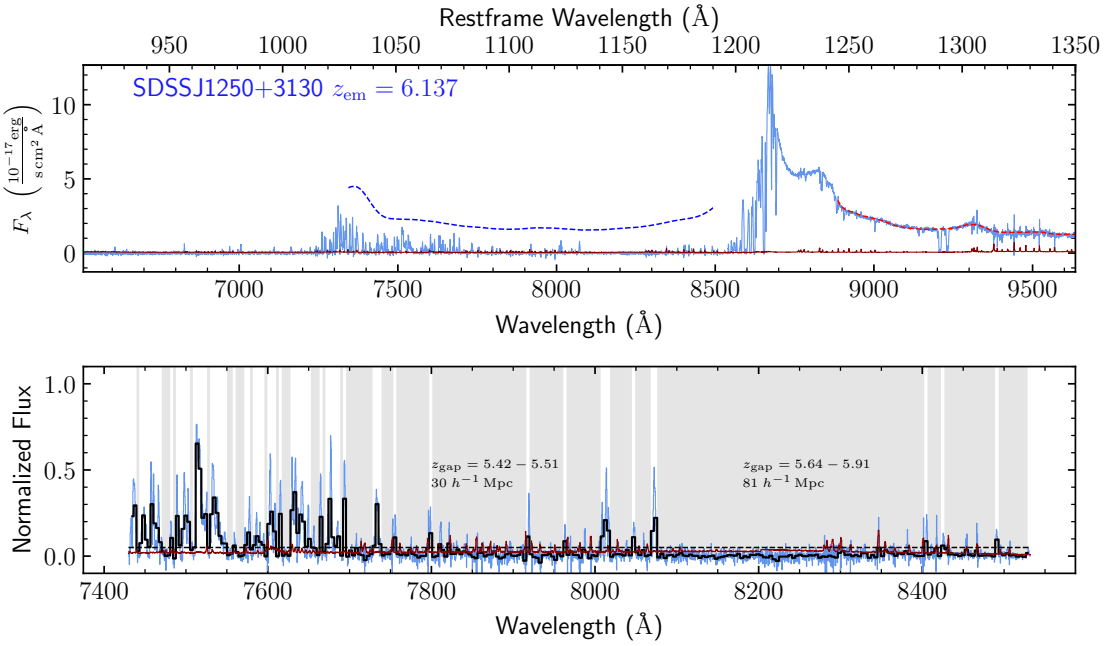


Figure 2.1: Spectrum, continuum fits and dark gap detection details of the $z_{\rm em} = 6.137$ quasar SDSS J1250+3130. **Top**: QSO spectrum and continuum fits. The light blue and dark red lines represents flux and flux error in the original binning. Dashed red and blue curves are the best-fitting QSO continuum based on PCA. **Bottom**: Lyα forest and dark gaps detected. The dashed black line labels the flux threshold of 0.05. The thick black line displays the flux binned to 1 h^{-1} Mpc. Light blue and dark red lines show the flux and flux error in the original binning. Dark gaps detected are shaded with gray. We also label the redshift range and length of each long dark gap ($L \geq 30\ h^{-1}$ Mpc), if any.
(The complete figure set (55 images) is available in the online journal and at https://ydzhuastro.github.io/Zhu21.html.)

dian ratio of unabsorbed QSO continuum to noise per 30 km s^{-1} near 1285 Å in the rest frame.

2.3.2 Continuum Fitting

The detection of dark gaps relies on the construction of the intrinsic continuum over the Lyα forest. In order to estimate QSO continua blueward of the Lyα emission line, we use Principal Component Analysis (PCA), which is less biased than the conventional power law fitting (e.g., Bosman et al. 2021). In this work, we apply the log-PCA method of Davies et al. (2018c) as implemented in the Lyα forest portion of the spectrum by Bosman et al. (2022), with 15 and 10 components used for the red-side (rest-frame wavelength $\lambda_0 > 1230$ Å) and blue-side ($\lambda_0 < 1170$ Å) continuum. For each QSO, we fit the red-side continuum with principal components, and map the corresponding red principal component coefficients to the blue side coefficients with a projection matrix. For X-Shooter spectra with observations from the NIR arm, we fit the red continuum over $1230 < \lambda < 2000$ Å in the rest frame.

The ESI spectra are fit using an optical-only PCA, which is presented in Bosman et al. (2021). QSOs with strong broad absorption lines (BALs) in their spectra were excluded from our sample. For QSOs with mild absorption features that interfere minimally with the Lyα forest, we mask out the absorption lines when fitting their spectra. In addition, we intentionally leave out the Lyα emission peak and the proximity zone when fitting and predicting the continuum on account of the large object-to-object variations in these regions. The typical 1σ uncertainty of the PCA continuum fitting over the Lyα forest is less than 10%. Continuum fits and blue-side predictions are shown in Figure Set 2.1 along with the

QSO spectra. We also verify that our dark gap statistics results do not significant change if we use power-law continua (see Appendix C), which have a typical bias of ∼10% over the Lyα forest (Bosman et al. 2021).

2.4 Dark gap statistics

2.4.1 Method

We define a dark gap to be a continuous spectral region in which all pixels binned to 1 h^{-1} Mpc have an observed normalized flux $F = F_{\rm obs}/F_{\rm c} < 0.05$, where $F_{\rm obs}$ is the observed flux and $F_{\rm c}$ is the continuum flux. The minimum length of a dark gap is 1 h^{-1} Mpc. We apply this definition when searching for dark gaps in both the real data and the mock spectra. The bin size and flux threshold were chosen to enable a uniform analysis over our large sample of spectra. A bin size of 1 h^{-1} Mpc (corresponding to a velocity interval of $\Delta v \simeq 150$ km s^{-1} at $z = 5.6$) provides a convenient scale that preserves most of the structure of the Lyα forest. The choice of the flux threshold $F_{\rm t}$ is mainly restricted by the quality of the data. Our choice of $F_{\rm t} = 0.05$ corresponds to non-detection of transmission lower than approximately twice the binned flux error (2σ) in the spectrum with the lowest S/N in our sample. Using such a threshold, all dark gaps longer than 30 h^{-1} Mpc have $\tau_{\rm eff} > 4$. [3] We have tested that using 0.1 or 0.025 for the flux threshold does not change our conclusions fundamentally when applying the same criteria to both the observed and

[3] Throughout this paper, $\tau_{\rm eff}$ of a dark gap is calculated based on flux averaged along the full length of the gap rather than over windows of a fixed length. Most low $\tau_{\rm eff}$ values for short dark gaps are caused by skyline subtraction or telluric correction residuals.

mock spectra. Setting $F_{\rm t} = 0.1$ tends to yield dark gaps that are less opaque, while setting $F_{\rm t} = 0.025$ would decrease the number of usable QSO sightlines from 55 to 37.

In order to avoid the QSO proximity region, we identify dark gaps in the Lyα forest starting from 7 proper Mpc (pMpc) blueward from the QSO, which is close to the size of the largest proximity zones of bright QSOs at these redshifts (Eilers et al. 2017, 2020). On the blue end, we limit our search to greater than 1041 Å in the rest frame in order to avoid contamination from associated Lyβ or O VI absorption (e.g., Becker et al. 2015). For the purpose of comparing our results to simulations, we wish to avoid dark gaps that may be truncated by transmission peaks within the proximity zone. When quantifying the fraction of lines of sight that intersect gaps of length $L \geq 30\ h^{-1}\,$Mpc (Sections 2.4.4 and 2.4.5), the highest redshift at which we register an individual sightline that shows a long gap, *if any*, is therefore 30 $h^{-1}\,$Mpc blueward of our proximity zone cut, although the gap may include pixels that extend up to the proximity zone. Nevertheless, we still record the full lengths of gaps extending to this 30 $h^{-1}\,$Mpc "buffer zone" when searching for the longest possible dark gaps in both data and simulations. Dark gaps completely located in the QSO proximity zone and/or in this "buffer zone", however, are discarded. This ensures that the pixel at the red end of each sightline *may* intersect a long ($L \geq 30\ h^{-1}\,$Mpc) dark gap.[4] Finally, we limit our analysis to $z < 6$ because the mean transmitted flux at $z > 6$ is so low that most spectra show long dark gaps, making the dark gap statistics less informative.

We note that there is no perfect way to handle the proximity zone effect. It is difficult to precisely define and measure the proximity zone size for each QSO, which partly

[4]If we do not introduce this "buffer zone", there is a possibility that the F_{30} (Section 2.4.4) is underestimated near the red end of a sightline, since there can exist otherwise $> 30\ h^{-1}\,$Mpc gaps that are truncated by the edge or peaks in the proximity zone.

motivates our choice to use a fixed proximity zone cut. The proximity zone for the brightest QSOs in our sample (e.g., SDSS J0100+2802 and VDES J0224+4711) may be larger than our adopted cut of 7 pMpc. Fortunately, the use of an additional 30 h^{-1} Mpc buffer zone minimizes the potential effect of the larger proximity zone of these objects. In addition, because we limit our statistics over $5 < z < 6$, proximity zone transmission at $z > 6$ towards some extremely bright QSOs does not impact our results. Still, one should treat dark gaps near the QSO proximity zone with caution.

Noisy residuals from skyline subtraction and telluric correction may divide an otherwise continuous region of depressed flux. To deal with this, when searching for dark gaps we mask out $\pm 75\,\mathrm{km\,s^{-1}}$ intervals of the spectra centered at peaks in the flux error array, which typically correspond to skyline residuals. The exception to this is that we do not mask out any pixels with $F > 3\sigma_F$. For consistency, we apply the same masking procedure to the mock spectra. [5] In Appendix B, we use the mock spectra to show that such masking only produces a minor change in the results. We also test that the impact of masking telluric correction residuals near 7600-7650 Å is neglectable.

As for contamination from damped Lyα systems (DLAs) or other metal-enriched absorbers, we made no correction for their effect on dark gap detection following, e.g., Fan et al. (2006). Even strong DLAs can hardly, on their own, produce dark gaps as long as 30 h^{-1} Mpc, which are the primary focus of this work. Nevertheless, in the results we label dark gaps with intervening metal systems for reference based on the systems identified by Chen et al. (2017) and Becker et al. (2019), as well as our own inspections. We visually

[5] Since we add noise to the mock spectra pixel-wise according to the noise array of each observed spectrum with a Gaussian distribution, the skyline residuals in the mock spectra are not actually modeled. However, masking $\pm 75\,\mathrm{km\,s^{-1}}$ intervals makes the profile of sky subtraction residuals unimportant.

searched all $L \geq 30\ h^{-1}\,\mathrm{Mpc}$ dark gaps for metal absorbers not listed in the literature. The systems were identified via the coincidence of multiple metal lines in redshift. The metal lines we used include C II $\lambda1334$, C IV $\lambda\lambda1548, 1550$, O I $\lambda1302$, Mg II $\lambda\lambda2797, 2803$, Al II $\lambda1670$, Si II $\lambda1527$, and Si IV $\lambda\lambda1394, 1403$. A detection required these metal lines (if available) to have significant absorption features and self-consistent velocity profiles at the same redshift. We have a good wavelength coverage for most metal lines mentioned above in QSO spectra taken with X-Shooter. Even for these objects, however, we caution that the list of metal absorbers may be still incomplete. A full list of metal absorbers in the XQR-30 spectra will be presented by R. Davies et al., in prep. We also note that the simulations we used do not include DLAs or other metal-enriched absorbers.

2.4.2 Notable dark gaps

Long dark gaps play an important role in characterizing the IGM in the later stages of reionization. Among 50 dark gaps with $L \geq 30\ h^{-1}\,\mathrm{Mpc}$ detected in our sample, Figure 2.2 displays some notable examples. They either extend down to or below $z \sim 5.5$, are extremely long ($L > 80\ h^{-1}\,\mathrm{Mpc}$), or both.

Two long dark gaps entirely at $z < 5.5$ are identified towards PSO J183-12 and PSOJ340-18. They span $z_{\mathrm{gap}} = 5.36 - 5.47$ and $z_{\mathrm{gap}} = 5.31 - 5.42$, corresponding to lengths of $L = 37\ h^{-1}\,\mathrm{Mpc}$ and $L = 34\ h^{-1}\,\mathrm{Mpc}$, respectively. Most spikes and sharp dips with negative flux in the un-binned spectra inside the two gaps are skyline subtraction residuals as indicated by the peaks in the flux error array. Both dark gaps are highly opaque, with $\tau_{\mathrm{eff}} > 6$. The spectra of both QSOs have a good coverage of redshifted common metal lines.

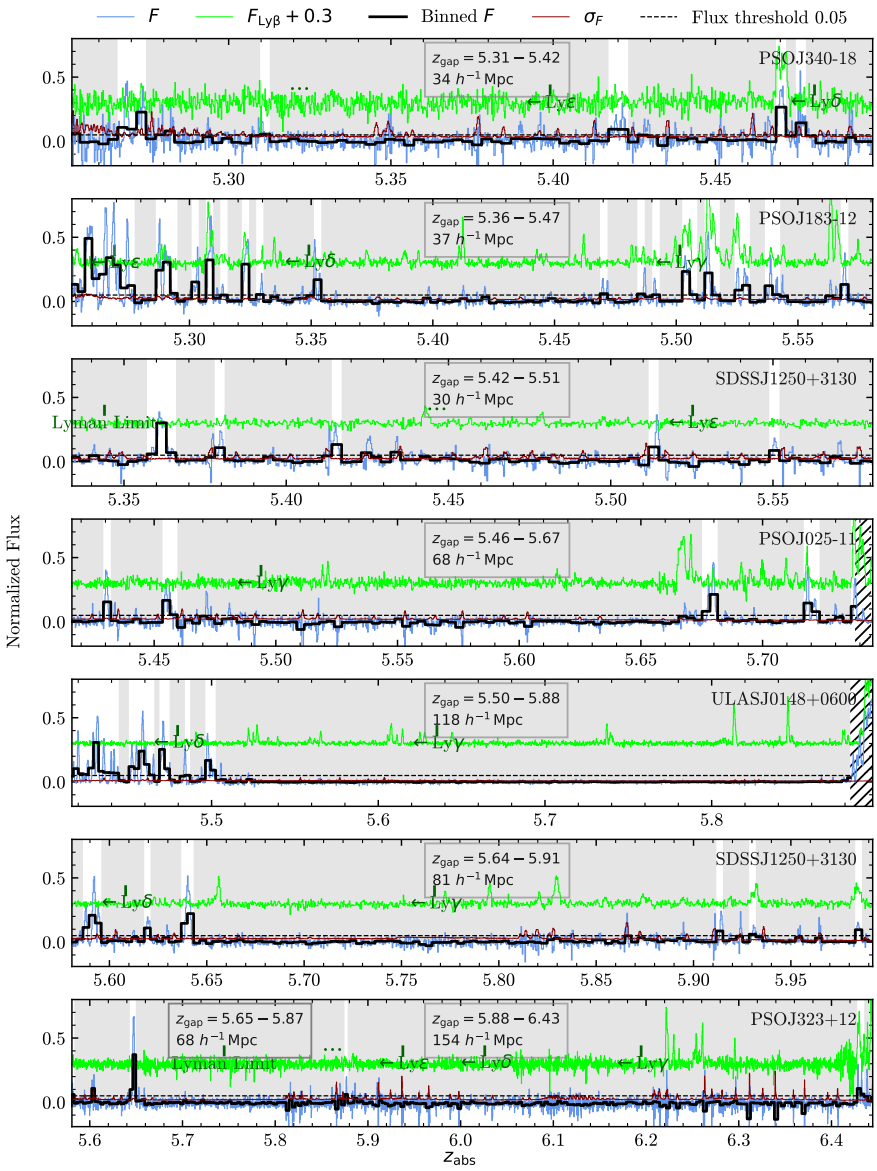


Figure 2.2: Examples of notable dark gaps identified. Dark gaps are labeled with gray shades. The thick black line is the binned flux with binning size of 1 h^{-1} Mpc. Dashed black horizontal line sets the flux threshold of 0.05. Un-binned flux and flux uncertainty are represented by thin blue and dark red lines, respectively. The text boxes display the redshift span and length of each long dark gap ($L \geq 30\ h^{-1}$ Mpc). Regions redward the proximity zone cut are labeled with hatches and are excluded from the statistics. For reference, the green line, offset by 0.3 in flux, is shifted in wavelength to show the Lyβ and higher-order Lyman forest at the same redshifts. Vertical ticks label the starting points of higher-order Lyman forests.

We searched their X-Shooter VIS and NIR spectra and found no metal absorption within the redshift ranges of the dark gaps. In addition, a 30 h^{-1} Mpc dark gap extending just above $z = 5.5$ is identified towards SDSS J1250+3130. Most of the spikes inside this gap are also probably due to sky lines as indicated by peaks in the flux error array.

The fourth through sixth rows in Figure 2.2 display three examples of long dark gaps extending down to $z \sim 5.5$. The long gap extending to $z = 5.46$ with a length of $L = 68$ h^{-1} Mpc towards PSOJ025-11 is one of the longest troughs below redshift six discovered in this work. The only weak transmission peaks in the un-binned flux array that seem to be real are the ones at $z_{\rm abs} \simeq 5.47$, 5.48, and 5.67. Overall, however, it is extremely dark, with $\tau_{\rm eff} \geq 6.4$. We also reproduce the detection of the long trough discovered towards ULAS J0148+0600 by Becker et al. (2015), which extends down to $z = 5.5$ with a total length $L > 110$ h^{-1} Mpc. Due to the use of a different definition of dark gap compared to Becker et al. (2015), the trough detected in this work includes an additional small transmission peak that appears in the un-binned spectrum near the blue end. This yields a slightly larger L but a comparable $\tau_{\rm eff}$ value. We also find a gap of $L = 81$ h^{-1} Mpc extending down to $z = 5.64$ towards SDSS J1250+3130. Spikes within the trough are skyline subtraction residuals, as shown by peaks in the error array. We do not see any strong metal absorbers that would indicate dense absorption systems such as DLAs or Lyman limit systems (LLSs), in any of these gaps. Finally, we find dark gaps longer than 110 h^{-1} Mpc towards several QSOs with the highest redshifts in our sample. This is not surprising because the IGM is more neutral at higher redshifts and therefore more likely to produce large Lyα opaque regions. For example, Barnett et al. (2017) identified a

240 h^{-1} Mpc gap at $z > 6.1$ towards the $z = 7.1$ QSO ULAS J1120+0641. Here we display a remarkably long dark gap towards PSO J323+12. It covers $z_{\rm gap} = 5.88$–6.43 and has a length of 154 h^{-1} Mpc, as shown in the bottom row of Figure 2.2.

For reference, we overplot in green the regions of spectra corresponding to Lyβ for the Lyα shown in Figure 2.2.[6] In many cases the Lyβ forest also includes higher order Lyman series absorption, as indicated in the figure. Although dark gaps are highly opaque to Lyα, there are often narrow transmission peaks corresponding to Lyβ. These peaks demonstrate that the dark gaps in Lyα typically cannot arise from continuous regions of neutral gas, which would be highly opaque to all Lyman series lines. Broken regions of neutral gas may still be present, however, with the Lyβ transmission corresponding to gaps between neutral sections (e.g., Keating et al. 2020b; Nasir & D'Aloisio 2020).

2.4.3 Overview of dark gaps

In total, we detected 1329 dark gaps from the sample, of which 50 have a length of $L \geq 30$ h^{-1} Mpc. Properties of all dark gaps detected are summarized in Table 2.2. Details on dark gap detection for each QSO sightline are shown in Figure Set 2.1.

As an overview, Figure 2.3 plots all dark gaps identified in this work according to their central redshift and length. Dark gaps with associated metal absorbers are labeled in red. This figure has excluded dark gaps that are completely inside the 7 proper-Mpc proximity zone and/or inside the 30 h^{-1} Mpc "buffer zone" beyond the proximity zone.

[6] We use a power-law fit to the continuum for regions blueward of the Lyα forest because our PCA implementation does not cover these wavelengths; however, this should not significantly affect the qualitative results for the higher-order Lyman series transmission shown in Figure 2.2.

Table 2.2: Properties of dark gaps

Index	QSO	$z_{\rm blue}$	$z_{\rm red}$	$L(h^{-1}{\rm Mpc})$	$\tau_{\rm eff}$	$z_{\rm absorber}$
26	ULASJ1319+0950	5.876	6.012[b]	≥ 40	≥ 6.674	
157	SDSSJ0100+2802	5.883	5.988	31	7.874 ± 0.339	5.945, 5.940
240	PSOJ108+08	5.661	5.836[b]	≥ 54	6.062 ± 0.145	
281	PSOJ183-12	5.332	5.350	6	4.193 ± 0.133	
292	PSOJ183-12	5.690	5.702	4[c]	3.951 ± 0.082	
294	SDSSJ1602+4228	5.065	5.071	2	3.580 ± 0.251	
350	ATLASJ025-33	5.285[a]	5.356	24	5.680 ± 0.226	
817	SDSSJ1148+5251	5.853	6.285[b]	≥ 124	≥ 7.558	6.258, 6.011, 6.131
959	SDSSJ1137+3549	5.683	5.686	1	≥ 3.569	

Table Comments: Columns: (1) index of the dark gap, (2) QSO name, (3) redshift at the blue end of the gap, (4) redshift at the red end of the gap, (5) dark gap length, (6) effective opacity of the dark gap based on the flux and flux error in the original binning, (7) redshift(s) of known metal absorber(s) in the dark gap, if any.
[a] Dark gap starting at the blue edge of the Lyα forest.
[b] Dark gap ending at the red edge of the Lyα forest.
[c] Dark gap located completely inside the buffer zone.
(This table is published in its entirety in the machine-readable format. A portion is shown here for guidance regarding its form and content.)

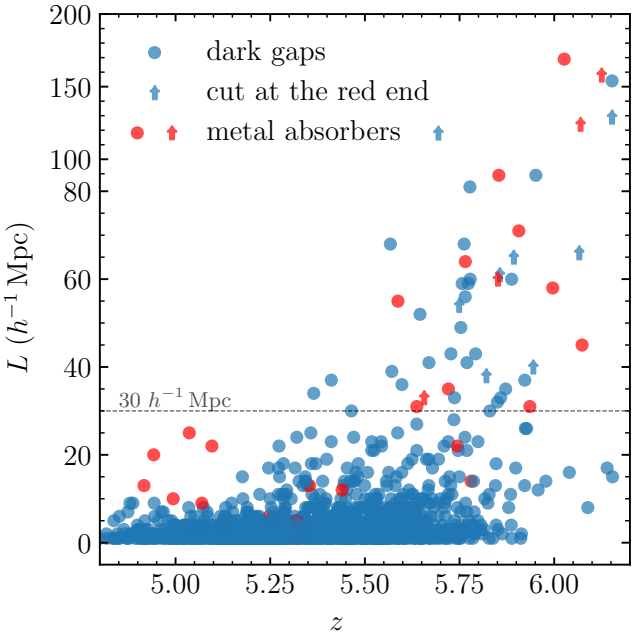


Figure 2.3: Gap length versus central redshift for dark gaps detected from our sample. Dark gaps located completely in the "buffer zone" are excluded from this plot. Arrows indicate dark gaps whose red edge lies within 7 pMpc from the QSO and are therefore potentially truncated by the proximity effect; lower limits on the length are therefore given for these gaps. Red symbols indicate dark gaps with one or more metal absorbers.

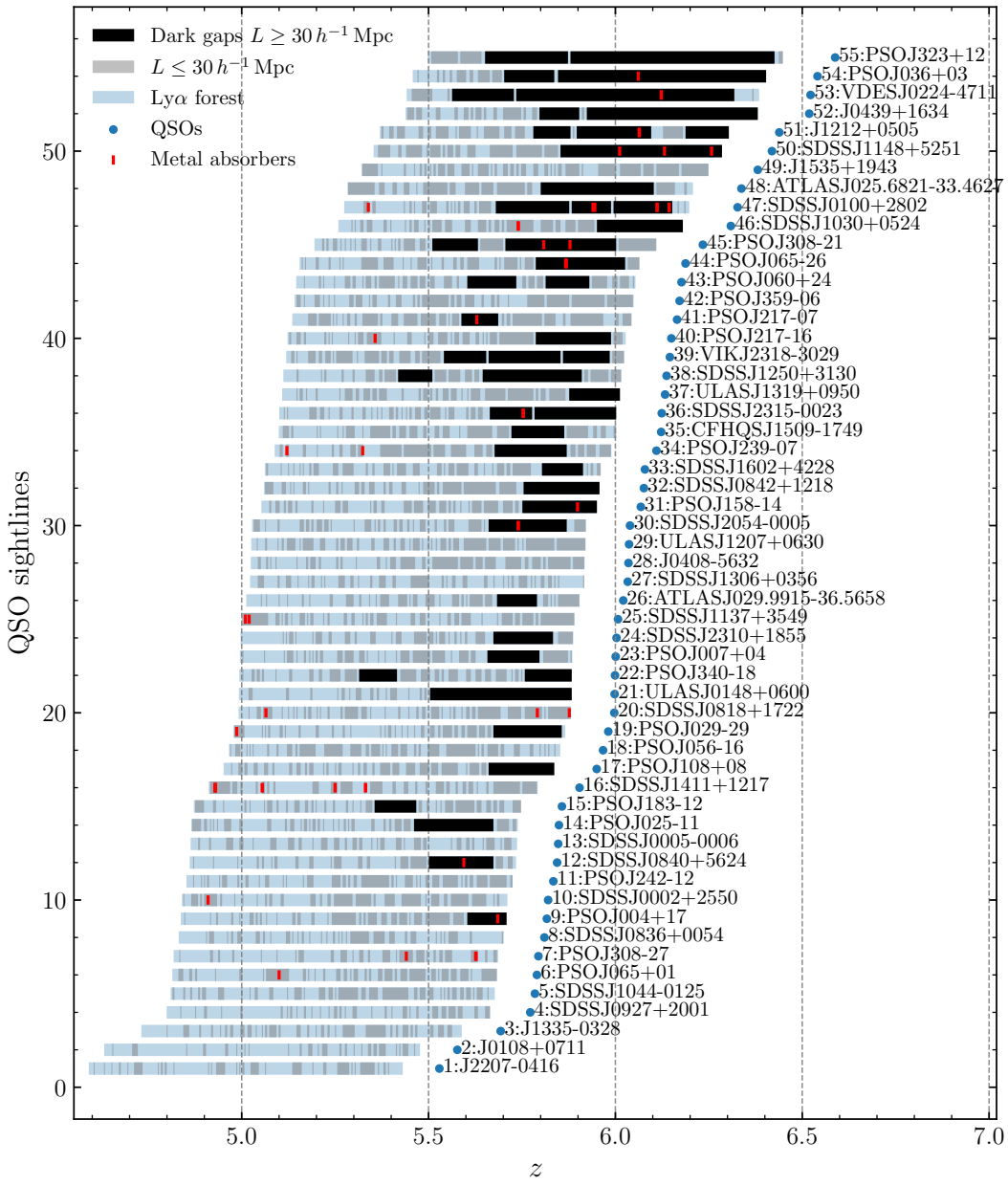


Figure 2.4: Overview of dark gaps identified in the Lyα forest of 55 QSO sightlines. Black bars and gray shaded regions represent dark gaps longer and shorter than 30 h^{-1} Mpc, respectively. Red short lines denote known associated metal absorbers intervening dark gaps. Light blue shaded regions indicate the redshift coverage of the Lyα forest. Redshifts of QSOs are marked with blue dots. The Lyα forest is truncated at 7 pMpc from the QSO. The Lyα forest shown in this figure *includes* the 30 h^{-1} Mpc buffer zone on the red end, which is excluded from the statistical analysis of dark gaps. See Section 2.4.1 for details.

Not surprisingly, as redshift increases, there are more long dark gaps and a larger scatter in dark gap length. The lowest-redshift gaps with $L \geq 30\ h^{-1}\,\mathrm{Mpc}$ appear around $z = 5.3$.

Figure 2.4 displays the Lyα forest coverage and all dark gaps identified for every line of sight in our sample. At $z \lesssim 5.2$, most QSO sightlines are highly transmissive; a few gaps with $L \sim 10$–$20\ h^{-1}\,\mathrm{Mpc}$ appear but these tend to contain metal absorbers and are likely to be DLAs. Dark gaps longer than $30\ h^{-1}\,\mathrm{Mpc}$ appear in the sightlines of PSO J340-18 and PSO J183-12 at $z \simeq 5.3$ and 5.4, respectively. The frequency of long dark gaps increases with redshift such that most lines of sight at $z \simeq 5.8$ show gaps longer than $30\ h^{-1}\,\mathrm{Mpc}$ in the Lyα forest. Interestingly, the J1535+1943 sightline is relatively transmissive at $z \sim 6$ compared to others at the same redshift. Although J1535 has a reddened spectrum, the continuum re-construction is acceptable and most of the transmission peaks in the Lyα forest appear to be real.

2.4.4 Fraction of QSO spectra exhibiting long dark gaps

We introduce the fraction of QSO spectra exhibiting long ($L \geq 30\ h^{-1}\,\mathrm{Mpc}$) gaps as a function of redshift, $F_{30}(z)$, as a new Lyα forest statistic. As mentioned in Section 2.4.1, in order to deal with the finite length of the spectra for this statistic we cut off each QSO sightline at the blue edge of the $30\ h^{-1}\,\mathrm{Mpc}$ buffer zone. F_{30} quantifies how common the large Lyman-alpha-opaque regions are and how they evolve with redshift. We choose $30\ h^{-1}\,\mathrm{Mpc}$ because we found that this length most effectively distinguishes between the models described in Section 2.5, especially between the `homogeneous-UVB` and other models. The comparison of the dark gap length distribution, $P(L)$, predicted by different models in Section 2.5.5 also implies that dark gaps with $L \geq 30\ h^{-1}\,\mathrm{Mpc}$ are potentially good

probes for H I if the late reionization scenario is indeed preferred. We note that we include all long dark gaps regardless of the presence of associated metal absorbers since the dense absorption systems alone are not likely to create troughs longer than 30 h^{-1} Mpc.

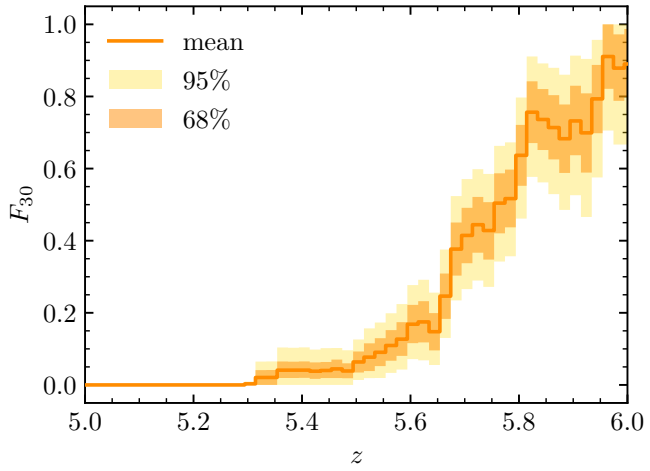


Figure 2.5: Measured fraction of QSO spectra exhibiting long ($L \geq 30\ h^{-1}$ Mpc) dark gaps as a function of redshift. We use bootstrap re-sampling to calculate the mean, 68% and 95% limits of F_{30} averaged over $\Delta z = 0.02$ bins, presented with the solid orange line, dark and light shaded regions, respectively.
(Tabular Data behind the Figure (DbF) is available online.)

Figure 2.5 displays the evolution of F_{30} with redshift measured from the QSO spectra. The result is averaged over $\Delta z = 0.02$ bins. The mean, 68% limits, and 95% limits of F_{30} are calculated based on 10000 bootstrap re-samplings of the whole sample. In each realization, we randomly select 55 QSO spectra, with replacement, and add up the number of sightlines showing $L \geq 30\ h^{-1}$ Mpc dark gaps at a given redshift. The total is then normalized by the number of QSO sightlines at each redshift, which yields $F_{30}(z)$ for this realization. F_{30} starts to be nonzero from $z \simeq 5.3$ and increases strongly with redshift. At $z = 6$, $\sim$90% of sightlines present long gaps.

We noted above that a deep spectrum of SDSS J1250+3130 was obtained based on preliminary indications from shallower data of a long gap in its spectrum. This is the only QSO in the sample for which the selection is related to the foreknowledge of dark gaps. We include J1250 for completeness, but note that excluding this line of sight from our sample would only decrease (increase) F_{30} by $\lesssim 0.02$ (0.05) over $5.50 < z < 5.90$ ($5.90 < z < 5.93$).

Finally, we test whether metal absorbers could be linking adjacent dark gaps in a way that would impact our F_{30} statistic. For this we calculate a "pessimistic" F_{30} by dividing dark gaps at the redshifts of DLAs and other metal systems (Appendix D). The resulting change in F_{30} is minor, with a maximum decrease of ~ 0.1 at $z \sim 5.8$. The differences between the observations and model predictions (Section 2.6.1; Figure D.1) can still be well distinguished. We therefore conclude that this potential impact of metal absorbers on F_{30} is not significant.

2.4.5 Distribution of dark gap length

In addition to F_{30}, we investigate the cumulative distribution function (CDF) of gap length, $P(< L)$. Figure 2.6 plots $P(< L)$ in redshift bins of $\Delta z = 0.2$. Dark gaps are assigned to a bin based on the central redshift of the gap, and we do not truncate gaps extending beyond the edges of the redshift windows. We treat the dark gaps truncated by the 7 proper-Mpc proximity zone cut by plotting the most pessimistic and optimistic bounds on $P(< L)$. The pessimistic bound is calculated by considering the lengths of dark gaps are as measured. The optimistic bound, however, is given by assuming the lengths of truncated dark gaps are infinite, which indicates the most extreme dark gap length possible

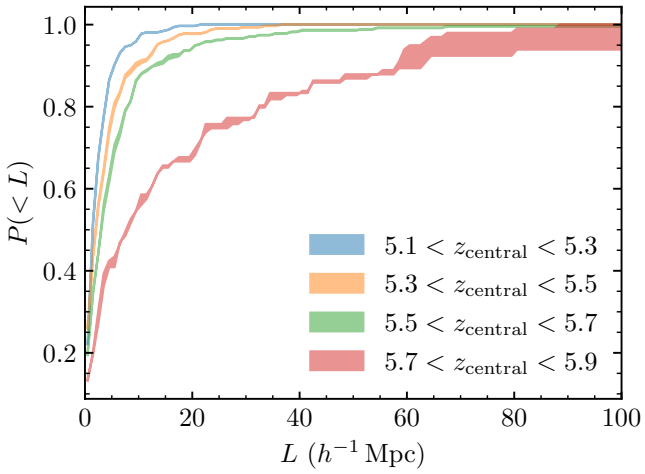


Figure 2.6: Cumulative distributions of dark gap length. We include all dark gaps regardless of the presence of associated metal absorbers. The upper and lower bounds of the shaded regions correspond to the most pessimistic and optimistic cases for $P(< L)$. See text for details.

in the absence of the QSO. In the latter case, we still use the measured central redshift of each dark gap to assign it to a redshift bin.

Figure 2.6 demonstrates that longer dark gaps become more common towards higher redshifts. This is consistent with the result of F_{30}. Moreover, similar to the rapid redshift evolution in F_{30} near $z \simeq 5.7$, $P(< L)$ shows a large change between $5.5 < z < 5.7$ and $5.7 < z < 5.9$.

To test the effects of metal absorbers on $P(< L)$, we calculate the distribution by excluding dark gaps with known associated metal absorbers. We find the difference is minor. The maximum increment on the most pessimistic $P(< L)$ over $5.7 < z < 5.9$ is less than 0.03, and the difference is less than 0.005 over the other redshift bins.

Table 2.3: Models used in this work

Model	Reionization	z_{95}	z_{50}
`homogeneous-UVB`	-	15	-
`K20-low-`$\tau_{\rm CMB}$	late	5.6	6.7
`K20-low-`$\tau_{\rm CMB}$`-hot`	late	5.6	6.7
`K20-high-`$\tau_{\rm CMB}$	late	5.9	8.4
`ND20-late-longmfp`	late	5.3	7.0
`ND20-late-shortmfp`	late	5.4	7.5
`ND20-early-shortmfp`	early	6.6	8.7

Columns: (1) name of the model, (2) qualitative description of the reionization model, (3) redshift at which the volume filling factor of ionized gas reaches 95%, (4) redshift at which the volume filling factor of ionized gas reaches 50%. We use `K20` for models from Keating et al. (2020a) and `ND20` for models from Nasir & D'Aloisio (2020). See Sections 2.5.2 & 2.5.3 for details.

2.5 Models and Simulations for Comparison

We compare our measurements to predictions from hydrodynamical simulations that span a range of reionization histories and UV backgrounds. Here we briefly describe the simulations. The key information is summarized in Table 2.3, with the redshift evolution of the volume-weighted neutral hydrogen fraction $\langle x_{\rm HI} \rangle$ for each simulation plotted in Figure 2.7.

2.5.1 Homogeneous UV Background

We first include a baseline model wherein reionization is fully completed at $z > 6$ and the UVB is spatially uniform. For this we use a run from the Sherwood simulation suite, which successfully reproduces multiple characteristics of the observed Lyα forest over $2.5 < z < 5$ (Bolton et al. 2017). The Sherwood suite uses a homogeneous Haardt & Madau (2012) UV background. Reionization occurs instantaneously at $z = 15$, allowing the

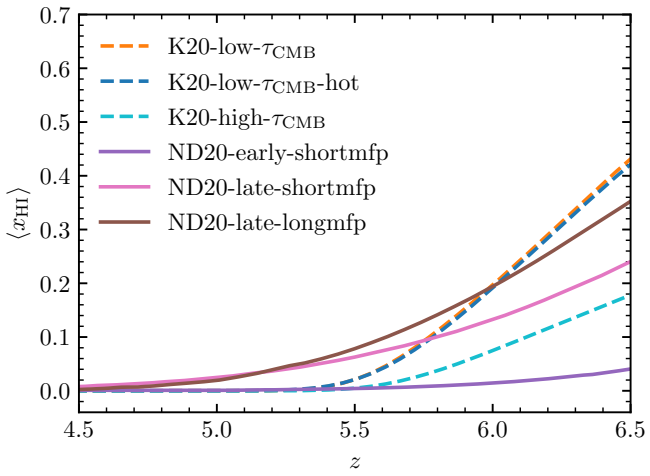


Figure 2.7: Redshift evolution of the volume-weighted average neutral hydrogen fraction of reionization models used in this work. See Table 2.3 for the key information of these models and details in Section 2.5.

IGM to fully relax hydrodynamically by $z = 6$. The simulations were run with the parallel smoothed particle hydrodynamics code `P-GADGET-3`, which is an updated and extended version of `GADGET-2` (Springel 2005). We use the simulation with 2×2048^3 particles and box size of $(160\ h^{-1}\,\mathrm{Mpc})^3$ to build mock spectra for the `homogeneous-UVB` model, as described in Section 2.5.4.

2.5.2 Late Reionization

We use two sets of models wherein reionization continues significantly below redshift six. In these models, long dark gaps in Lyα transmission at $z < 6$ arise from a combination of neutral islands and regions of suppressed UVB, which are often adjacent to one another.

The first late reionization models are from Keating et al. (2020a). They include three models with different ionization and/or thermal histories. We denote the fiducial

model as `K20-low-`$\tau_{\rm CMB}$, wherein the volume filling fraction of ionized gas reaches 95% at $z = 5.6$ and 99.9% at $z = 5.2$. Two other runs, the `K20-low-`$\tau_{\rm CMB}$`-hot` and `K20-high-`$\tau_{\rm CMB}$ models, are also included. Briefly, the `K20-low-`$\tau_{\rm CMB}$`-hot` model uses a higher temperature for the input blackbody ionizing spectrum, namely $T = 40000$ K instead of $T = 30000$ K as used in the `K20-low-`$\tau_{\rm CMB}$ model. They have a volume-weighted mean temperature at the mean density at $z = 6$ of $T_0 \simeq 10000$ K and 7000 K, respectively. The `K20-high-`$\tau_{\rm CMB}$ model shares a similar IGM thermal history with the `K20-low-`$\tau_{\rm CMB}$ model, but it has an earlier reionization midpoint of $z_{\rm mid} = 8.4$.

The `K20` simulations are modified versions of the late reionization model published in Kulkarni et al. (2019a). The model was modified such that (i) the IGM temperature evolution is in better agreement with recent observations (Boera et al. 2019; Walther et al. 2019; Gaikwad et al. 2021), and (ii) the mean Lyα transmission is in better agreement with data at $z < 4.7$ (Becker et al. 2015). The ionization state of the IGM is modeled using the radiative transfer code `ATON` (Aubert & Teyssier 2008, 2010) that post-processes underlying hydrodynamic simulations performed with `P-GADGET-3`. The simulations use the identical initial condition and box size of the Sherwood simulation suite. The radiative transfer, however, leads to an extended and self-consistent reionization history. This produces scatter in the Lyα $\tau_{\rm eff}$. The simulations also contain fluctuations in temperature and photoionization rates. A lightcone from the radiative transfer simulation were extracted on the fly. Using sightlines through this lightcone, Keating et al. (2020a) computed the optical depths continuously spanning $4.0 \lesssim z \lesssim 7.5$ for each model, which allows us to avoid having to do any interpolation.

The second set of late reionization models is from Nasir & D'Aloisio (2020). In these models, the volume filling factor of ionized gas reaches 95% at $z = 5.3$–5.4. As in the Keating et al. (2020a) models, fluctuations in both the UVB and temperature are present. The UVB fluctuations are driven by a short and spatially variable mean free path, similar to the model in Davies & Furlanetto (2016). In the two Nasir & D'Aloisio (2020) models, which we denote as `ND20-late-longmfp` and `ND20-late-shortmfp`, the volume-weighted average mean free path for 912 Å photons at $z = 5.6$ is $\langle \lambda_{\rm mfp}^{912} \rangle = 30\ h^{-1}\,{\rm Mpc}$ and $10\ h^{-1}\,{\rm Mpc}$, respectively. As a result of the shorter mean free path, `ND20-late-shortmfp` contains stronger fluctuations in the UVB. The shorter $\langle \lambda_{\rm mfp}^{912} \rangle$ is also more consistent with the recent mean free path measurement of Becker et al. (2021).

The Nasir & D'Aloisio (2020) simulations use a modified version of the Eulerian hydrodynamics code from Trac & Pen (2004). They use 2×2048^3 gas and dark matter resolution elements and a box size of $L = 200\ h^{-1}\,{\rm Mpc}$. To model the effects of reionization on the forest, they post-process the hydrodynamics simulations using semi-numeric methods. Optical depth skewers are available at $z = 5.6$, 5.8, and 6.0, and neutral fraction information is available at $z = 5.6$ and 5.8. A sample of 4000 lines of sight were extracted at each redshift, with each optical depth skewer having a length of $500\ h^{-1}\,{\rm Mpc}$ by making use of the periodic boundary conditions (F. Nasir, private communication).

2.5.3 Early reionization with a fluctuating UVB

Finally, we include a model from Nasir & D'Aloisio (2020) wherein the volume filling factor of ionized gas reaches $\sim$ 98% by $z = 6$ but the UVB retains large spa-

tial fluctuations to somewhat lower redshifts.[7] It has $\langle \lambda_{\rm mfp}^{912} \rangle = 10\ h^{-1}\,{\rm Mpc}$ as in the `ND20-late-shortmfp` model. We refer to this model as `ND20-early-shortmfp`. It is essentially a modified version of the fluctuating UVB model proposed by Davies & Furlanetto (2016) with temperature fluctuations included. Compared to the `ND20-late-shortmfp` model mentioned previously, the `ND20-early-shortmfp` model has a similarly broad UVB distribution but a much earlier end of reionization. In this model, long dark gaps at $z < 6$ primarily correspond to regions with a low UVB. Since the IGM is not technically fully ionized in this model until down to $z \simeq 5$, however, a small fraction of dark gaps may still contain some neutral hydrogen.

2.5.4 Construction of mock spectra

In order to directly compare the observations to the models we construct mock spectra from the simulations with properties similar to the real data. We firstly describe how we create mock spectra for the `homogeneous-UVB` model.

The snapshots for the `homogeneous-UVB` model are available on every $\Delta z = 0.1$ interval over $3.9 \leq z \leq 8.9$. To be consistent with the simulations from Nasir & D'Aloisio (2020) we only use snapshots from every $\Delta z = 0.2$, and the same snapshots are used for every sightline. We have verified, however, that using snapshots spaced every $\Delta z = 0.1$ would not significantly impact our results. Each snapshot was used to extract 5000 160 $h^{-1}\,{\rm Mpc}$ skewers along which the native Lyα optical depths have been calculated (Bolton et al. 2017). For a mock spectrum centered at redshift z_0 we combine skewers from redshifts $z_0 - 0.2$, z_0,

[7] The volume filling factor of ionized gas no longer increases significantly at $z < 6$. Although it has not reached 99% strictly by $z = 6$, we still consider this model as an early reionization model.

and $z_0 + 0.2$ [8] after shifting the periodic lines of sight by random amounts. The resulting mock spectra are still 160 h^{-1} Mpc in length but contain information about the redshift evolution of the Lyα-opaque regions. We fit the $\tau_{\rm eff}$ evolution over $5 \leq z \leq 6$ from Bosman et al. (2018) with a power law of $\tau_{\rm eff} \propto (1+z)^{12.34}$ and re-normalize the optical depths of the mock spectra such that their average Lyα transmission matches this evolution. We have also checked that the mean transmission measured directly from our observed sample is within the 1σ uncertainties of the measurement in Bosman et al. (2018). We create 5000 mock spectra matching each of our 55 lines of sight. For each QSO, we bin the mock spectra using exactly the same wavelength array as the observed spectrum. We then add Gaussian noise to the mock spectra based on the corresponding flux error array.

Because each optical depth skewer from the Nasir & D'Aloisio (2020) models has a length of 500 h^{-1} Mpc, we first clip them to 160 h^{-1} Mpc and then follow a similar procedure to build the mock spectra set at $z_0 = 5.8$ as described above, including rescaling the effective optical depth. In order to cover the full redshift range of the simulation outputs, we extend the mock spectra down to $z = 5.6$ and up to $z = 6.0$ by making use of the unclipped skewers to create mock spectra sets centered at $z_0 = 5.6$ and 6.0. However, since the spatial structure of the IGM is only recovered over $5.6 \leq z \leq 6.0$, we restrict our dark gap analysis to this redshift range. As for `K20` models, Keating et al. (2020a) ran many radiative transfer simulations until converged on a reionization history that self-consistently reproduces the mean flux of the Lyα forest as measured by Bosman et al. (2018). We therefore only needed to re-bin the skewers and add noise in order to match them to each individual observed

[8] We cut the skewers into three pieces and then stitch the corresponding pieces with those from the adjacent redshifts. Only a portion of a 160 h^{-1} Mpc skewer from a given snapshot is therefore used for a mock spectrum centered at z_0.

QSO spectrum. We note that continuum errors are not considered for the mock spectra. This is because the continuum errors for the observed spectra are estimated to be small ($\lesssim$10%; Section 2.3.2), and partly because we are primarily concerned with very low flux levels, which are less affected in an absolute sense by continuum uncertainties.

In Figure 2.8, we display mock spectra randomly selected from all the models with S/N chosen to match the the Lyα forest of ULAS J0148+0600 as examples. The `homogeneous-UVB` model exhibits more small transmission peaks than the other models, as expected because the IGM is fully ionized by a uniform UVB. The other models tend to show longer dark gaps interspersed with regions of high transmission.

2.5.5 Neutral islands and dark gaps

Here we examine the connection between dark gaps and regions of neutral hydrogen. For this we calculate the dark gap length distribution $P(L)$ predicted by models. We use the method described in Section 2.4.1 to find dark gaps in mock spectra generated in Section 2.5.4, but with no noise added, and identify gaps that contain regions of neutral hydrogen. The frequency of dark gaps with length L for each model in each redshift bin is calculated based on 10000 realizations and normalized by the total count of dark gaps in each redshift bin, with $P(L)$ averaged over bins of $\Delta L = 5\ h^{-1}\,\mathrm{Mpc}$. We consider a dark gap to contain neutral hydrogen if any pixels inside this gap have $x_{\mathrm{HI}} > 0.9$. Over each redshift bin, dark gaps extending beyond the boundaries of the $\Delta z = 0.2$ window are truncated at the edge. We do so to avoid artifacts in $P(L)$ caused by the finite length of the mock spectra.

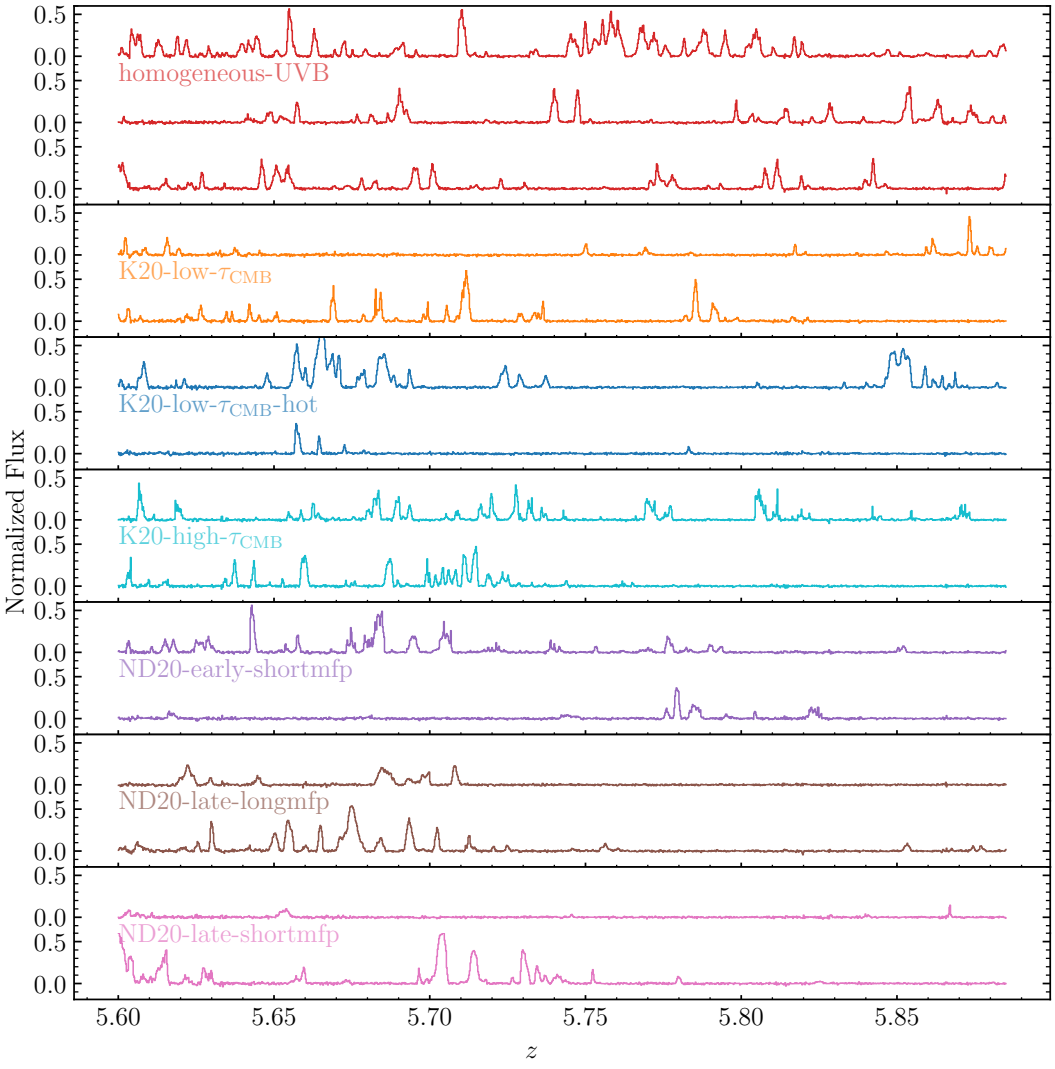


Figure 2.8: Example mock spectra with binning and S/N chosen to match the Lyα forest of ULAS J0148+0600. We randomly select mock sightlines from each model here. The colored lines represent the flux arrays.

As shown in Figure 2.9, $P(L)$ varies significantly between models. Firstly, no dark gaps with neutral pixels are found in the `homogeneous-UVB` model because the IGM is fully ionized. In the `ND20-early-shortmfp` model, the IGM is 98% ionized by $z = 6$, and therefore only a small fraction of dark gaps contain neutral islands. Dark gaps with no neutral islands also dominate in the `K20-high-`$\tau_{\rm CMB}$ model that has an extended reionization history. The situation is very different in the rapid late reionization scenarios, however. Dark gaps with neutral islands become dominant for $L \geq 15$–$20\ h^{-1}\,$Mpc in both `ND20-late` models. Similarly, in the `K20-low-`$\tau_{\rm CMB}$`(-hot)` model, dark gaps with neutral islands start to be the majority for $L \gtrsim 25$–$30\ h^{-1}\,$Mpc at $z > 5.4$. Long dark gaps with $L \gtrsim 30\ h^{-1}\,$Mpc are therefore of potentially high interest in terms of identifying regions of the IGM that may contain neutral gas. This paper is therefore largely focused on these long gaps.

We further investigate the correlation between neutral islands coverage and dark gap length in the `K20-low-`$\tau_{\rm CMB}$ model at different redshift, as shown in Figure 2.10. The histogram is calculated based on 10000 realizations, and we include all dark gaps regardless of whether they contain neutral pixels. The neutral islands coverage shown here is the sum of the line-of-sight length of neutral pixels inside a dark gap. The mean neutral islands coverage is proportional to the dark gap length, meaning that long dark gaps may contain more neutral gas. Nevertheless, the neutral islands coverage is, on average, significantly less than the dark gap length. This suggests that UVB fluctuations also play a significant role in producing the dark gaps in the late reionization models.

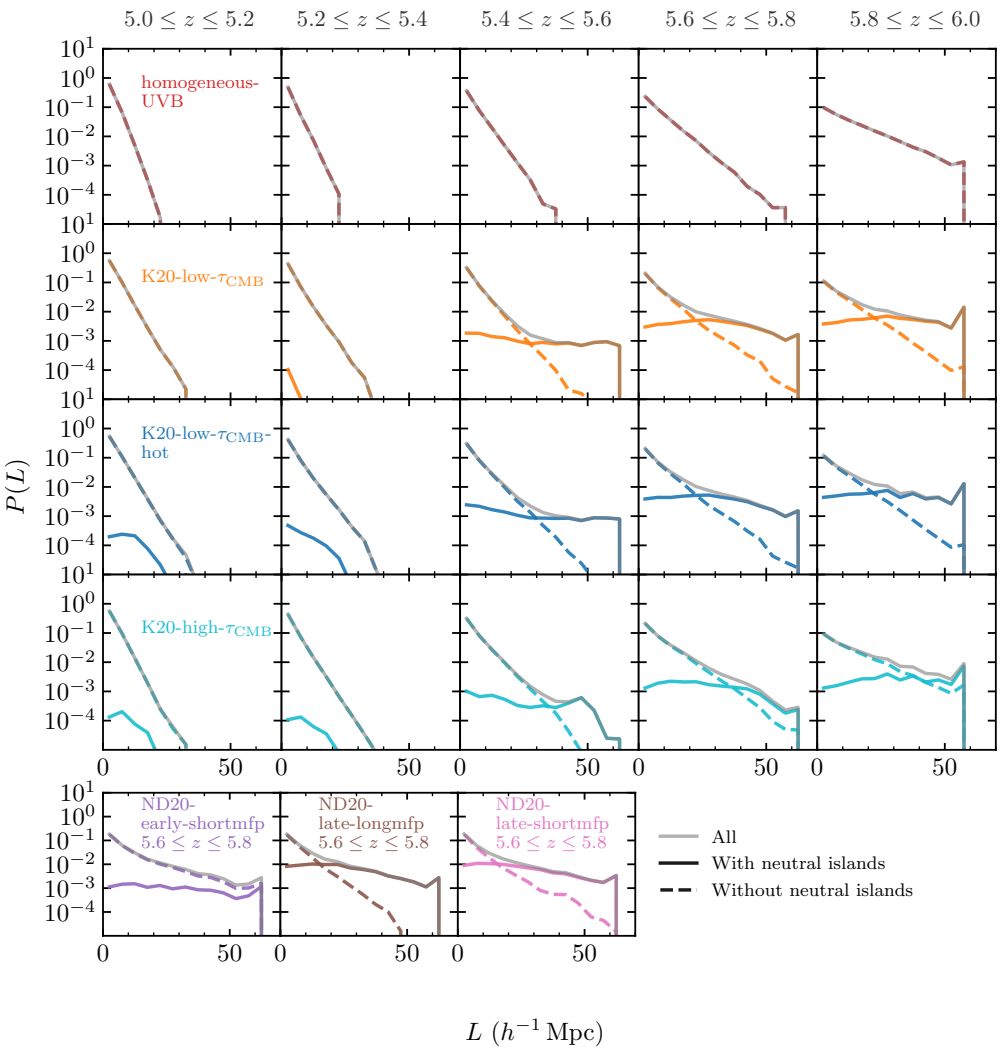


Figure 2.9: Distribution of dark gaps with and without neutral hydrogen predicted by simulations. A dark gap is considered to contain neutral gas if any pixels inside this gap have a neutral fraction of $x_{\rm HI} > 0.9$. $P(L)$ is calculated with count of dark gaps over $\Delta L = 5\ h^{-1}$ Mpc bins divided by the total count of dark gaps. The distribution is calculated with dark gaps detected in 10000 sets of mock spectra (Section 2.5.4) for each simulation, but with no noise added. We note that the volume neutral fraction information of Nasir & D'Aloisio (2020) models is only available at $z = 5.6$ and 5.8. The plots for `ND20` models are therefore not extended to $z = 6.0$.

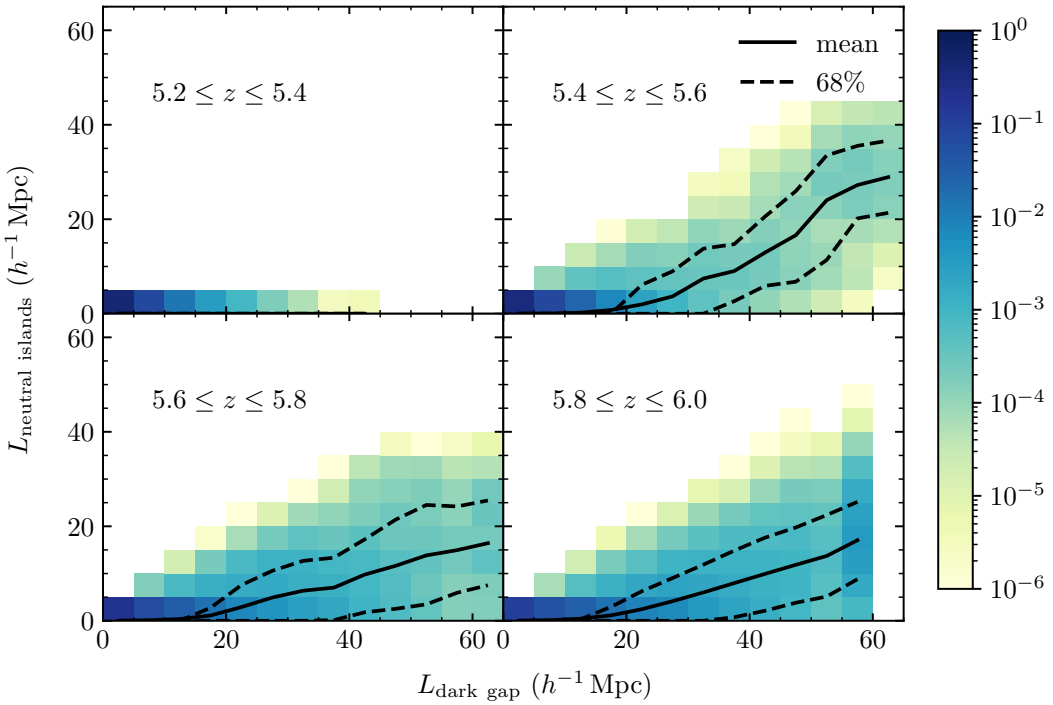


Figure 2.10: Correlation between neutral islands coverage and dark gap length in the `K20-low-`τ_{CMB} model based on 10000 realizations. The histogram is calculated on 5×5 (h^{-1} Mpc)^2 bins and color indicates the normalized probability, and all dark gaps are included regardless of whether they contain neutral gas. Solid and dashed lines show the mean and 68% interval of the neutral islands coverage.

2.6 Discussion

2.6.1 Model comparisons

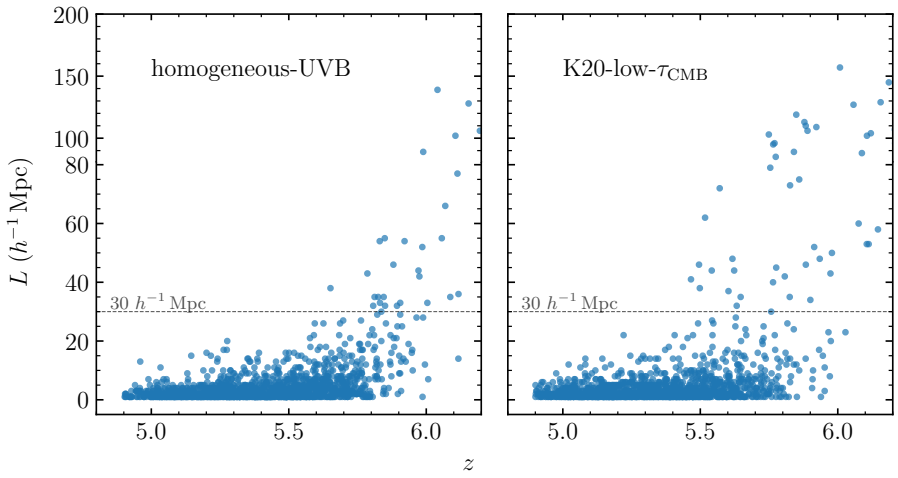


Figure 2.11: Gap length versus central redshift for dark gaps detected in mock spectra. For both models, the results shown here are based on one randomly selected set of mock spectra that matches our QSO sample in redshift and S/N ratio.

We now compare our results to predictions from the simulations described in Section 2.5. Figure 2.11 plots the dark gap length versus central redshift for representative mock samples drawn from the `homogeneous-UVB` model and the `K20-low-`$\tau_{\rm CMB}$ model. Qualitatively, as redshift increases, the `homogeneous-UVB` model predicts a milder increase in long dark gaps than is seen in either the `K20-low-`$\tau_{\rm CMB}$ model or the observations (Figure 2.3). To quantify the differences, we compute the relevant statistics by drawing mock samples from the simulations that match our observed QSO spectra in redshift and S/N ratio. We then compute the dark gap statistics described in Section 2.4. We repeat this

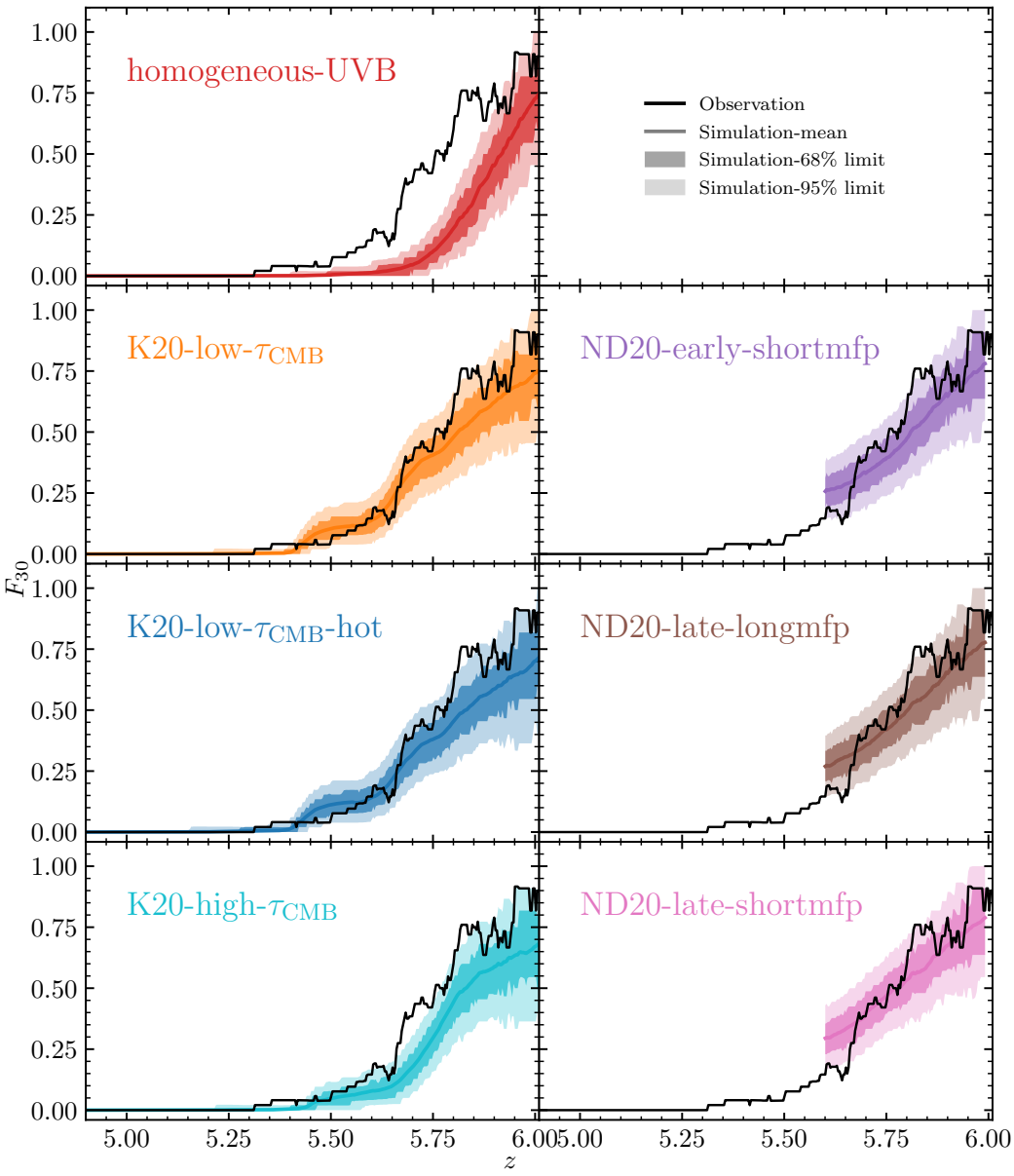


Figure 2.12: The fraction of sightlines located in dark gaps with $L \geq 30\ h^{-1}\,\mathrm{Mpc}$ as a function of redshift. Both the observations (solid black line) and simulations (colored solid line and shaded regions) include gaps that are truncated at the red end of the Lyα forest. Note that we discard dark gaps that are entirely in the 7 proper-Mpc proximity zone and/or in the buffer zone that covers 30 $h^{-1}\,\mathrm{Mpc}$ blueward the proximity zone cut. Dark and light shaded regions show the 68% and 95% intervals, respectively, spanned by the models.

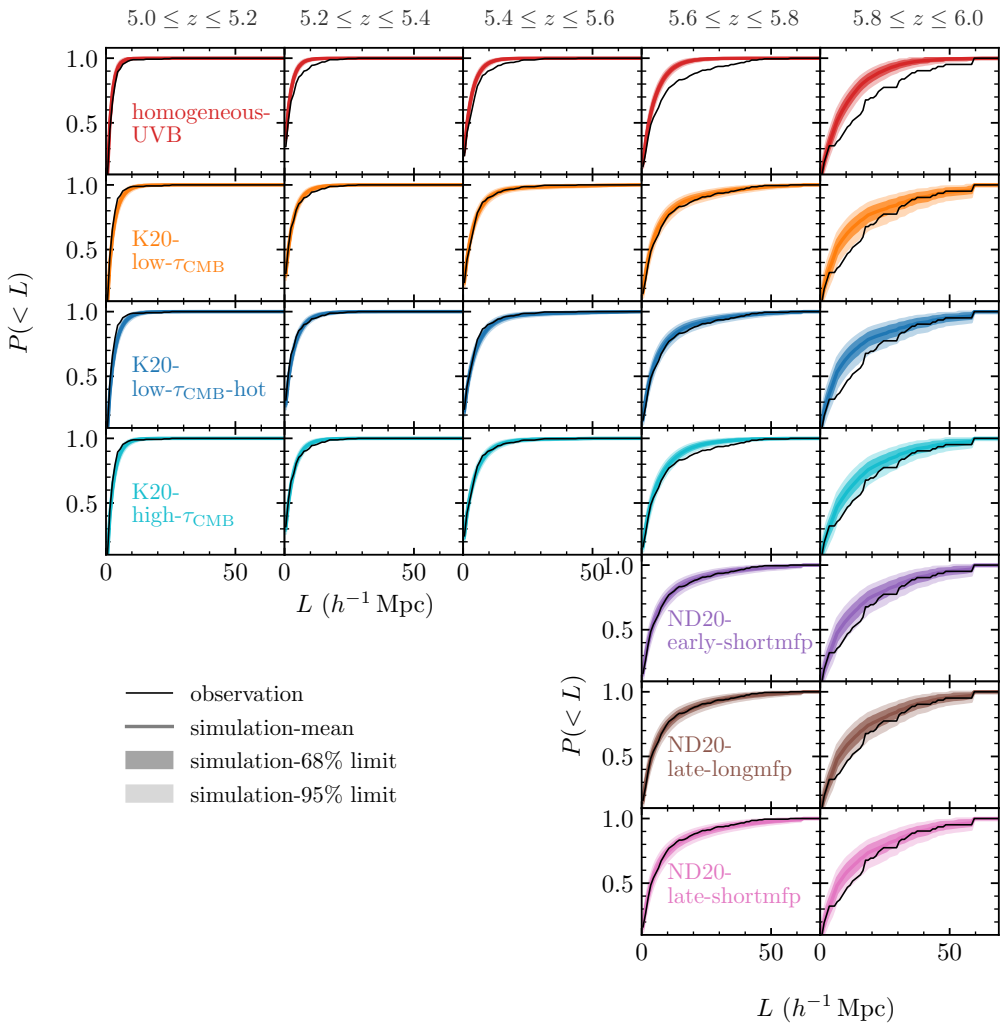


Figure 2.13: Cumulative distributions of dark gap length. In each redshift bin, the black line shows $P(< L)$ of the observed dark gaps. Dark gaps extending beyond the boundaries of the $\Delta z = 0.2$ window are truncated at the edge. The colored lines and dark/light shaded regions represent the mean and 68%/95% limits of $P(< L)$ in mock samples drawn from the models. In this figure, we use a high-mass-resolution run from the Sherwood Simulation Suite (Bolton et al. 2017) for the `homogeneous-UVB` model instead of the fiducial configuration. See text for details.

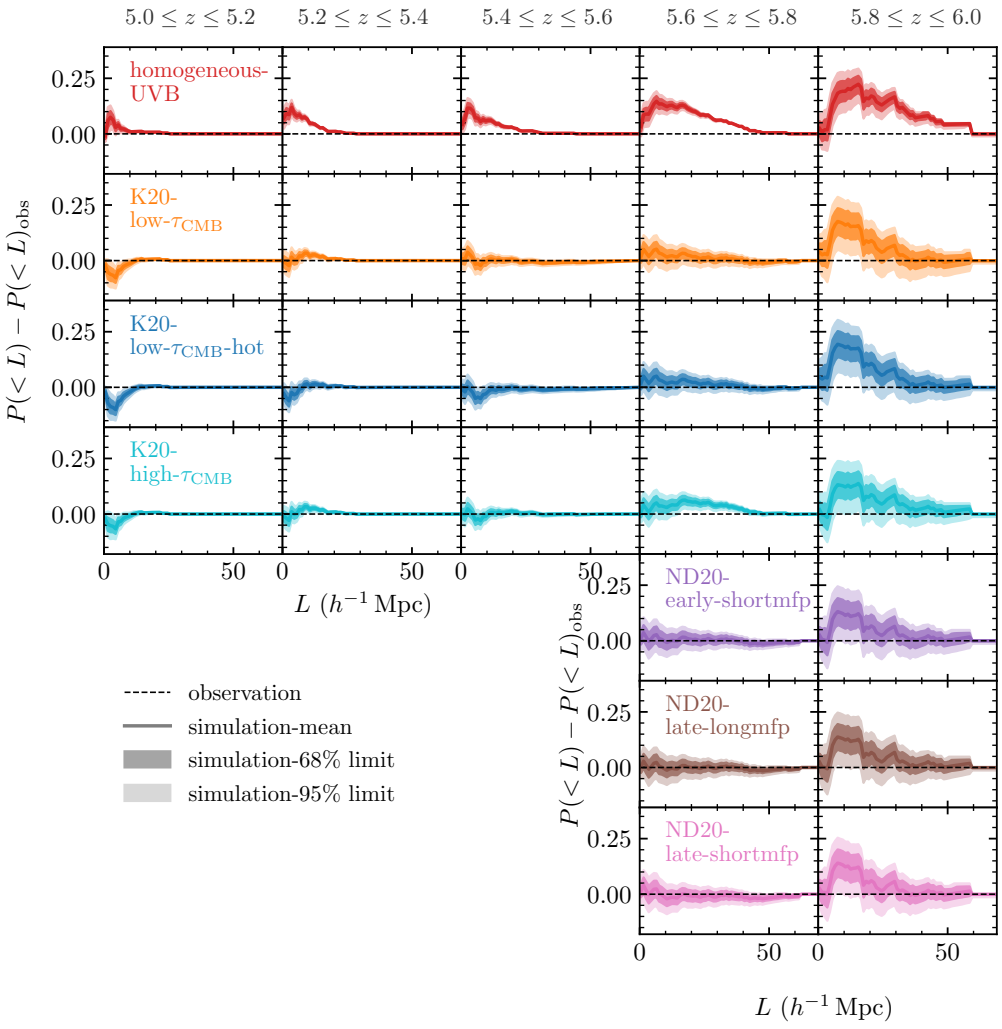


Figure 2.14: Similar to Figure 2.13, but showing the differences on cumulative distributions of dark gap length between the models and the observations.

process 10000 times for each model and compute the mean, 68% and 95% limits on the expected scatter for the present sample size. Figure 2.12 compares F_{30} predicted by models to that calculated from data. The jagged edges of the simulation confidence intervals are caused by the combined effects of step changes in the number of sightlines with redshift and the quantization of F_{30} for a finite sample size.

The top left panel shows that the `homogeneous-UVB` model is highly inconsistent with the observations over $5.3 \lesssim z \lesssim 5.9$. At $z \sim 5.8$, the `homogeneous-UVB` model under-predicts F_{30} by a factor of three. At $z \sim 5.4$ and over $5.5 \lesssim z \lesssim 5.8$, this model is rejected by the data with $> 99.9\%$ confidence.

On the other hand, the `K20-low-`$\tau_{\rm CMB}$ and `K20-low-`$\tau_{\rm CMB}$`-hot` models, wherein reionization ends at $z \simeq 5.3$, produce F_{30} results that are generally consistent with the observations over $5 < z < 6$. One exception is that these models under-predict the small number of long dark gaps observed at $z \sim 5.4$. The `K20-high-`$\tau_{\rm CMB}$ model is consistent with the observations at $z \geq 5.75$ but under-predicts F_{30} at lower redshifts. This is a natural consequence of the earlier reionization in this model, which leads to a lower neutral hydrogen fraction and smaller UVB fluctuations at these redshifts.

As shown in the right panels, F_{30} values from the Nasir & D'Aloisio (2020) models are consistent with the observations within their 95% limits over the available redshift range. Among the `ND20` models, `ND20-early-shortmfp` gives lower F_{30} values compared to `ND20-late`, but the difference is within the 68% range for the present sample size.

We compare the cumulative distributions of dark gap length in Figure 2.13, and give the differences between the observation and the model predictions in Figure 2.14. In

order to facilitate a direct comparison between the observations and simulations, we divide the data into redshift bins of $\Delta z = 0.2$. Here, dark gaps extending beyond the boundaries of a redshift bin are truncated at the edge when calculating $P(< L)$ for both the observation and models. Similar to our approach in Section 2.5.5, we do this to avoid artifacts from the finite length of the mock spectra.

We present numerical convergence tests for the `homogeneous-UVB` model in Appendix A. We find that the results for both F_{30} and $P(< L)$ are relatively insensitive to box size, but that the number of small gaps increases with increasing mass resolution. The impact of mass resolution is more significant for $P(< L)$ at smaller gap lengths than for F_{30}. For $P(< L)$ measured from the `homogeneous-UVB` model, therefore, we display predictions based on a higher-resolution run with 2×2048^3 particles and a box size of $L = 40\ h^{-1}\,\mathrm{Mpc}$ (hereafter `40_2048`) instead of the fiducial configuration of 2×2048^3 particles and box size of $L = 160\ h^{-1}\,\mathrm{Mpc}$ (hereafter `160_2048`). Because Keating et al. (2020a) use post-processed radiative transfer simulations, and Nasir & D'Aloisio (2020) simulations are based on an Eulerian code instead of a SPH code, mass resolution effects may be significantly different for these models than for the `homogeneous-UVB` model. We therefore present results as they are, although mass resolution corrections may be needed.

Over $z = 5.6$–6.0, the `homogeneous-UVB` model predicts significantly fewer long gaps than are observed in the data. The discrepancies between the data and the `homogeneous-UVB` model persist down to the $z = 5.2$–5.4 bin.

In contrast, the late reionization models, `K20-low-`τ_{CMB}, `K20-low-`τ_{CMB}`-hot`, and `K20-high-`τ_{CMB}, predict $P(< L)$ values that are generally consistent with the data. Never-

theless, over $z = 5.7$–5.9, we note that these models, especially the `K20-high-`$\tau_{\rm CMB}$ model, systematically yield higher $P(< L)$, i.e. fewer long gaps, than the observed for some L, though the discrepancies are less conspicuous compared to those for the `homogeneous-UVB`. At lower redshifts, there are minor differences between the `K20` models and the observation. The `ND20-early-shortmfp` and `ND20-late` models are generally consistent with the observation in the redshift range ($5.6 \leq z \leq 6.0$) currently probed by the simulations.

2.6.2 Implications for reionization

Combining the results for F_{30} and $P(< L)$, it is evident that a fully ionized IGM with a homogeneous UV background is disfavored by the observations down to $z \sim 5.3$. This result is consistent with the large-scale inhomogeneities in IGM Lyα opacity seen in recent $\tau_{\rm eff}$ measurements (Becker et al. 2015; Bosman et al. 2018; Eilers et al. 2018; Yang et al. 2020b; Bosman et al. 2022).

The late reionization models from Keating et al. (2020a) and Nasir & D'Aloisio (2020) are generally consistent with dark gap statistics in the Lyα forest. In these models, the residual neutral islands at $z < 6$ coupled with UVB fluctuations can naturally explain the appearance of long dark gaps in the Lyα forest. Among these models, the data tend to prefer those with later and more rapid reionization histories. For example, the `K20-low-`$\tau_{\rm CMB}$ and `K20-low-`$\tau_{\rm CMB}$`-hot` models, which have a reionization midpoint of $z_{50} = 6.7$, is somewhat more consistent (see curves and shades near $5.6 \leq z \leq 5.8$ in Figures 2.12, 2.13, and 2.14) with the dark gap statistics at $z < 6$ than the `K20-high-`$\tau_{\rm CMB}$ mode, for which $z_{50} = 8.4$. A late and rapid reionization is also suggested by the recent mean free path measurement from Becker et al. (2021) (see also Cain et al. 2021; Davies et al. 2021).

Alternatively, long dark gaps can arise from a fully reionized IGM provided that there are large UVB fluctuations. The early reionization model from Nasir & D'Aloisio (2020), which retains post-reionization fluctuations in the UV background and IGM temperature, is consistent with the data over at least $5.6 < z < 6.0$, where the available simulation outputs allow mock spectra to be compared to the data using the methods described above. Extending these simulations down to lower redshifts would be helpful for testing the pure fluctuating UVB model further.

2.7 Summary

In this paper, we present a search for dark gaps in the Lyα forest over $5 < z < 6$. We use high-S/N spectra of 55 QSOs at $z_{\rm em} > 5.5$ taken with Keck ESI and VLT X-Shooter, including data from the new XQR-30 VLT Large Programme. We focus on two statistics: the fraction of sightlines containing dark gaps of length $L \geq 30\ h^{-1}\,$Mpc as a function of redshift, F_{30}, which we introduce here for the first time, and the dark gap length distribution, $P(< L)$. Our primary goal is to quantify the persistence of large Lyα-opaque regions in the IGM below redshift six, and to evaluate the consistency between the observed dark gap statistics and predictions from various models. We include a model with a fully ionized IGM and a uniform ionizing UV background, and others with large islands of neutral gas and/or UVB fluctuations. Our main results can be summarized as follows:

- We identify 50 long dark gaps ($L \geq 30\ h^{-1}\,$Mpc) in the Lyα forest from our sample. Two long dark gaps are found at $z < 5.5$, with one extending down to $z \simeq 5.3$. We also report new ultra-long dark gaps ($L > 80\ h^{-1}\,$Mpc) below $z = 6$, similar to the one

previously reported towards ULAS J0148+0600 by Becker et al. (2015). The presence of long dark gaps at these redshifts demonstrates that large regions of the IGM remain opaque to Lyα down to $z \simeq 5.3$.

- In terms of both F_{30} and $P(< L)$, a fully ionized IGM with a homogeneous UVB is disfavored by the data down to $z \simeq 5.3$.

- Models wherein reionization ends significantly below redshift six (Keating et al. 2020a; Nasir & D'Aloisio 2020) are broadly consistent with the data. Among these, the data favor models with a reionization midpoint near $z \sim 7$ and an end at $z \simeq 5.3$ or later. In these models, dark gaps arise from a combination of neutral patches in the IGM and regions of low ionizing UV background, which are often adjacent to one another.

- We also find consistency with a model wherein reionization ends by $z = 6$ but the IGM retains large fluctuations in the UV background (Nasir & D'Aloisio 2020), at least over $5.6 < z < 6.0$.

Overall, the evolution of dark gaps observed at $z < 6$ suggests that signatures of reionization remain present in the IGM until at least $z \simeq 5.3$ in the form of neutral hydrogen islands and/or fluctuations in the ionizing UV background. We note that this work focuses on dark gaps in the Lyα forest. Given its lower optical depth, however, Lyβ may also be a useful tool. For example, islands of neutral gas may tend to produce more long Lyβ troughs than are created by fluctuations in the UV background (e.g., Nasir & D'Aloisio 2020). These and other statistics should provide further details on how the IGM evolves near the end of reionization.

Chapter 3

Long Dark Gaps in the Lyβ Forest at $z < 6$: Evidence of Ultra Late Reionization from XQR-30 Spectra

3.1 Abstract

We present a new investigation of the intergalactic medium (IGM) near reionization using dark gaps in the Lyman-β (Lyβ) forest. With its lower optical depth, Lyβ offers a potentially more sensitive probe to any remaining neutral gas compared to commonly used Lyα line. We identify dark gaps in the Lyβ forest using spectra of 42 QSOs at $z_{\rm em} > 5.5$, including new data from the XQR-30 VLT Large Programme. Approximately 40% of these QSO spectra exhibit dark gaps longer than 10 $h^{-1}\,$Mpc at $z \simeq 5.8$. By comparing the results to predictions from simulations, we find that the data are broadly consistent both

with models where fluctuations in the Lyα forest are caused solely by ionizing ultraviolet background (UVB) fluctuations and with models that include large neutral hydrogen patches at $z < 6$ due to a late end to reionization. Of particular interest is a very long ($L = 28\ h^{-1}\,\mathrm{Mpc}$) and dark ($\tau_{\mathrm{eff}} \gtrsim 6$) gap persisting down to $z \simeq 5.5$ in the Lyβ forest of the $z = 5.85$ QSO PSO J025−11. This gap may support late reionization models with a volume-weighted average neutral hydrogen fraction of $\langle x_{\mathrm{HI}} \rangle \gtrsim 5\%$ by $z = 5.6$. Finally, we infer constraints on $\langle x_{\mathrm{HI}} \rangle$ over $5.5 \lesssim z \lesssim 6.0$ based on the observed Lyβ dark gap length distribution and a conservative relationship between gap length and neutral fraction derived from simulations. We find $\langle x_{\mathrm{HI}} \rangle \leq 0.05$, 0.17, and 0.29 at $z \simeq 5.55$, 5.75, and 5.95, respectively. These constraints are consistent with models where reionization ends significantly later than $z = 6$. [1]

3.2 Introduction

Determining when and how reionization occurred is essential for understanding the IGM physics and galaxy formation in the early Universe (e.g., Muñoz et al. 2022). Recent observations have made significant progress on establishing the timing of reionization and largely point to a midpoint of $z \sim 7 - 8$. These observations include the electron optical depth to the cosmic microwave background (CMB) photons (Planck Collaboration et al. 2020, see also de Belsunce et al. 2021), the Lyman-α (Lyα) damping wing in $z \gtrsim 7$ QSO spectra (e.g., Bañados et al. 2018; Davies et al. 2018b; Wang et al. 2020; Yang et al. 2020a; Greig et al. 2022), the decline in observed Lyα emission from $z > 6$ galaxies (e.g., Mason

[1] A version of this chapter is published as Zhu et al. (2022) — *Long Dark Gaps in the Lyβ Forest at $z < 6$: Evidence of Ultra-late Reionization from XQR-30 Spectra, The Astrophysical Journal, 932:76.*

et al. 2018, 2019; Hoag et al. 2019; Hu et al. 2019, and references therein, but see Jung et al. 2020; Wold et al. 2021), and the IGM thermal state measurements at $z > 5$ (e.g., Boera et al. 2019; Gaikwad et al. 2021).

The observations noted above are generally consistent with reionization ending by $z \sim 6$, a scenario supported by existing measurement of the fraction of dark pixels in the Lyman series forests (e.g., McGreer et al. 2011, 2015). Other observations, however, suggest a significantly later end of reionization. The large-scale fluctuations in the measured Lyα effective optical depth, $\tau_{\rm eff} = -\ln \langle F \rangle$, where F is the continuum-normalized transmission flux (e.g., Fan et al. 2006; Becker et al. 2015; Eilers et al. 2018; Bosman et al. 2022; Yang et al. 2020b), together with long troughs extending to or below $z \simeq 5.5$ in the Lyα forest (e.g., Becker et al. 2015; Zhu et al. 2021) potentially indicate the existence of large neutral IGM islands (e.g., Kulkarni et al. 2019a; Keating et al. 2020b; Nasir & D'Aloisio 2020). The underdensities around long dark gaps traced by Lyα emitting galaxies (LAEs) are also consistent with a late reionization model wherein reionization ends at $z < 6$ (Becker et al. 2018; Kashino et al. 2020; Christenson et al. 2021).

These Lyα forest and LAE results are potentially consistent with an alternative scenario wherein the IGM is ionized by $z = 6$ but retains large-scale fluctuations in the ionizing UV background down to lower redshifts (Davies et al. 2018a). On the other hand, recent measurements of the mean free path of ionizing photons measured at $z = 5.1$ and 6.0 (Becker et al. 2021) are difficult to reconcile with models wherein reionization ends at $z > 6$, and may instead prefer models wherein the IGM is still 20% neutral or more at $z = 6$ (Becker et al. 2021; Cain et al. 2021; Davies et al. 2021). In addition, a reionization

ending at $z < 6$ is consistent with models that reproduce a variety of observations (e.g., Weinberger et al. 2019; Choudhury et al. 2021; Qin et al. 2021).

One way of searching for signatures of late ($z_{\rm end} < 6$) reionization in the Lyα forest is by investigating dark gaps, i.e., contiguous regions of strong absorption (e.g., Songaila & Cowie 2002; Furlanetto et al. 2004; Paschos & Norman 2005; Fan et al. 2006; Gallerani et al. 2008; Gnedin et al. 2017; Nasir & D'Aloisio 2020). In Zhu et al. (2021, hereafter Paper I), we explored long dark gaps in the Lyα forest and found that a fully ionized IGM with a homogeneous UVB is strongly ruled out down to $z \simeq 5.3$. In contrast, late reionization models and a model wherein reionization ends by $z \sim 6$ but retains large-scale UVB fluctuations are consistent with the observations. Predictions for the Lyα dark gap statistics are similar between the two types of models. This is largely because realistic late reionization models also include UVB fluctuations, which are often associated with the neutral islands. Lyα also tends to saturate at a relatively low ($x_{\rm HI} \sim 10^{-3}$) neutral fraction, limiting its sensitivity to neutral gas.

Given its lower optical depth [2], Lyβ should be a more sensitive probe of neutral gas in the $z \lesssim 6$ IGM. As a result, ultra-late reionization models wherein neutral islands persist down to $z < 5.5$ may produce more long Lyβ dark gaps than can be produced by UVB fluctuations alone. Based on this feature, we can potentially use dark gaps in the Lyβ forest to place stronger constraints on the timing of reionization and distinguish the late reionization models from the pure fluctuating UVB models. As presented in Nasir & D'Aloisio (2020), distributions of dark gaps in the Lyβ forest differ between these models

[2]Lyβ absorption has a shorter wavelength ($\lambda_{{\rm Ly}\beta} = 1025.72\,$Å) and a lower oscillator strength ($f_{{\rm Ly}\beta} = 0.0791$) compared to those of Lyα absorption ($\lambda_{{\rm Ly}\alpha} = 1215.67\,$Å, $f_{{\rm Ly}\alpha} = 0.4164$). The ratio of optical depth is given by $\tau_{{\rm Ly}\beta}/\tau_{{\rm Ly}\alpha} = (f_{{\rm Ly}\beta}\lambda_{{\rm Ly}\beta})/(f_{{\rm Ly}\alpha}\lambda_{{\rm Ly}\alpha}) \simeq 0.16$.

most strongly on scales of $L \gtrsim 10\ h^{-1}\,\mathrm{Mpc}$. We are therefore particularly interested in these long dark gaps.

In this work, we use 42 high-quality QSO spectra that allow us to search for dark gaps in the Lyβ forest over the redshift range $5.5 < z < 6.0$. The sample includes high-quality X-Shooter spectra from the XQR-30 VLT large program [3] (D'Odorico et al., in prep.). In addition to comparing the results to model predictions, we also constrain $\langle x_{\rm HI} \rangle$ based on a conservative relationship between dark gap length and neutral fraction derived from simulations.

This paper is organized as follows. In Section 3.3 we describe the data and results from the observations. Section 3.4 compares our results to model predictions, discusses the implications for reionization, and infers constraints on $x_{\rm HI}$. Finally, we conclude the findings in Section 3.5. Throughout this paper, we assume a ΛCDM cosmology with $\Omega_{\rm m} = 0.308$, $\Omega_\Lambda = 0.692$, and $h = 0.678$ (Planck Collaboration et al. 2014). Distances are quoted in comoving units unless otherwise noted.

3.3 The Data and Results

3.3.1 QSO Spectra

Our sample includes the 42 out of 43 spectra of QSOs at $5.77 \lesssim z_{\rm em} \lesssim 6.31$ that were used for Paper I. The exception is PSO J004+17, whose spectrum has lower S/N that does not meet the requirement of our flux threshold for the Lyβ forest (Section 3.3.3). The spectra are taken with the Echellette Spectrograph and Imager (ESI) on Keck (Sheinis

[3] https://xqr30.inaf.it

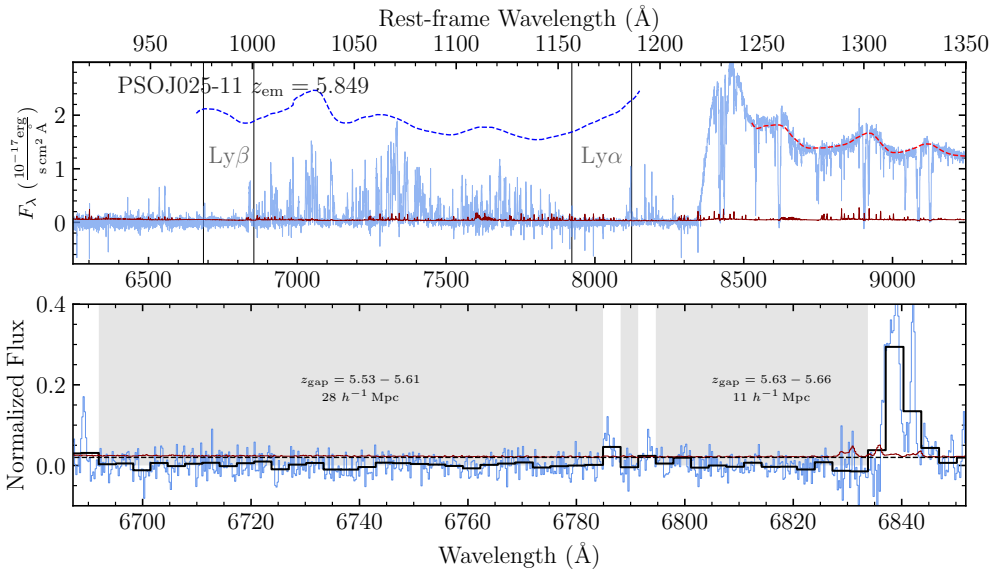


Figure 3.1: **Top panel:** Spectrum and continuum fits for the $z_{\rm em} = 5.849$ QSO PSO J025−11. The light blue and dark red lines represents flux and flux error in the original binning. Dashed curves redward and blueward of the Lyα peak show the best-fitting and predicted QSO continuum in absence of absorption, respectively. The fitting and prediction are based on Principal Component Analysis (PCA). The continuity between the Lyα forest and the Lyβ forest continuum is broken intentionally near 1020 Å in the rest frame. We label the wavelength range over which we search for dark gaps in the Lyβ forest and its corresponding Lyα forest in redshift. **Bottom panel:** Lyβ forest and dark gaps detected. The dashed black line labels the flux threshold of 0.02. The thick black line displays the flux binned to 1 h^{-1} Mpc. Light blue and dark red lines show the flux and flux error in the original binning. Dark gaps detected are shaded with gray. We also label the redshift range and length of each long ($L \geq 10\ h^{-1}$ Mpc) dark gap.
(The complete figure set [42 images] is available online at https://ydzhuastro.github.io/lyb.html.)

et al. 2002) and the X-Shooter spectrograph on the Very Large Telescope (VLT; Vernet et al. 2011). Among these, 19 X-Shooter spectra are from the XQR-30 VLT large program (D'Odorico et al., in prep). The details of the data reduction procedures are given in Paper I and Becker et al. (2019). We note that the targets are selected without foreknowledge of dark gaps in the Lyβ forest. Figure Set 3.1 displays the spectra, continuum fits, and dark gaps detected in the Lyβ forest for each QSO (see details below). An example is given in Figure 3.1.

3.3.2 Continuum Fitting

We use Principal Component Analysis (PCA) to predict the unabsorbed QSO continuum and normalize the transmission in the Lyβ forest. We follow a similar method as described in Paper I to fit and predict the continuum. Briefly, we implement and apply the log-PCA method of Davies et al. (2018c) in the Lyα and Lyβ forest portion of the spectrum following Bosman et al. (2022). The continuity between the Lyα forest and the Lyβ forest continuum is broken intentionally to correct for the effect of overlapping Lyα absorption in the Lyβ forest in the PCA training sample. We fit the red-side (rest-frame wavelength $\lambda_0 > 1230$ Å) continuum up to 2000 Å in the rest frame for X-Shooter spectra with NIR observations. The ESI spectra are fit using an optical-only PCA, which is presented in Bosman et al. (2021). The Lyβ dark gap detection is not very sensitive to the continuum, and we also test that using a power-law continuum does not significantly change the dark gap results in this work.

3.3.3 Dark Gap Detection

Similar to the definition of a dark gap in the Lyα forest in Paper I, we define a dark gap in the Lyβ forest to be a continuous spectral region in which all pixels binned to 1 h^{-1} Mpc have an observed normalized flux $F = F_{\rm obs}/F_{\rm c} < 0.02$, where $F_{\rm obs}$ is the observed flux and $F_{\rm c}$ is the continuum flux. The minimum length of a dark gap is 1 h^{-1} Mpc following Paper I. A bin size of 1 h^{-1} Mpc enables us to retain significant transmission profiles while reducing the influence of very small peaks on dark gap detection. The precise bin size should have relatively little impact on our results provided that the observations and mock spectra are treated consistently. We have verified that using bin sizes of 0.5 h^{-1} Mpc or 1.5 h^{-1} Mpc does not change our major conclusions, although the lengths of some dark gaps would change. A flux threshold of 0.02 is used here instead of 0.05, which we used for the Lyα gaps, because spectra in this sub-sample have higher signal-to-noise (S/N) levels. In addition, the Lyβ forest at the redshifts that we are interested in is less contaminated by sky lines than the Lyα forest. We have tested that using a threshold of 0.05 will not change our results fundamentally, although the difference between the models (Section 3.4) may become less significant. In order to reduce false detections caused by foreground Lyα absorption, we further require that all Lyβ dark gaps correspond to Lyα dark gaps as defined in Paper I over the same redshifts for both the observed and mock (Section 3.4.1) spectra. That is to say, each 1 h^{-1} Mpc bin in the Lyβ dark gap also has a normalized flux less than 0.05 in the Lyα forest at the same redshift. [4] For reference, we present the relationship between Lyβ dark gaps and Lyα dark gaps in Appendix E.

[4] Based on our test, whether requiring gaps to be also dark in the Lyα forest or not only affects a small fraction of gaps and does not change the results in this paper significantly. Although this requirement may not remove all false detection, it partially avoids contamination from random foreground density fluctuations.

For each QSO sightline, dark gaps are detected from 976 Å in the rest frame to 11 proper-Mpc blueward of the QSO's redshift, which corresponds to approximately 1000 Å in the rest frame. The lower wavelength limit ensures that the detection is not affected by the Lyγ absorption. We use the higher limit to avoid the QSO proximity zone transmission, and the cut is comparable to the choice in, e.g., Bosman et al. (2021). Following Paper I, we also exclude from the statistical analysis an additional 10 h^{-1} Mpc buffer zone blueward of the proximity zone cut. This allows the pixel at the red end of each sightline to intersect a possible long ($L \geq 10\ h^{-1}$ Mpc) dark gap, and hence helps to mitigate potential truncation issues. [5]

To avoid the contamination from sky line subtraction residuals, we mask out ± 75 km s^{-1} intervals of the spectra centered at sharp peaks in the flux error array when searching for dark gaps. The exception is that we do not mask transmission with a $> 3\sigma$ detection. In addition, we make no attempts to correct for the effects of damped Lyα systems (DLAs) or metal-enriched absorbers, although known systems in a sub-sample of the spectra with a relatively complete identification of metal-enriched systems are discussed in Appendix F. Figure 3.2 displays an overview of dark gaps detected in the Lyβ forest from our sample.

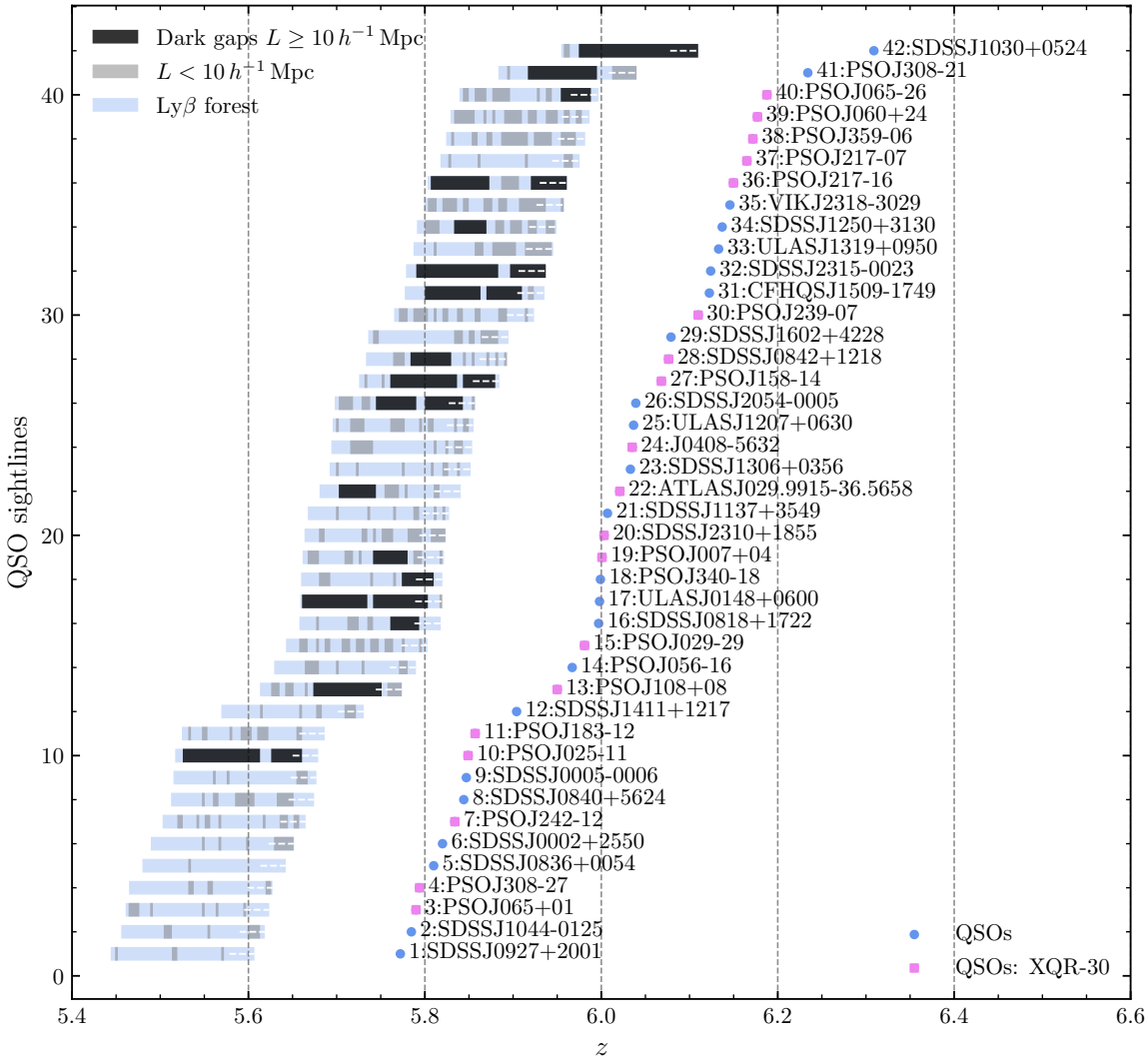
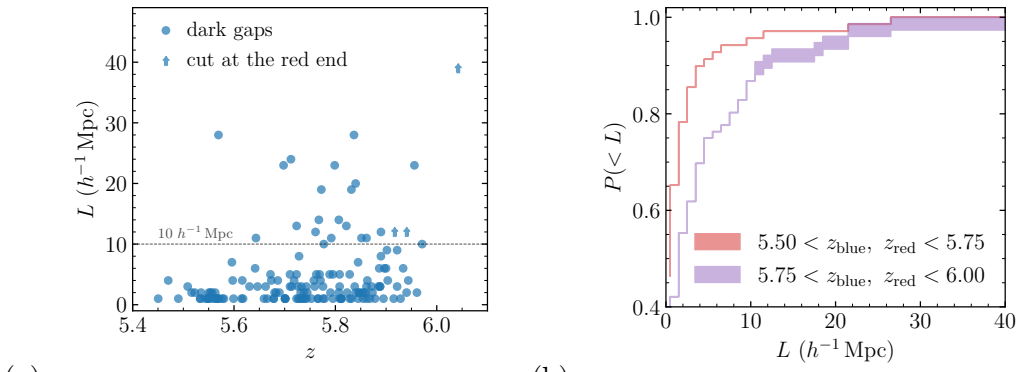
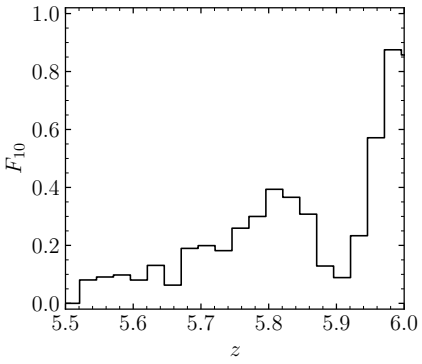


Figure 3.2: Overview of dark gaps identified in the Lyβ forest from our sample of 42 QSO spectra. Black (gray) bars represent dark gaps longer (shorter) than 10 h^{-1} Mpc. Pink squares label redshifts of XQR-30 targets, and blue dots mark the redshifts of the rest of QSOs. Light blue shaded regions demonstrate the redshift coverage of the Lyβ forest. We note that the Lyβ forest is truncated at 11 pMpc from the QSO. The Lyβ forest shown in this figure includes the 10 h^{-1} Mpc buffer zone labeled with dashed white line at the red end.



(a) (b)

Figure 3.3:



(c)

Figure 3.4: Observed Lyβ dark gaps. **(a)** Gap length versus central redshift. Dark gaps cut at the red end by the proximity zone cut are labeled with arrows. **(b)** Cumulative distribution of dark gap length for two redshift intervals. The upper/lower bounds of the shaded region are described in Section 3.3.3. **(c)** The fraction of QSO spectra showing dark gaps longer than 10 h^{-1} Mpc as a function of redshift, F_{10}. F_{10} is plotted with a binning of $\Delta z = 0.025$. See text and Appendix G for discussions about the drop of F_{10} near $z = 5.9$. (The data used to create this figure are available online at `https://ydzhuastro.github.io/lyb.html`.)

Dark Gaps Statistics

Figure 3.4 displays the statistical properties of dark gaps detected in the Lyβ forest from our sample. We detect 195 dark gaps in total, of which 24 have $L \geq 10\ h^{-1}\,\mathrm{Mpc}$. Panel (a) plots length versus central redshift of these dark gaps. Long dark gaps become less common as redshift decreases. Nevertheless, some long gaps still exist down to $z < 5.6$.

We calculate the cumulative distribution function of dark gap length, $P(< L)$. Dark gaps are sorted into two redshift bins according to their redshifts at both ends. For distributions that include dark gaps cut at the red end by the proximity zone limit, we calculate a lower bound on $P(< L)$ by assuming a infinite length for these gaps, and an upper bound by assuming the gap length that would appear in the absence of any proximity effect is the same as the one measured. As shown in Figure 3.4 (b), longer dark gaps become more numerous over $5.75 < z < 6.00$ compared to $5.50 < z < 5.75$. This significant evolution of $P(< L)$ is consistent with the results shown in panel (a).

Following Paper I, we quantify the spatial coverage of large Lyβ-opaque regions by calculating the fraction of QSO spectra showing long ($L \geq 10\ h^{-1}\,\mathrm{Mpc}$) dark gaps as a function of redshift, $F_{10}(z)$. Here we use $10\ h^{-1}\,\mathrm{Mpc}$ as the threshold because dark gaps longer than this in the late reionization models (see Section 3.4) are dominated by those containing neutral islands. Based on our tests, the number of dark gaps predicted by different models differs the most for $L \gtrsim 10\ h^{-1}\,\mathrm{Mpc}$, as also suggested by Nasir & D'Aloisio (2020).

[5] Without this additional buffer zone, it is possible that the F_{10} (Section 3.3.3) is underestimated near the red end of a sightline, since there can exist otherwise $\geq 10\ h^{-1}\,\mathrm{Mpc}$ gaps that are truncated by the edge or peaks in the proximity zone.

We calculate F_{10} at each redshift and average over $\Delta z = 0.025$ bins. As shown in Figure 3.4 (c), F_{10} has a significant redshift evolution over $5.5 < z < 6.0$. It increases rapidly with redshift over $5.5 < z < 5.8$, from $\sim 10\%$ to $\sim 40\%$, and climbs up to $\sim 80\%$ by $z = 6.0$ after a drop at $z \sim 5.9$. The reason of the drop is unclear, but the limited number of QSO sightlines may produce large statistical fluctuations (Appendix G), as also shown in the model predictions in Section 3.4. For comparison, we compute F_{01}, the fraction of QSO sightlines exhibiting dark gaps with $L \geq 1\ h^{-1}\,\mathrm{Mpc}$, and find no significant drop near $z = 5.9$ (see Appendix G).

Long Dark Gap toward PSO J025−11

We find a particularly interesting Lyβ gap toward the $z = 5.849$ QSO PSO J025−11 (Figure 3.1). This dark gap is within the redshift interval of a long ($68\ h^{-1}\,\mathrm{Mpc}$) trough in the Lyα forest. It spans $5.526 \leq z \leq 5.613$ with a length of $28\ h^{-1}\,\mathrm{Mpc}$. This is longer and extending to a even lower redshift than the 19 and $23\ h^{-1}\,\mathrm{Mpc}$ Lyβ troughs with $z_{\mathrm{min}} = 5.66$ within the extreme ($110\ h^{-1}\,\mathrm{Mpc}$) Ly$\alpha$ trough over $5.523 \leq z \leq 5.879$ toward ULAS J0148+0600 (Becker et al. 2015). This dark gap toward J025−11 contains no apparent transmission peaks, even in the unbinned data. The 2σ lower limit of $\tau_{\mathrm{eff}} \geq 6.067$ measured over the complete trough indicates that it is highly opaque.

There is a possibility that part of the trough may be due to either Lyβ or foreground Lyα absorption from the circum-galatic medium (CGM) around intervening galaxies. In this case, corresponding metal lines may be present. We check for potential CGM absorption using the XQR-30 metal absorber catalog (Davies et al., in prep; see Appendix F for technique details). We find no intervening metal systems within the redshift range

of the gap. We note that this line of sight has a DLA near the redshift of the QSO, as evidenced by the damping wing at the blue edge of the Lyα proximity zone. The Lyβ gap described here is at a velocity separation of >3000 km s^{-1} from the QSO, however, and is unaffected by the DLA. The XQR-30 catalog does include a C IV system towards J025−11 at $z = 4.5138$, for which Lyα would fall at the blue end of the Lyβ trough. Overall, however, the general lack of metal absorbers associated with this long dark gap may suggest that the gap probes a low-density region. This would favor the association of highly opaque sightlines with galaxy underdensities, as seen in imaging surveys for galaxies surrounding long Lyα troughs (Becker et al. 2018; Kashino et al. 2020; Christenson et al. 2021).

We examine the possible role of metal-enriched absorbers more broadly in Appendix F, finding little evidence for a strong correlation with long Lyβ troughs. We also examined a sample of lower-redshift lines of sight in Appendix H, finding that metal-enriched absorbers in the foreground Lyα alone are unlikely to create such a long dark gap. We emphasize that this gap falls in redshift within a long Lyα trough spanning $5.461 \leq z \leq 5.674$ with $L = 68\ h^{-1}$ Mpc that does not appear to be affected by metal absorbers (Paper I). This combination of factors gives us confidence that the $L = 28\ h^{-1}$ Mpc dark gap is genuinely caused by IGM opacity [6].

[6] In an extreme case where this foreground absorber links two shorter dark gaps, although very unlikely, one of these two shorter dark gaps would still have a size of $L \sim 25\ h^{-1}$ Mpc.

3.4 Models and Discussion

3.4.1 Models and Mock Spectra

Here we compare our results to predictions from models based on hydrodynamical simulations. We use the following models, which were also used in Paper I:

1. the `homogeneous-UVB` model from the Sherwood Simulation Suite (Bolton et al. 2017), which uses a homogeneous Haardt & Madau (2012) UV background;

2. late reionization models wherein reionization ends at $z \lesssim 5.3$, including `K20-low-`$\tau_{\rm CMB}$, `K20-low-`$\tau_{\rm CMB}$`-hot`, `K20-high-`$\tau_{\rm CMB}$ models from Keating et al. (2020a), and `ND20-late-longmfp`, `ND20-late-shortmfp` models from Nasir & D'Aloisio (2020); and

3. an early reionization model wherein the IGM is mostly ionized by $z = 6$ but large scale fluctuations in the UVB, which are amplified by a short mean free path of ionizing photons ($\lambda_{\rm mfp}^{912} = 10\ h^{-1}\,{\rm Mpc}$ at $z = 5.6$), persist down to lower redshifts (`ND20-early-shortmfp`, Nasir & D'Aloisio 2020).

These models were chosen, in part, because they reproduce at least some other observations of the Lyα forest. The `homogeneous-UVB` model agrees well with observations at $z < 5$ including the IGM temperature and flux power spectra (Bolton et al. 2017), although it fails to predict the Lyα opacity distribution at $z > 5$ (e.g., Bosman et al. 2022). The late reionization and fluctuating UVB models are broadly consistent with observations of IGM temperature, mean Lyα transmission, and fluctuations in Lyα opacity over the redshift range we are interested in (Keating et al. 2020a; Nasir & D'Aloisio 2020). Moreover, these models are able to produce long Lyα troughs at $z < 6$ (e.g., Paper I). We note that,

nevertheless, that none of the models we use can self-consistently predict the mean free path evolution over $5 < z < 6$ as measured in Becker et al. (2021).

In the `homogeneous-UVB` model, the IGM is instantaneously reionized at $z = 15$. At $z < 6$, therefore, the IGM in this model is fully ionized and the gas is hydrodynamically relaxed. A homogeneous UVB model that produced a later reionization (e.g., Puchwein et al. 2019; Villasenor et al. 2021) would mainly alter the temperature and pressure smoothing at $z < 6$. These are small-scale effects, however, that should only minimally impact our measurements. We would generally expect that any homogeneous UVB model that reionizes by $z = 6$ would produce similar dark gap statistics as the Haardt & Madau (2012) UVB once the ionization rates at $z < 6$ are rescaled to reproduce the observed mean flux.

The `K20-low-`$\tau_{\rm CMB}$`-hot` model shares a similar reionization history with the `K20-low-`$\tau_{\rm CMB}$ model, but it has a different thermal history with a volume-weighted mean temperature at the mean density at $z = 6$ of $T_0 = 10,000$ K compared to that of the latter being 7000 K. The `K20-high-`$\tau_{\rm CMB}$ model has an earlier mid-point of reionization at $z = 8.4$, which is at the upper end of the value suggested by CMB measurements (Planck Collaboration et al. 2020). As for the late reionization models from Nasir & D'Aloisio (2020), the main difference is that the `ND20-late-shortmfp` model includes stronger post-reionization UVB fluctuations than the `ND20-late-longmfp` model, while they both have neutral islands surviving at $z < 6$. The mean free path of ionizing photons at $z = 5.6$ in these two models are $\lambda_{\rm mfp}^{912} = 10$ and 30 h^{-1} Mpc, respectively.

The box sizes we use in this work are $L = 160$, 160, and 200 h^{-1} Mpc, for simulations in Bolton et al. (2017), Keating et al. (2020a), and Nasir & D'Aloisio (2020),

respectively. We note that the `K20` models are from radiative transfer simulations run in post-processing and that the `ND20` models are semi-numeric models. For more details on the models, see Paper I.

We rescale the optical depths in the simulations as needed in order to match the observed mean flux in the Lyα forest (see, e.g., §2.2 in Bolton et al. 2017, and references therein). We scale to the measurements of Bosman et al. (2018), which are consistent with the mean Lyα fluxes obtained from our sample. The same rescaling factor is then applied to both the Lyα and corresponding Lyβ optical depths. We note that this rescaling mainly applies to the `homogeneous-UVB` model, for which scaling by factors of $\sim$0.4$-$0.6 is required over $5 < z < 6$. We are therefore explicitly testing only a `homogeneous-UVB` model that also matches the observed mean Lyα flux. The Keating et al. (2020a) models already produce a mean Lyα transmission consistent with the measurements of Bosman et al. (2018). The mock spectra from this simulation are continuous in redshift, with a smoothly evolving mean flux. Nasir & D'Aloisio (2020) also calibrated their simulations to the observed $\tau_{\rm eff}$ from Bosman et al. (2018) but provide one-dimensional skewers at discrete redshifts. For these simulations we therefore only need to rescale the optical depths such that the mock spectra described below have a mean flux that evolves smoothly with redshift.

We derive dark gap predictions from forward-modeled mock spectra that are matched to the observed sample in QSO redshift, resolution, and S/N. Keating et al. (2020a) provide mock spectra of the Lyβ forest including the foreground Lyα absorption. For the `homogeneous-UVB` model and models from Nasir & D'Aloisio (2020) we follow the methods

described in Paper I to build the mock Lyβ forest and foreground Lyα forest. In all cases we re-bin the mock spectra and apply the noise arrays according to each observed spectrum.

3.4.2 Model Comparisons

Comparisons of F_{10}

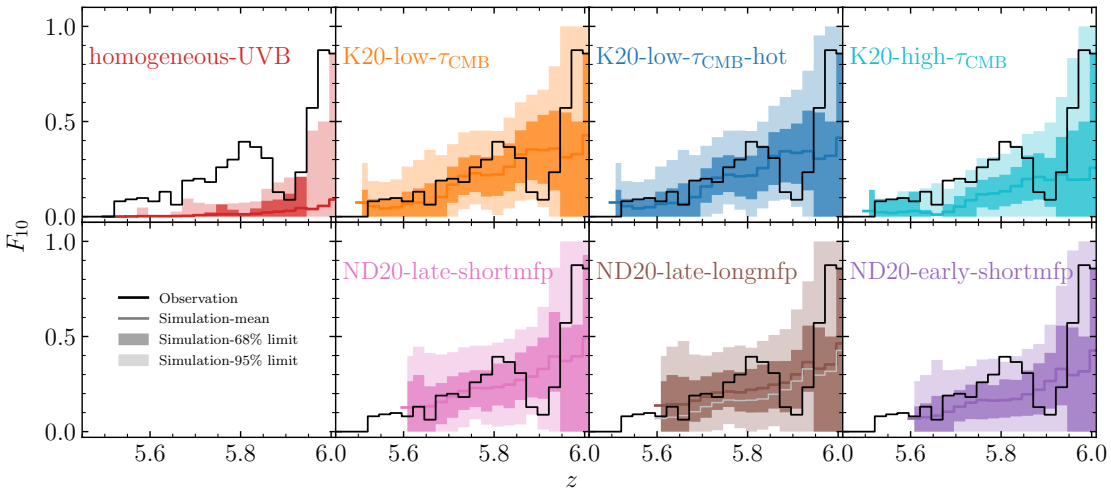


Figure 3.5: Fraction of QSO spectra showing long ($L \geq 10\ h^{-1}\,$Mpc) Lyβ dark gaps as a function of redshift predicted by different models. The colored lines, dark-shaded regions, and light-shaded regions show the mean, 68%, and 95% limits of F_{10} predicted by models based on 10,000 realizations. The black lines plot F_{10} from the observations. For comparison, we also overplot the mean F_{10} predicted by the `ND20-early-shortmfp` in the panel of the `ND20-late-longmfp` model with a gray line.

We compute the predicted F_{10} of each model based on 10,000 randomly selected sets of mock spectra of the same size as the observed sample. Their mean, 68%, and 95 % limits are plotted in Figure 3.5, along with the observations. Similar to F_{30} [7] for the Lyα

[7] F_{30} is defined as the fraction of QSO spectra exhibiting gaps longer than 30 $h^{-1}\,$Mpc as a function of redshift.

forest in Paper I, F_{10} shows jagged features due to the combined effects of step changes in the number of sightlines with redshift and the quantization of F_{10} for a finite sample size. We note that 68 and 95 percentile limits can share their upper and/or lower bounds at some redshifts, for the same reason. These features, on the other hand, show the constraining ability of the current sample size and data quality. The drop at $z \sim 5.9$ seen in the observed F_{10} is also broadly included in the 95% limits for most of the models.

The `homogeneous-UVB` model is not supported by the data at the $\geq 95\%$ level. This is consistent with the conclusion based on the Lyα forest in Paper I that a fully ionized IGM with a homogeneous UVB scenario is disfavored by the data at $z < 6$ down to $z \simeq 5.3$. In contrast, the late reionization models are still consistent with the data, except for the `K20-high-`$\tau_{\rm CMB}$ model, which covers the observed F_{10} just within its 95% upper limit. This supports the conclusion of Paper I that this extended reionization model is less favored by the data due to its insufficient neutral hydrogen and/or UVB fluctuations at $z < 6$.

Our results further show that dark gaps in Lyβ are more sensitive probes of neutral regions than gaps in Lyα. For dark gaps in the Lyα forest, we see little difference between the `ND20-early-shortmfp` model and the `ND20-late` models (Paper I). In the Lyβ forest, however, the former predicts a smaller F_{10} than the latter by ~ 0.05 at most redshifts. This difference is not enough for us to distinguish them based on the current sample, although the Lyβ gaps put some pressure on the early reionization model. Nasir & D'Aloisio (2020) note that these models are different in their Lyβ dark gap length distributions while they cannot be distinguished in the Lyα forest. Nevertheless, we compute the differential dark gap length distribution for individual ΔL bins, $L\Delta P(L)/\Delta L$, and find their differences are

minor compared to the scatter of the data. Looking ahead to the era of ELTs, we forecast that ~ 100 lines of sight with the Lyβ forest covering $z \sim 5.8$ would be needed to distinguish the two models at $\sim$95% confidence based on F_{10}. A signal-to-noise ratio of 50 per 10 km s^{-1} for the spectra would be adequate according to our tests using mock spectra.

Total Number of Long Dark Gaps at $z \leq 5.8$

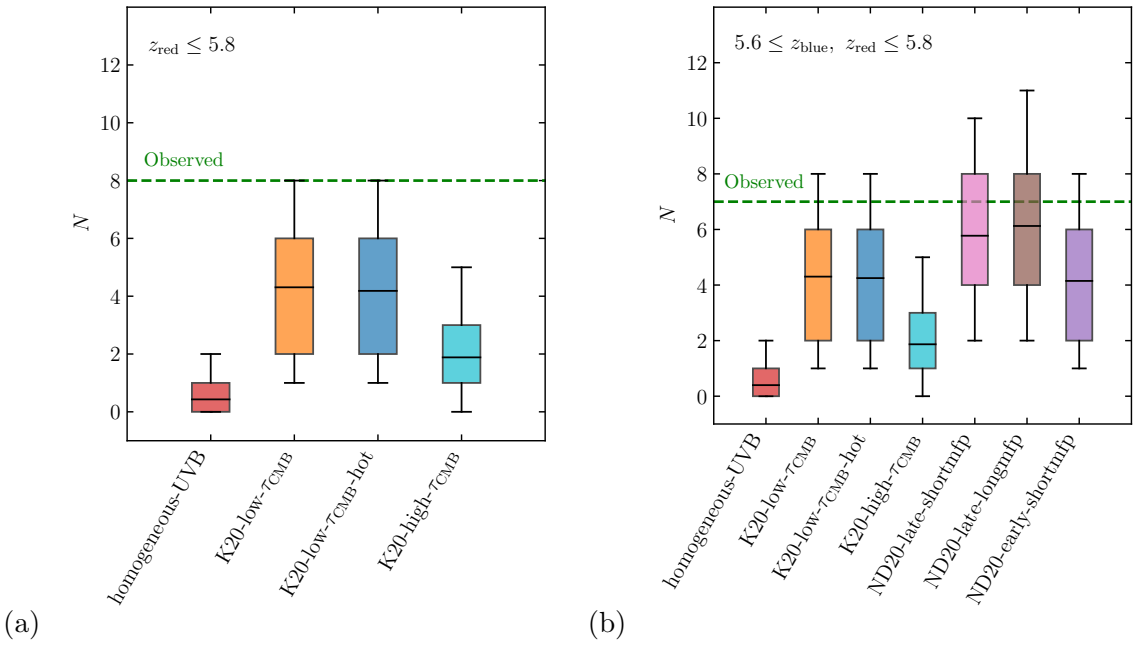


Figure 3.6: **(a)** Number of long ($L \geq 10\ h^{-1}$ Mpc) Lyβ dark gaps that lie entirely below $z = 5.8$ in the mock sample from each model. **(b)** Number of long dark gaps entirely over $5.6 \leq z \leq 5.8$. In both panels the center lines, boxes, and error bars show the mean, 68% limits, and 95% limits, respectively. The observed numbers of long dark gaps in each redshift range are indicated by dashed green lines.

To further illustrate the differences between models, we use our mock data to calculate the total number of long dark gaps predicted to lie entirely at $z < 5.8$. Figure 3.6

compares the model results to the observations. Given that the `ND20` models only have outputs down to $z = 5.6$, we exclude these models when counting the total number of long dark gaps below $z = 5.8$. We nevertheless include the `ND20` models for dark gaps that fall entirely over $5.6 \leq z \leq 5.8$ for reference.

The results are consistent with those from F_{10}. As shown in Figure 3.6 (a), the 95% upper limit of the predicted number of long dark gaps by the `homogeneous-UVB` model is 2. This is a factor of 4 smaller than the observed value, which is 8. The `K20-high-`$\tau_{\rm CMB}$ model is also disfavored by the data at $> 95\%$ confidence given its deficit of long dark gaps. In contrast, the rapid late reionization models, i.e. `K20-low-`$\tau_{\rm CMB}$ models, agree with the observations within their 95% limits.

Over $5.6 \leq z \leq 5.8$ the observed number of long dark gaps decreases by one while the simulation predictions have little change. In this case, rapid late reionization models from Keating et al. (2020a) are still consistent with the data. The observations also support both fluctuating UVB and late reionization models from Nasir & D'Aloisio (2020). We note that the difference between the predicted mean numbers and the observed value is smallest for the `ND20-late` models, wherein $\langle x_{\rm HI} \rangle$ is still higher than $\sim 5\%$ by $z = 5.6$. On the other hand, the `K20-high-`$\tau_{\rm CMB}$ model is disfavored by the data also in this redshift range. This would suggest that very extended reionization scenarios in which insufficient neutral hydrogen and/or UVB fluctuations remain at $z < 6$ may be disfavored.

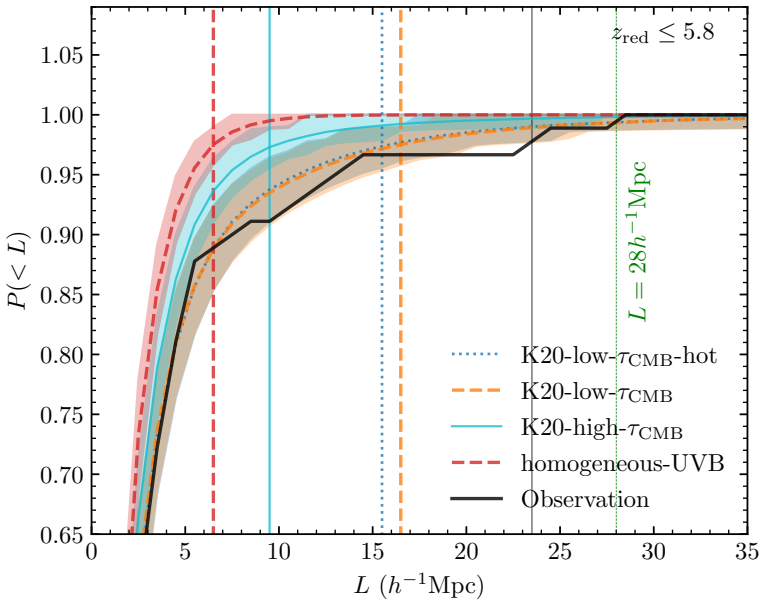


Figure 3.7: Cumulative distribution of dark gaps that are entirely below $z = 5.8$. Vertical colored lines indicate the 97.5 percentile limit for each model. The color shaded regions plot the 68% limit for each model. From left to right are the `homogeneous-UVB` model, the `K20-high-`$\tau_{\rm CMB}$ model, and the `K20-low-`$\tau_{\rm CMB}$`-hot` model (almost completely overlapped with the `K20-low-`$\tau_{\rm CMB}$ model), respectively.

Detection Rate of an $L \geq 28\ h^{-1}\,\mathrm{Mpc}$ Dark Gap

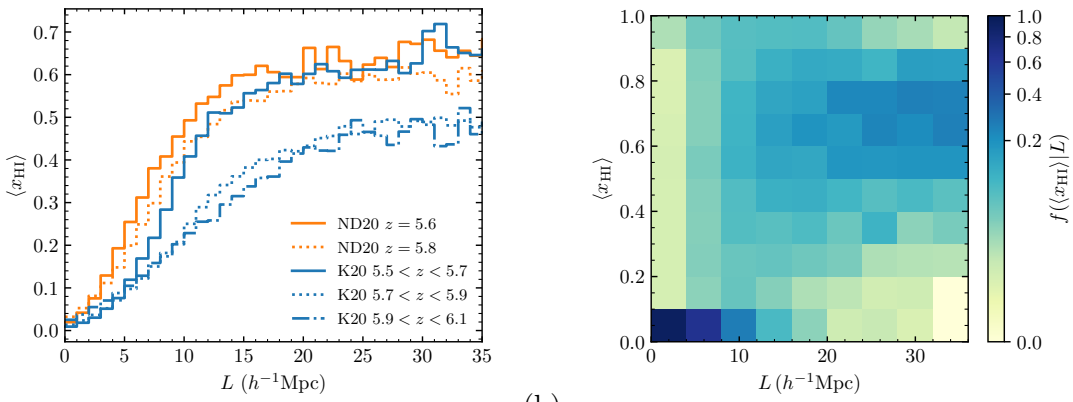
Perhaps the most conspicuous feature in the observations is the $L = 28\ h^{-1}\,\mathrm{Mpc}$ dark gap toward PSO J025−11 that extends down to $z \simeq 5.5$. The appearance of this gap may indicate that significant neutral hydrogen islands and/or UVB fluctuations persist down to $z \simeq 5.5$, and provide further leverage for discriminating between models. As the outputs of the `ND20` models have no redshift coverage for this dark gap, we only compare the `K20` models and the `homogeneous-UVB` model for this section.

For each model, we use 10,000 bootstrapping realizations to calculate the cumulative distribution function (CDF) of dark gap length, $P(< L)$. Figure 3.7 compares the observed and predicted $P(< L)$ for dark gaps that are entirely below $z = 5.8$. As indicated by the vertical lines, the observed dark gap with $L = 28\ h^{-1}\,\mathrm{Mpc}$ is well beyond the 95% limits of all the models shown here. These results suggest that the $L = 28\ h^{-1}\,\mathrm{Mpc}$ gap we observed in the Lyβ forest toward PSO J025−11 is extremely rare in these models. We perform Mann-Whitey U tests (Mann & Whitney 1947) for the hypotheses that the distributions of L in the data and predicted by models are equal, for each model respectively. The hypothesis is rejected with p-values < 0.05 for the `homogeneous-UVB` model.

We further calculate the detection rate of at least one $L = 28\ h^{-1}\,\mathrm{Mpc}$ gap entirely below $z = 5.61$ in the mock samples from each model with the required redshift coverage. We note that in the data there are 10 QSO spectra where the Lyβ forest (excluding the proximity zone) covers the full central redshift range of the $L = 28\ h^{-1}\,\mathrm{Mpc}$ dark gap. We find zero detections of such long dark gaps in the `homogeneous-UVB` model out of 10,000 trails. The `K20-high-`τ_{CMB} model yields a detection rate of 4%. Both the `K20-low-`τ_{CMB}

and `K20-low-`$\tau_{\rm CMB}$`-hot` models give higher detection rate of 10%. These results suggest that in the case of a late reionization, models with a volume-weighted average neutral hydrogen fraction $\langle x_{\rm HI} \rangle \gtrsim 5\%$ at $z = 5.6$ are more consistent with the observations. In addition, the relatively rare presence of $L \geq 28\ h^{-1}\,$Mpc gaps in the models could also be related to the size of the simulation volume ($160\ h^{-1}\,$Mpc for the `K20` simulations). Simulations run in larger volumes may be needed to compute more accurate statistics on the incidence of these rare, long troughs in late reionization models.

3.4.3 Neutral Hydrogen Fraction



(a) (b)

Figure 3.8: (a) The mean volume-weighted neutral fraction ($\langle x_{\rm HI} \rangle$) over a Ly$\beta$ dark gap for a given dark gap length in the mock data. In this figure, "ND20" and "K20" refer to the `ND20-late-longmfp` model and the `K20-low-`$\tau_{\rm CMB}$ model, respectively. (b) Distribution of $\langle x_{\rm HI} \rangle$ for a given Lyβ dark gap length, $f(x_{\rm HI}|L)$, in the `ND20-late-longmfp` model at $z = 5.6$. $f(x_{\rm HI}|L)$ is normalized for each L interval.

We can further use dark gaps to infer an upper limit on $\langle x_{\rm HI} \rangle$. One can set a strict upper limit on the neutral fraction by assuming that all dark gaps correspond to neutral

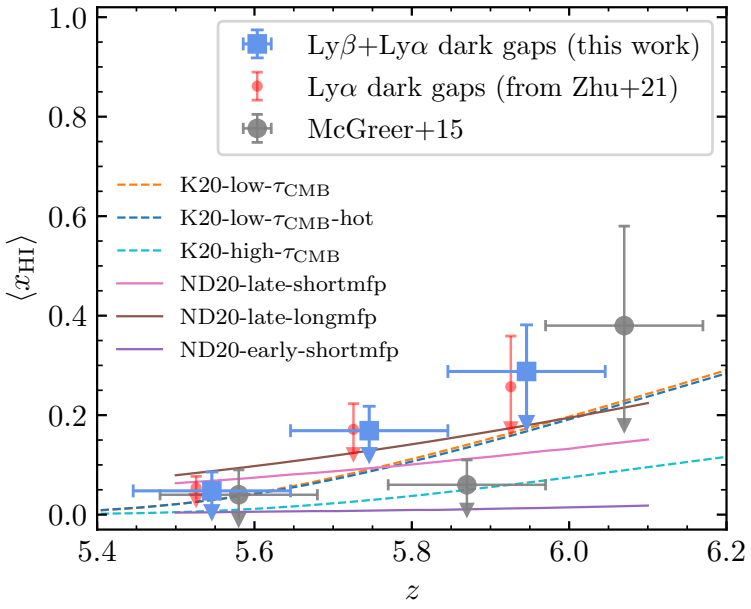


Figure 3.9: Inference on the neutral hydrogen fraction ($\langle x_{\rm HI} \rangle$) from Ly$\beta$ dark gaps, which are also dark in the Lyα forest by definition. We show constraints based on Lyα dark gaps from Paper I with red error bars, which are shifted by -0.02 in redshift for display. Gray markers plot the constraints based of fraction of pixels that are dark in both the Lyα and Lyβ forests from McGreer et al. (2015). The vertical error bars show the 68% (1σ) limits. The horizontal error bars indicate the $\Delta z = 0.2$ redshift bins. For reference, colored lines plot the redshift evolution of $\langle x_{\rm HI} \rangle$ for the reionization models used in this work.

gas (e.g., McGreer et al. 2011, 2015). At the end of reionization, however, a combination of density and UVB fluctuations will tend to produce dark gaps even once the gas is ionized. We therefore wish to use insights from reionization models to derive a more physically motivated but still conservative upper limit on $\langle x_{\rm HI} \rangle$ from the covering fraction of dark gaps. As described below, we use the fact that dark gaps in the late reionization models tend to show a correlation between the volume-averaged neutral fraction within a gap and the gap length. By applying this relationship to the observed gap length distribution we can set constraints on $\langle x_{\rm HI} \rangle$.

Our goal is to set physically reasonable constraints on $\langle x_{\rm HI} \rangle$ while minimizing the model dependency. We therefore wish to identify the maximum average neutral fraction for a given gap length that is allowed by the models. We first explore the distribution of neutral fractions for a given dark gap length, $f(x_{\rm HI}|L)$. We focus on two models wherein neutral regions contribute significantly to forming dark gaps, the `ND20-late-longmfp` model and the `K20-low-`$\tau_{\rm CMB}$ model. Using the mock data, we calculate $x_{\rm HI}$ for each dark gap by averaging the neutral fraction pixel-wise. Figure 3.8 (a) plots the mean neutral fraction of dark gaps as a function of length, $\langle x_{\rm HI} \rangle_L$, at different redshifts. It is related to $f(x_{\rm HI}|L)$ by

$$\langle x_{\rm HI} \rangle_L = \int_0^1 x_{\rm HI} f(x_{\rm HI}|L) dx_{\rm HI}. \tag{3.1}$$

As shown in the figure, dark gaps of a given length tend to be more neutral as redshift decreases. This is largely because the opacity of the ionized IGM tends to decrease with decreasing redshift, making it more difficult to produce long gaps through density and/or UVB fluctuations alone. In order to set conservative upper limits of $\langle x_{\rm HI} \rangle$ we adopt the $\langle x_{\rm HI} \rangle_L$ relationship from `ND20-late-longmfp` at $z = 5.6$. The normalized $f(x_{\rm HI}|L)$ for

each dark gap length interval is plotted in Figure 3.8 (b). This is similar to but slightly higher than the relationship from `K20-low-`$\tau_{\rm CMB}$ at the same redshift. We also note that the redshift evolution in $\langle x_{\rm HI}\rangle_L$ in these models is relatively modest, up to a factor of $\sim$2 in the `K20-low-`$\tau_{\rm CMB}$ model between $z \sim 6$ and 5.6.

In order to translate the observed gap length distribution into a $\langle x_{\rm HI}\rangle$ constraint, we calculate $\mathscr{F}_L$, the fraction of QSO spectra showing dark gaps with length L as a function of redshift. At a certain redshift, the total mean neutral hydrogen fraction is then given by

$$\langle x_{\rm HI}\rangle = \sum_{L=1}^{\infty} \mathscr{F}_L \langle x_{\rm HI}\rangle_L. \tag{3.2}$$

Here we use a sum for L instead of an integral because we measure dark gap lengths in increments of 1 h^{-1} Mpc. To estimate the uncertainty in $\langle x_{\rm HI}\rangle$, we randomly select the observed sightlines with replacement and calculate the corresponding $\mathscr{F}_L$. We use bootstrapping to randomly sample the neutral hydrogen fraction from $f(x_{\rm HI}|L)$ given by models and multiply by the observed $\mathscr{F}_L$ of this sample, then sum up for all dark gap lengths. The final uncertainty in $\langle x_{\rm HI}\rangle$ is calculated by repeating this process 10,000 times.

The results are shown in Figure 3.9. We calculate $\langle x_{\rm HI}\rangle$ in Equation (3.2) over $\Delta z = 0.2$ bins. The inferred upper limits on $\langle x_{\rm HI}\rangle$ are $0.05^{+0.04}_{-0.04}$, $0.17^{+0.05}_{-0.05}$, and $0.29^{+0.09}_{-0.10}$ at $z \simeq 5.55$, 5.75, and 5.95, respectively. We also calculate $\langle x_{\rm HI}\rangle$ following the same method based on Lyα dark gaps presented in Paper I, as shown with red symbols in Figure 3.9. The Lyα dark gaps yield $\langle x_{\rm HI}\rangle \leq 0.05$, 0.17, and 0.26 at $z \simeq 5.55$, 5.75, and 5.95, respectively. The measurements based on Lyα and Lyβ dark gaps are highly consistent with each other. Compared to the measurements based the fraction of dark pixels by McGreer et al. (2015), our results potentially allow a higher neutral fraction over $5.6 \lesssim z \lesssim 6.0$ and a later

reionization. The difference in $\langle x_{\rm HI} \rangle$ might be due to cosmic variance and/or the different definitions of dark gaps and dark pixels used in these works. The $\langle x_{\rm HI} \rangle$ measurement at $z \sim 5.9$ in McGreer et al. (2015), moreover, may be biased by transmission peaks in the QSO proximity zone given that their wavelength range for both the Lyα and Lyβ forests ends at $z_{\rm QSO} - 0.1$, which is less than 6.5 pMpc from the QSO at $z \sim 6$ (see proximity zone size measurements in e.g., Eilers et al. 2017, 2020).

3.5 Conclusion

In this work, we explore the IGM near the end of reionization using dark gaps in the Lyβ forest over $5.5 \lesssim z \lesssim 6.0$. We show that about 10%, 40%, and 80% of QSO spectra exhibit long ($L \geq 10\ h^{-1}\,$Mpc) dark gaps in their Lyβ forest at $z \simeq 5.6$, 5.8, and 6.0, respectively. Among these gaps, we detect a very long ($L = 28\ h^{-1}\,$Mpc) and dark ($\tau_{\rm eff} \gtrsim 6$) Lyβ gap extending down to $z \sim 5.5$ toward the $z_{\rm em} = 5.85$ QSO PSO J025−11.

A comparison between the observed Lyβ dark gap statistics for the whole sample of 42 lines of sight and predictions from multiple reionization models (Bolton et al. 2017; Keating et al. 2020a; Nasir & D'Aloisio 2020) confirms that evidence of reionization in the form of neutral islands and/or a fluctuating UV background persists down to at least $z \sim 5.5$. This supports the conclusions in Paper I and Bosman et al. (2022). In Paper I we noted a possible tension between Lyα gap statistics and a model wherein reionization ends by $z < 6$ but has a relatively early mid-point of $z = 8.4$ (Keating et al. 2020a). With Lyβ this tension becomes more significant ($> 95\%$ level) based on the count of long dark gaps at $z \leq 5.8$, suggesting that very extended reionization scenarios with insufficient remaining

neutral hydrogen and/or UVB fluctuations at $z < 6$ may be disfavored. In contrast, rapid late reionization models with $\langle x_{\rm HI} \rangle \gtrsim 5\%$ at $z = 5.6$ (Keating et al. 2020a; Nasir & D'Aloisio 2020) are consistent with the observations. A model wherein reionization ends early but retains large-scale fluctuations in the ionizing UV background (Nasir & D'Aloisio 2020) is also permitted by the dark gap data. We note, however, that recent IGM temperature measurements from Gaikwad et al. (2020) disfavor this model.

A caveat is that we are testing only specific reionization models, including only one with a fluctuating UVB in which reionization ends at $z > 6$. By comparison, Gnedin et al. (2017) showed that their full radiative transfer simulations, which reionized near $z \sim 7$, were able to reproduce the Lyα dark gap distribution measured from ESI spectra of a set of twelve $z \sim 6$ QSOs. Because Lyβ dark gaps are correlated with Lyα opacities (Appendix E), it is possible that some early reionization scenarios with UVB fluctuations can reproduce our Lyβ dark gap distributions while also matching the observed evolution of the mean Lyα transmission.

Finally, we use the observed Lyβ gaps to place constraints on the neutral hydrogen fraction based on the association between neutral islands and dark gaps seen in reionization simulations. Our results are broadly consistent with, but more permissive than the constraints from McGreer et al. (2011, 2015) that are based on the dark pixel fraction. Notably, we find an upper limit at $z \simeq 5.75$ of $\langle x_{\rm HI} \rangle \leq 0.17$. This constraint is consistent with scenarios wherein reionization extends significantly below $z = 6$.

Chapter 4

Probing Ultra-late Reionization: Direct Measurements of the Mean Free Path over $5 < z < 6$

4.1 Abstract

The mean free path of ionizing photons, $\lambda_{\rm mfp}$, is a critical parameter for modeling the intergalactic medium (IGM) both during and after reionization. We present direct measurements of $\lambda_{\rm mfp}$ from QSO spectra over the redshift range $5 < z < 6$, including the first measurements at $z \simeq 5.3$ and 5.6. Our sample includes data from the XQR-30 VLT large program, as well as new Keck/ESI observations of QSOs near $z \sim 5.5$, for which we also acquire new [C II] 158μm redshifts with ALMA. By measuring the Lyman continuum transmission profile in stacked QSO spectra, we find $\lambda_{\rm mfp} = 9.33^{+2.06}_{-1.80}$, $5.40^{+1.47}_{-1.40}$, $3.31^{+2.74}_{-1.34}$,

and $0.81^{+0.73}_{-0.48}$ pMpc at $z = 5.08$, 5.31, 5.65, and 5.93, respectively. Our results demonstrate that $\lambda_{\rm mfp}$ increases steadily and rapidly with time over $5 < z < 6$. Notably, we find that $\lambda_{\rm mfp}$ deviates significantly from predictions based on a fully ionized and relaxed IGM as late as $z = 5.3$. By comparing our results to model predictions and indirect $\lambda_{\rm mfp}$ constraints based on IGM Lyα opacity, we find that the $\lambda_{\rm mfp}$ evolution is consistent with scenarios wherein the IGM is still undergoing reionization and/or retains large fluctuations in the ionizing UV background well below redshift six. [1]

4.2 Introduction

When and how reionization occurred is a fundamental question about the early universe and the first galaxies. The appearance of transmitted flux in the Lyman-α (Lyα) forest of high-redshift QSOs (e.g., Fan et al. 2006) has long been interpreted as evidence that hydrogen in the intergalactic medium (IGM) was largely reionized by $z = 6$. In terms of the ionizing photon budget, however, an end of reionization at $z \geq 6$ is challenging to reconcile with a midpoint of $z \sim 7-8$ suggested by e.g., cosmic microwave background (CMB) observations (Planck Collaboration et al. 2020, see also de Belsunce et al. 2021). In particular, star-forming galaxies at $z > 6$ would have to emit ionizing photons extremely efficiently in order to complete reionization within such a short interval. This leaves two possibilities: the ionizing efficiency of galaxies at $z > 6$ is remarkably high, and/or reionization extends to lower redshifts.

[1] A version of this chapter is published as Zhu et al. (2023) — *Probing Ultralate Reionization: Direct Measurements of the Mean Free Path over* $5 < z < 6$*, The Astrophysical Journal, 955:115.*

A number of observations have now been used to constrain the timeline of reionization. These observations include the Lyα damping wing in $z \gtrsim 7$ QSO spectra (e.g., Bañados et al. 2018; Davies et al. 2018b; Wang et al. 2020; Yang et al. 2020a; Greig et al. 2022), the decline in observed Lyα emission from galaxies at $z > 6$ (e.g., Mason et al. 2018, 2019; Hoag et al. 2019; Hu et al. 2019, and references therein, but see Jung et al. 2020; Wold et al. 2021), and measurements of the thermal state of the IGM at $z > 5$ (e.g., Boera et al. 2019; Gaikwad et al. 2021). These observations support a midpoint of reionization at $z \sim 7-8$ and are generally consistent with an ending point at $z \sim 6$, as constrained by the fraction of dark pixels in the Lyα forest (e.g., McGreer et al. 2015, but see Zhu et al. 2022).

Other observations, however, suggest that reionization may have extended to significantly lower redshifts. Large-scale fluctuations in the measured Lyα effective optical depth[2] (e.g., Fan et al. 2006; Becker et al. 2015; Eilers et al. 2018; Bosman et al. 2018, 2022; Yang et al. 2020b), together with long troughs extending down to or below $z \simeq 5.5$ in the Lyα forest (e.g., Becker et al. 2015; Zhu et al. 2021), potentially indicate the existence of large neutral IGM islands (e.g., Kulkarni et al. 2019a; Keating et al. 2020b; Nasir & D'Aloisio 2020; Qin et al. 2021). This interpretation is further supported by dark gaps in the Lyβ forest (Zhu et al. 2022). Reionization extending to $z < 6$ is also consistent with the observed underdensities around long dark gaps traced by Lyα emitting galaxies (LAEs, Becker et al. 2018; Kashino et al. 2020; Christenson et al. 2021). In addition, such a late-ending reionization scenario is consistent with the evolution of metal-enriched absorbers at $z \sim 6$ (e.g., Becker et al. 2019; Davies et al. 2023b,a), as well as numerical models that

[2]Defined as $\tau_{\rm eff} = -\ln \langle F \rangle$, where F is the continuum-normalized transmission flux.

reproduce a variety of observations (e.g., Weinberger et al. 2019; Choudhury et al. 2021; Qin et al. 2021; Gaikwad et al. 2023).

A potentially decisive clue for establishing when reionization ended comes from recent measurements of the mean free path of ionizing photons ($\lambda_{\rm mfp}$). Becker et al. (2021, herein referred to as B21) found that the $\lambda_{\rm mfp}$ increases by a factor of around ten between $z = 6.0$ and 5.1, and the $\lambda_{\rm mfp}$ at $z = 6.0$ is about eight times shorter than what would be expected based on its evolution at $z \lesssim 5$ (see also constraints from Bosman 2021). Such a rapid evolution in the $\lambda_{\rm mfp}$ is expected to occur near the end of reionization due to (i) the growth and merger of ionized bubbles, and (ii) the photo-evaporation of dense, optically thick sinks (e.g., Shapiro et al. 2004; Furlanetto & Oh 2005; Sobacchi & Mesinger 2014; Park et al. 2016; D'Aloisio et al. 2020). Furthermore, the $\lambda_{\rm mfp}$ measurements in B21 are difficult to reconcile with models where reionization completes at $z > 6$, and may instead support models where the IGM is still $\gtrsim 20\%$ neutral at $z = 6$ (Becker et al. 2021; Cain et al. 2021; Davies et al. 2021).

Our understanding of how $\lambda_{\rm mfp}$ evolves over $5 < z < 6$ is highly incomplete, however. The measurements of B21 were restricted to $z \simeq 5.1$ and $\simeq 6.0$ by a lack of high-quality spectra at intermediate redshifts. This was due to a historical redshift gap in the discovery of QSOs near $z \sim 5.5$, which have overlap with brown dwarfs in their visible colors. This gap, however, has been filled by Yang et al. (2017, 2019b) using near- and mid-infrared photometry, making it possible to obtain a significant sample of high-quality $z \sim 5.5$ QSO spectra for the first time.

In this work, we report the first measurements of $\lambda_{\rm mfp}$ at multiple redshifts between $z = 6$ and 5. In addition to archival QSO spectra used in B21, our sample includes new QSO spectra from the XQR-30 VLT large program (e.g., D'Odorico et al. 2023; Bischetti et al. 2022; Bosman et al. 2022; Chen et al. 2022; Davies et al. 2023b; Lai et al. 2022; Satyavolu et al. 2023b; Zhu et al. 2021, 2022) as well as from new Keck/ESI observations. This paper is organized as follows. In Section 4.3, we describe the data and observations. Section 4.4 briefly introduces the methods we use to measure the $\lambda_{\rm mfp}$. We present our results and discuss their implications for reionization in Section 4.5. Finally, we summarize our findings in Section 4.6. Throughout this paper we assume a ΛCDM cosmology with $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$, and $h = 0.7$. Distances are quoted in proper units unless otherwise noted. We also use 912 Å to represent the Lyman limit wavelength of 911.76 Å.

4.3 Data and Observations

To measure $\lambda_{\rm mfp}$ over $5 < z < 6$, we employ a large sample of 97 spectra of QSOs at $4.9 < z < 6.1$. Our sample includes 23 LRIS spectra and 35 GMOS spectra of QSOs at $z \lesssim 5.3$ used in B21. For higher redshifts, we use 18 and 6 spectra from the Keck/ESI (Sheinis et al. 2002) and VLT/X-Shooter (Vernet et al. 2011) archives, respectively. We include 7 high-quality spectra with sufficient wavelength coverage from the XQR-30 VLT large program (D'Odorico et al. 2023). The rest of the data are new spectra of 8 QSOs near $z \sim 5.5$ from our ESI observations. A summary of our QSO sample is provided in Table 4.1.

Table 4.1: QSO Spectra Used for $\lambda_{\rm mfp}$ Measurements

QSO	RA (J2000)	DEC (J2000)	$z_{\rm qso}$	Ref	Instrument	M_{1450}	Ref	$R_{\rm eq}$ (pMpc)
J0015−0049	00:15:29.86	−00:49:04.3	4.931	a	LRIS	−25.2	α	$3.8^{+1.1}_{-0.8}$
J0256+0002	02:56:45.75	+00:02:00.2	4.960	a	LRIS	−24.6	α	$2.8^{+0.9}_{-0.6}$
J0236−0108	02:36:33.83	−01:08:39.2	4.974	a	LRIS	−25.0	α	$3.4^{+1.1}_{-0.7}$
J0338+0018	03:38:30.02	+00:18:40.0	4.988	a	LRIS	−25.1	α	$3.5^{+1.2}_{-0.7}$
J2226−0109	22:26:29.28	−01:09:56.6	4.994	a	LRIS	−24.6	α	$2.8^{+0.9}_{-0.5}$
SDSSJ1341+4611	13:41:41.46	+46:11:10.3	5.003	b	GMOS	−25.4	β	$4.0^{+1.3}_{-0.8}$
J0129−0028	01:29:07.45	−00:28:45.6	5.015	a	LRIS	−25.1	α	$3.5^{+1.2}_{-0.7}$
SDSSJ1337+4155	13:37:28.82	+41:55:39.9	5.018	b	GMOS	−26.6	β	$7.0^{+2.2}_{-1.4}$
J0221−0342	02:21:12.33	−03:42:31.6	5.024	a	LRIS	−24.9	α	$3.2^{+1.0}_{-0.7}$
SDSSJ0846+0800	08:46:27.84	+08:00:51.7	5.028	b	GMOS	−26.9	β	$8.1^{+2.6}_{-1.6}$
J2111+0053	21:11:58.02	+00:53:02.6	5.034	a	LRIS	−25.3	α	$3.9^{+1.2}_{-0.8}$
SDSSJ1242+5213	12:42:47.91	+52:13:06.8	5.036	b	GMOS	−25.7	β	$4.7^{+1.5}_{-1.0}$
J0023−0018	00:23:30.67	−00:18:36.6	5.037	a	LRIS	−25.1	α	$3.5^{+1.1}_{-0.7}$
SDSSJ0338+0021	03:38:29.31	+00:21:56.2	5.040	b	GMOS	−26.7	β	$7.4^{+2.3}_{-1.5}$
J0321+0029	03:21:55.08	+00:29:41.6	5.041	a	LRIS	−24.9	α	$3.2^{+1.0}_{-0.7}$
SDSSJ0922+2653	09:22:16.81	+26:53:59.1	5.042	b	GMOS	−26.0	β	$5.4^{+1.6}_{-1.1}$
SDSSJ1534+1327	15:34:59.76	+13:27:01.4	5.043	b	GMOS	−25.0	β	$3.4^{+1.0}_{-0.7}$
SDSSJ1101+0531	11:01:34.36	+05:31:33.9	5.045	b	GMOS	−27.7	β	$11.8^{+3.6}_{-2.5}$
SDSSJ1340+3926	13:40:15.04	+39:26:30.8	5.048	b	GMOS	−26.8	β	$7.8^{+2.4}_{-1.6}$
SDSSJ1423+1303	14:23:25.92	+13:03:00.7	5.048	b	GMOS	−27.1	β	$8.9^{+2.8}_{-1.8}$
SDSSJ1154+1341	11:54:24.73	+13:41:45.8	5.060	b	GMOS	−25.6	β	$4.5^{+1.3}_{-1.0}$
J1408+5300	14:08:22.92	+53:00:20.9	5.072	a	LRIS	−25.5	α	$4.3^{+1.3}_{-0.9}$
SDSSJ1614+2059	16:14:47.04	+20:59:02.8	5.081	b	GMOS	−26.6	β	$7.2^{+2.0}_{-1.6}$
J2312+0100	23:12:16.44	+01:00:51.6	5.082	a	LRIS	−25.6	α	$4.5^{+1.3}_{-1.0}$
J2239+0030	22:39:07.56	+00:30:22.5	5.092	a	LRIS	−25.2	α	$3.8^{+1.0}_{-0.9}$
SDSSJ1204−0021	12:04:41.73	−00:21:49.5	5.094	b	GMOS	−27.4	β	$10.5^{+2.9}_{-2.4}$
J2233−0107	22:33:27.65	−01:07:04.5	5.104	a	LRIS	−25.0	α	$3.5^{+1.0}_{-0.8}$
J0108−0100	01:08:29.97	−01:00:15.7	5.118	a	LRIS	−24.6	α	$2.9^{+0.8}_{-0.7}$
SDSSJ1222+1958	12:22:37.96	+19:58:42.9	5.120	b	GMOS	−25.5	β	$4.4^{+1.2}_{-1.0}$
SDSSJ0913+5919	09:13:16.55	+59:19:21.7	5.122	b	GMOS	−25.3	β	$4.0^{+1.1}_{-0.9}$
SDSSJ1209+1831	12:09:52.71	+18:31:47.0	5.127	b	GMOS	−26.8	β	$8.0^{+2.2}_{-1.8}$
SDSSJ1148+3020	11:48:26.17	+30:20:19.3	5.128	b	GMOS	−26.3	β	$6.3^{+1.7}_{-1.5}$
SDSSJ1334+1220	13:34:12.56	+12:20:20.7	5.130	b	GMOS	−26.8	β	$8.0^{+2.2}_{-1.9}$
J2334−0010	23:34:55.07	−00:10:22.2	5.137	a	LRIS	−24.6	α	$2.9^{+0.8}_{-0.7}$

Continued on next page

Table 4.1 – *Continued from previous page*

QSO	RA (J2000)	DEC (J2000)	$z_{\rm qso}$	Ref	Instrument	M_{1450}	Ref	$R_{\rm eq}$ (pMpc)
J0115+0015	01:15:44.78	+00:15:15.0	5.144	a	LRIS	−25.1	α	$3.7^{+1.0}_{-0.9}$
SDSSJ2228−0757	22:28:45.14	−07:57:55.3	5.150	b	GMOS	−26.1	β	$5.8^{+1.6}_{-1.3}$
SDSSJ1050+5804	10:50:36.47	+58:04:24.6	5.151	b	GMOS	−26.5	β	$7.0^{+1.9}_{-1.6}$
SDSSJ1054+1633	10:54:45.43	+16:33:37.4	5.154	b	GMOS	−26.4	β	$6.6^{+1.8}_{-1.5}$
SDSSJ0957+0610	09:57:07.67	+06:10:59.6	5.167	b	GMOS	−27.6	β	$11.6^{+3.1}_{-2.7}$
J2238−0027	22:38:50.19	−00:27:01.8	5.172	a	LRIS	−25.1	α	$3.7^{+1.0}_{-0.8}$
SDSSJ0854+2056	08:54:30.37	+20:56:50.9	5.179	b	GMOS	−27.0	β	$8.8^{+2.4}_{-2.1}$
SDSSJ1132+1209	11:32:46.50	+12:09:01.7	5.180	b	GMOS	−27.2	β	$9.6^{+2.5}_{-2.2}$
J1414+5732	14:14:31.57	+57:32:34.1	5.188	a	LRIS	−24.8	α	$3.2^{+0.9}_{-0.7}$
SDSSJ0915+4924	09:15:43.64	+49:24:16.6	5.199	b	GMOS	−26.9	β	$8.4^{+2.2}_{-1.9}$
SDSSJ1221+4445	12:21:46.42	+44:45:28.0	5.203	b	GMOS	−25.8	β	$5.1^{+2.3}_{-0.7}$
SDSSJ0824+1302	08:24:54.01	+13:02:17.0	5.207	b	GMOS	−26.2	β	$6.1^{+2.7}_{-0.8}$
J0349+0034	03:49:59.42	+00:34:03.5	5.209	a	LRIS	−25.3	α	$4.1^{+1.8}_{-0.5}$
SDSSJ0902+0851	09:02:45.76	+08:51:15.9	5.226	b	GMOS	−25.9	β	$5.4^{+2.3}_{-0.8}$
SDSSJ1436+2132	14:36:05.00	+21:32:39.2	5.227	b	GMOS	−26.8	β	$8.2^{+3.3}_{-1.2}$
J2202+0131	22:02:33.20	+01:31:20.3	5.229	a	LRIS	−24.6	α	$3.0^{+1.2}_{-0.5}$
J0208−0112	02:08:04.31	−01:12:34.4	5.231	a	LRIS	−25.3	α	$4.1^{+1.7}_{-0.6}$
J2211+0011	22:11:41.02	+00:11:19.0	5.237	a	LRIS	−24.8	α	$3.3^{+1.3}_{-0.6}$
SDSSJ1053+5804	10:53:22.98	+58:04:12.1	5.250	b	GMOS	−27.0	β	$9.3^{+3.5}_{-1.6}$
SDSSJ1341+3510	13:41:54.02	+35:10:05.8	5.252	b	GMOS	−26.6	β	$7.7^{+2.9}_{-1.4}$
SDSSJ1026+2542	10:26:23.62	+25:42:59.4	5.254	b	GMOS	−26.5	β	$7.4^{+2.7}_{-1.3}$
SDSSJ1626+2751	16:26:26.50	+27:51:32.5	5.265	b	GMOS	−27.8	β	$13.6^{+4.7}_{-2.5}$
SDSSJ1202+3235	12:02:07.78	+32:35:38.8	5.298	a	ESI	−28.0	β	$15.8^{+4.7}_{-3.5}$
SDSSJ1233+0622	12:33:33.47	+06:22:34.2	5.300	b	GMOS	−26.2	β	$6.8^{+2.0}_{-1.5}$
SDSSJ1614+4640	16:14:25.13	+46:40:28.9	5.313	b	GMOS	−25.8	β	$5.7^{+1.6}_{-1.2}$
SDSSJ1659+2709	16:59:02.12	+27:09:35.1	5.316	a	ESI	−27.9	β	$14.6^{+4.2}_{-3.2}$
SDSSJ1437+2323	14:37:51.82	+23:23:13.4	5.320	a	ESI	−26.8	β	$9.0^{+2.6}_{-2.1}$
J1656+4541	16:56:35.46	+45:41:13.5	5.336	a	ESI	−27.6	γ	$12.9^{+3.6}_{-2.9}$
SDSSJ1340+2813	13:40:40.24	+28:13:28.1	5.349	a	ESI	−26.6	β	$8.3^{+2.3}_{-2.0}$
J0306+1853	03:06:42.51	+18:53:15.8	5.3808	c	ESI‡	−28.9	γ	$24.5^{+6.6}_{-5.9}$
J0155+0415	01:55:33.28	+04:15:06.7	5.382	a	ESI	−27.0	δ	$10.3^{+2.7}_{-2.5}$
SDSSJ0231−0728	02:31:37.65	−07:28:54.5	5.420	a	ESI	−26.6	β	$8.6^{+2.0}_{-2.2}$
J1054+4637	10:54:05.32	+46:37:30.2	5.469	a	ESI‡	−27.0	δ	$10.3^{+2.3}_{-2.8}$
SDSSJ1022+2252	10:22:10.04	+22:52:25.4	5.4787	c	ESI	−27.3	ϵ	$11.9^{+2.4}_{-3.2}$
J1513+0854	15:13:39.64	+08:54:06.5	5.4805	c	ESI‡	−26.8	δ	$9.6^{+2.0}_{-2.7}$

Continued on next page

Table 4.1 – *Continued from previous page*

QSO	RA (J2000)	DEC (J2000)	$z_{\rm qso}$	Ref	Instrument	M_{1450}	Ref	$R_{\rm eq}$ (pMpc)
J0012+3632	00:12:32.88	+36:32:16.1	5.485	c	ESI‡	−27.2	δ	$11.8^{+2.4}_{-3.3}$
J2207−0416	22:07:10.12	−04:16:56.3	5.5297	c	ESI‡	−27.8	δ	$15.3^{+7.9}_{-2.6}$
J2317+2244	23:17:38.25	+22:44:09.6	5.5580	c	ESI‡	−27.4	δ	$12.9^{+6.5}_{-2.2}$
J1500+2816	15:00:36.84	+28:16:03.0	5.5727	c	ESI‡	−27.6	δ	$14.4^{+6.9}_{-2.6}$
J1650+1617	16:50:42.26	+16:17:21.5	5.5769	c	ESI‡	−27.2	δ	$12.3^{+5.8}_{-2.3}$
J0108+0711	01:08:06.59	+07:11:20.7	5.580	a	ESI‡	−27.2	δ	$11.9^{+5.7}_{-2.2}$
J1335−0328	13:35:56.24	−03:28:38.3	5.699	c	X-Shooter	−27.7	δ	$18.0^{+4.1}_{-5.8}$
SDSSJ0927+2001	09:27:21.82	+20:01:23.6	5.7722	d	X-Shooter	−26.8	ζ	$12.8^{+1.6}_{-4.8}$
SDSSJ1044−0125	10:44:33.04	−01:25:02.2	5.7847	e	ESI	−27.2	η	$15.6^{+1.8}_{-5.9}$
PSOJ308−27	20:33:55.91	−27:38:54.6	5.798	a	X-Shooter†	−26.8	ζ	$13.3^{+1.3}_{-5.3}$
SDSSJ0836+0054	08:36:43.85	+00:54:53.3	5.805	a	ESI	−27.8	ζ	$21.4^{+12.1}_{-5.0}$
SDSSJ0002+2550	00:02:39.40	+25:50:34.8	5.824	a	ESI	−27.3	ζ	$17.3^{+9.2}_{-4.2}$
PSOJ065+01	04:23:50.15	+01:43:24.8	5.8348	f	X-Shooter†	−26.6	ζ	$12.4^{+6.3}_{-3.1}$
PSOJ025−11	01:40:57.03	−11:40:59.5	5.8414	f	X-Shooter†	−26.9	ζ	$14.4^{+7.4}_{-3.7}$
SDSSJ0840+5624	08:40:35.09	+56:24:19.8	5.8441	g	ESI	−27.2	ζ	$16.8^{+8.4}_{-4.4}$
PSOJ242−12	16:09:45.53	−12:58:54.1	5.8468	f	X-Shooter†	−26.9	ζ	$14.8^{+7.4}_{-3.9}$
PSOJ023−02	01:32:01.70	−02:16:03.1	5.848	a	X-Shooter†	−26.5	ζ	$12.0^{+6.1}_{-3.0}$
SDSSJ0005−0006	00:05:52.33	−00:06:55.6	5.851	a	ESI	−25.7	ζ	$8.4^{+4.1}_{-2.2}$
PSOJ183−12	12:13:11.81	−12:46:03.5	5.899	a	X-Shooter†	−27.5	ζ	$20.1^{+8.8}_{-5.8}$
SDSSJ1411+1217	14:11:11.28	+12:17:37.3	5.920	a	ESI	−26.7	ζ	$14.0^{+5.8}_{-4.1}$
PSOJ340−18	22:40:48.98	−18:39:43.8	6.0007	h	X-Shooter	−26.4	ζ	$12.4^{+5.1}_{-3.8}$
SDSSJ0818+1722	08:18:27.40	+17:22:52.0	6.001	a	X-Shooter	−27.5	ζ	$20.6^{+8.2}_{-6.4}$
SDSSJ1137+3549	11:37:17.72	+35:49:56.9	6.030	a	ESI	−27.4	ζ	$19.7^{+8.0}_{-6.0}$
SDSSJ1306+0356	13:06:08.26	+03:56:26.2	6.0330	i	X-Shooter	−26.8	ζ	$14.9^{+6.1}_{-4.6}$
ULASJ1207+0630	12:07:37.43	+06:30:10.1	6.0366	j	X-Shooter	−26.6	ζ	$13.6^{+5.4}_{-4.3}$
SDSSJ2054−0005	20:54:06.49	−00:05:14.6	6.0391	e	ESI	−26.2	ζ	$11.3^{+4.6}_{-3.5}$
SDSSJ0842+1218	08:42:29.43	+12:18:50.6	6.0763	j	X-Shooter†	−26.9	ζ	$15.6^{+6.2}_{-4.9}$
SDSSJ1602+4228	16:02:53.98	+42:28:24.9	6.084	a	ESI	−26.9	ζ	$15.6^{+6.2}_{-4.7}$

Columns: (1) QSO name; (2 & 3) QSO coordinates; (4) QSO redshift; (5) reference for QSO redshift; (6) instrument used for $\lambda_{\rm mfp}$ measurements: † and ‡ denote XQR-30 spectra and spectra from our new ESI observations, respectively; (7) absolute magnitude corresponding to the mean luminosity at rest-frame 1450 Å; (8) reference for M_{1450}; (9) QSO proximity zone size defined by Equation 4.2. **References:** Redshift lines and references — a. apparent start of the Lyα forest: Becker et al. (2019, 2021) and updated measurements in this work (after correction, see text); b. adopted from Worseck et al. (2014); c. [C II]

158μm: this work; d. CO: Carilli et al. (2007); e. [C II] 158μm: Wang et al. (2013); f. [C II] 158μm: Bosman et al. (in prep.); g. CO: Wang et al. (2010); h. Lyα halo: Farina et al. (2019); i. [C II] 158μm: Venemans et al. (2020); j. [C II] 158μm: Decarli et al. (2018);. M_{1450} references — α. McGreer et al. (2013, 2018); β. Becker et al. (2021), in which M_{1450} values are calculated from the flux-calibrated spectra published by Worseck et al. (2014); γ. Wang et al. (2016); δ. Yang et al. (2017, 2019b); ϵ. measured in this work; ζ. Bañados et al. (2016, 2023); η. Schindler et al. (2020).

In 2021 and 2022, we targeted bright QSOs with z-band magnitude $m_z \leq 20$ presented in Yang et al. (2017, 2019b) using ESI. With a typical exposure time of $\sim 1-3$ hours and using the $1.00''$ and $0.75''$ slits, we acquired spectra for 11 objects (Figure 4.1), including 10 QSOs at $z_{\rm qso} \sim 5.5$ (8 of them are included in our sample; see below) and 1 QSO at a lower redshift for replacing its archival spectrum. We followed Becker et al. (2019) to reduce the data, using a custom pipeline that includes optimal techniques for sky subtraction (Kelson 2003), one-dimensional spectral extraction (Horne 1986), and telluric absorption corrections for individual exposures using models based on the atmospheric conditions measured by the Cerro Paranal Advanced Sky Model (Noll et al. 2012; Jones et al. 2013). We extracted the spectra with a pixel size of 15 km s^{-1}. The typical resolution full width at half maximum (FWHM) is approximately 45 km s^{-1}.

All targets in our sample were selected without any foreknowledge of the Lyman continuum (LyC) transmission. We include all usable spectra as long as the QSO is free from very strong associated metal absorption and/or associated Lyα damping wing absorption, which may bias the $\lambda_{\rm mfp}$ measurements. We also reject objects with strong broad absorption lines (BALs) near the systemic redshift (see Bischetti et al. 2022 for the XQR-30 spectra). Because the LyC transmission is very weak at $z > 5.3$, we only use spectra with a signal-to-noise ratio (SNR) of $\gtrsim 20$ per 30 km s^{-1} near 1285 Å in the rest frame. Among the objects

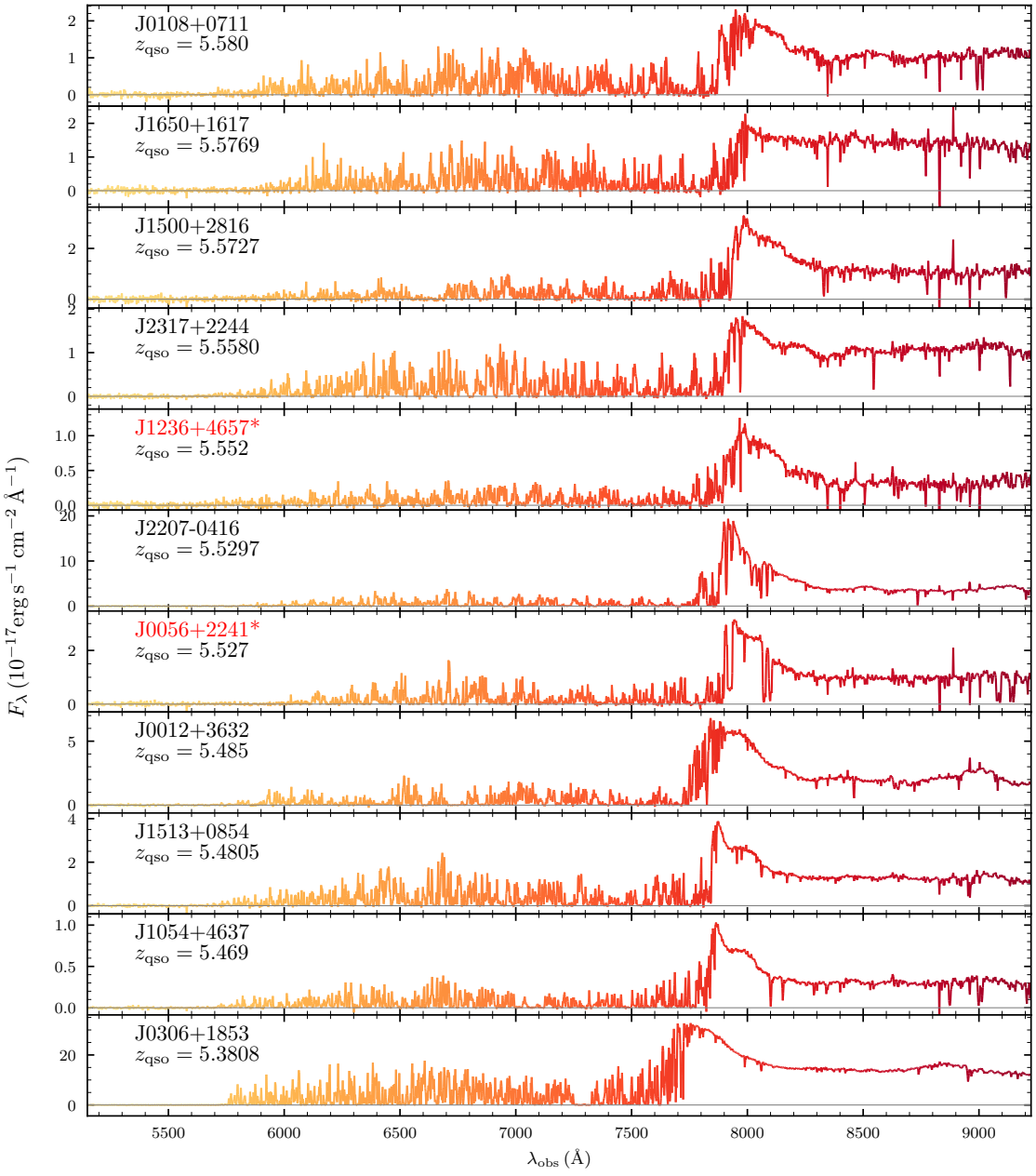


Figure 4.1: New Keck/ESI QSO spectra obtained for this work. Orange-red and gray curves plot the flux and zero flux, respectively. J0056+2241 and J1236+4657 (labelled with *) are not included in our sample of $\lambda_{\rm mfp}$ measurements (see text for details). The spectra are re-binned to 2 Å for display.

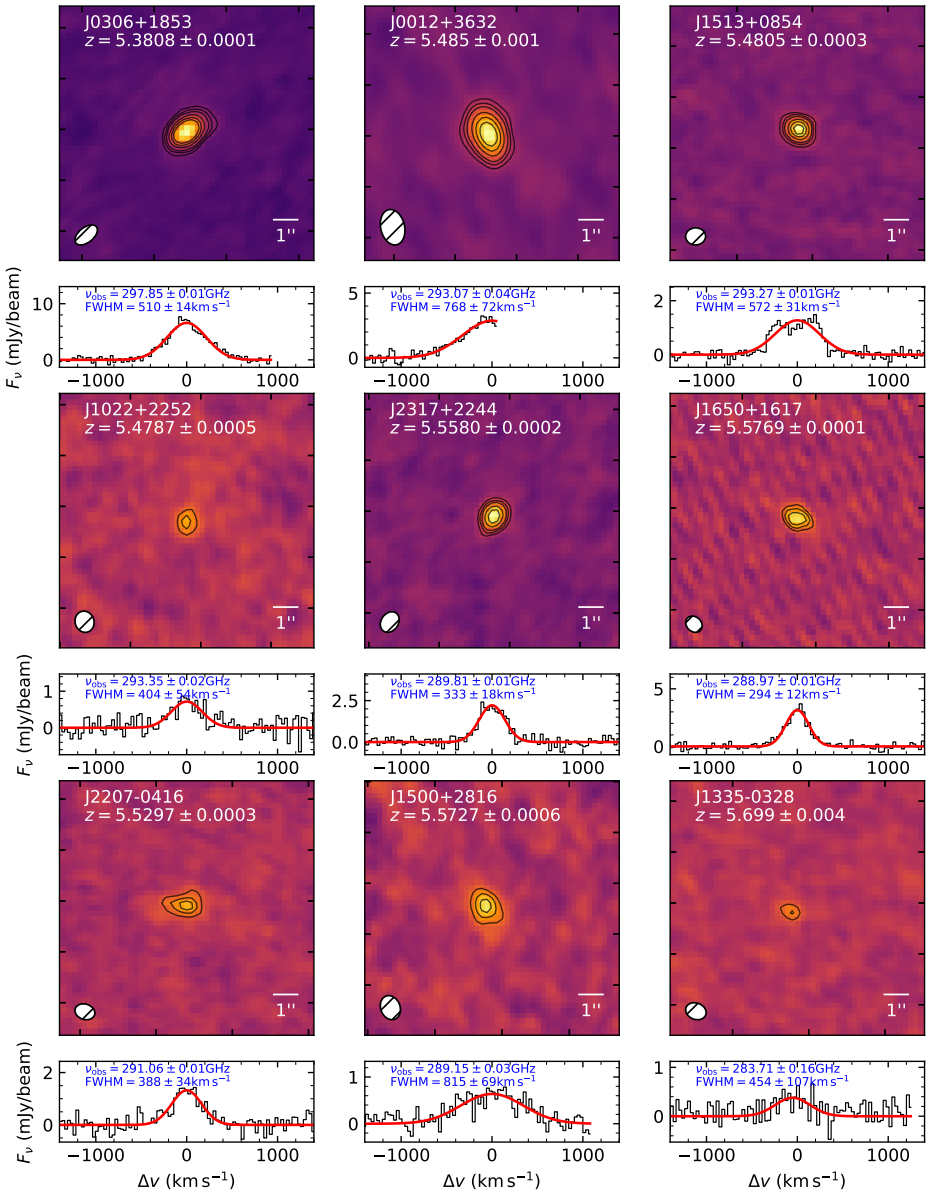


Figure 4.2: [C II] 158μm emission maps and spectra of QSOs with our new Keck/ESI observations. Contours show (2σ, 3σ, 4σ, 6σ, 8σ, 10σ) levels. Measured redshifts are labelled for each QSO and red curves show the best Gaussian fits. Observed frequency and FWHM of the [C II] emission are also provided for reference.

with new ESI spectra, we exclude J0056 due to its strong associated absorber, and J1236 for its low SNR.

For QSO redshifts, we employ measurements based on sub-millimeter observations, whenever available. Additionally, we carried out ALMA Band 7 observations of our new ESI targets in Cycle 9 and determined the systemic redshifts by fitting the [C II] 158μm line. For each object, we used two overlapping spectral windows to cover the [C II] line based on the estimated redshift and another two spectral windows to cover the dust continuum. With C43-(1, 2, 3) configurations, the typical angular resolution is $\sim 1''$. The data are calibrated and reduced using the default procedures of the CASA pipeline (version 6.4.1.12; McMullin et al. 2007; CASA Team et al. 2022). We follow the procedures described in Eilers et al. (2020) to generate the data cube and image the [C II] line: the [C II] emission is continuum subtracted with `uvconstsub`, and imaged with the `tclean` procedure using Briggs cleaning and a robust parameter of 2 (natural weighting) to maximize the sensitivity. We use a robust parameter of 0.5 for J1650+1617 to achieve a best data product. The mean RMS noise of our data set is 0.25 mJy beam^{-1} per 30 MHz bin. Figure 4.2 displays [C II] maps along with Gaussian fits to the emission. For each QSO, we extract the spectrum within one beam size centered at the target. We create the [C II] map by stacking the data cube within 1 standard deviation of the Gaussian fit from the line center. We note that the emission line of J0012+3632 is incomplete because the [C II] line is at the edge of our spectral window, which was chosen based on a preliminary redshift estimate (Yang et al. 2019b). J1513+0854 shows a double-peak emission line, which may be due to the rotating

disk of the QSO host galaxy. Future observations with higher spatial resolution may help resolve the disk.

For the rest of our sample, we employ redshifts measured from the apparent start of the Lyα forest, which are determined for each line of sight by visually searching for the first Lyα absorption line blueward the Lyα peak (e.g., Worseck et al. 2014; Becker et al. 2019). We do not use redshifts measured from Mg II emission because of their large offsets ($\sim$ 500 km s^{-1}) from the systemic redshifts (e.g., Venemans et al. 2016; Mazzucchelli et al. 2017; Schindler et al. 2020). Based on 42 QSOs at $5.3 < z_{\rm qso} < 6.6$ with [C II] or CO redshifts, we find that the redshifts we measure from the apparent start of the Lyα forest are blueshifted from the [C II] or CO redshifts by $\sim$ 185 km s^{-1} on average, with a standard deviation of $\sim$ 370 km s^{-1}. Such a redshift offset can be explained by the strong proximity zone effect close to the QSO: the first significant absorber may typically occur slightly blueward of the QSO redshift due to ionization effects. This offset is also consistent with that measured in B21. Thus, we shift redshifts measured from the apparent start of the Lyα forest by +185 km s^{-1} when measuring $\lambda_{\rm mfp}$, and the corrected values are listed in Table 4.1.

We generate rest-frame composite spectra for QSOs in each $\Delta z = 0.3$ bin, starting from $z = 4.9$. The redshift bins with a mean redshift of $\langle z \rangle = 5.08$, 5.31, 5.65, and 5.93 consist of 44, 26, 9, and 18 spectra, respectively. Following B21, each spectrum is shifted to rest-frame wavelength before being normalized. The normalization is done by dividing each spectrum by its *continuum* flux measured over wavelengths in the rest frame where the flux from broad emission lines is minimal. Here, we use the *continuum* flux over 1270Å $< \lambda <$ 1380Å in the rest frame. We have tested that using a different wavelength window does

not significantly change our results. Additionally, we identify and mask wavelength regions affected by skyline subtraction residuals. To reject spurious bad pixels, we apply a light median filter using a 3-pixel sliding window. Mean composite spectra are then computed in 120 km s^{-1} bins (as shown in the left-hand panel of Figure 4.3).

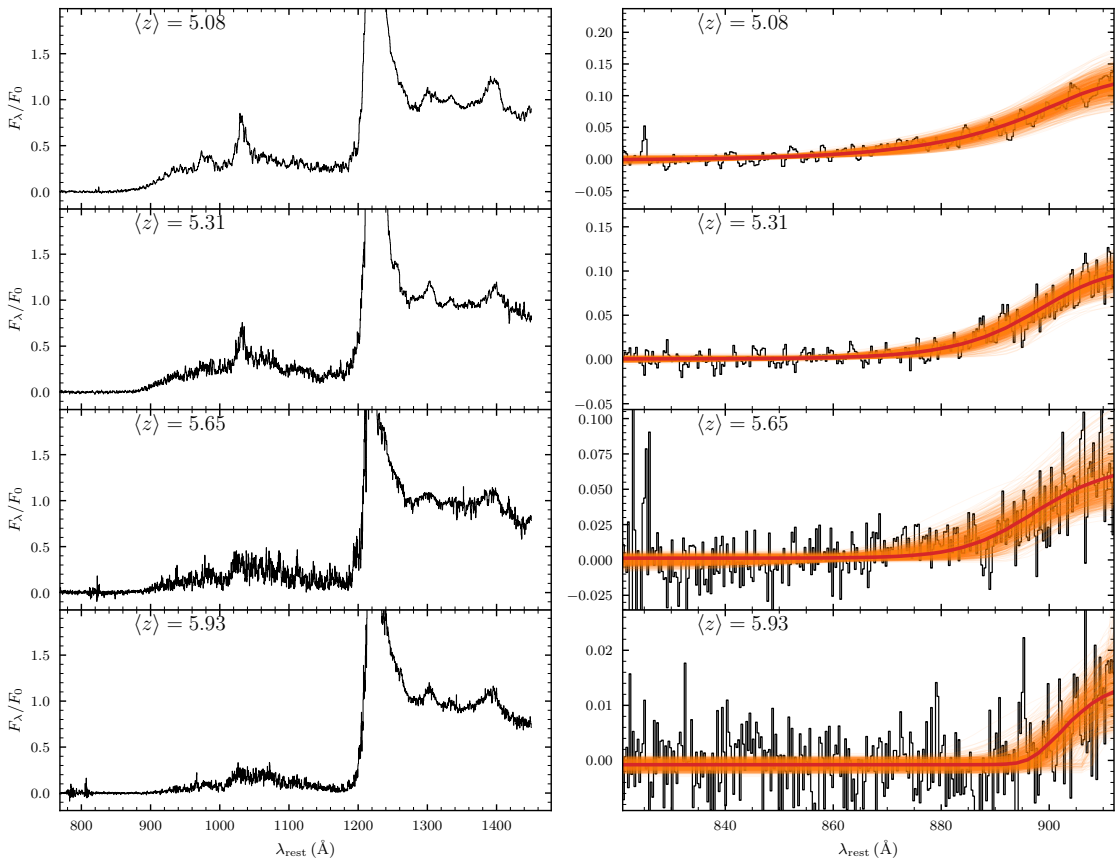


Figure 4.3: Composite QSO spectra for each redshift bin (left-hand panel) and fits to the LyC transmission (right-hand panel). Black curves show observed flux normalized by the median flux over 1270-1380 Å in the rest frame. The red curve shows our best-fit model. Thin orange curves show the fits from bootstrap realization (only 1000 curves are plotted here for display).

4.4 Methods

We measure the $\lambda_{\rm mfp}$, which is defined as the distance travelled by ionizing photons that would be attenuated by a factor of $1/e$ by LyC absorption, by fitting the transmitted flux profile blueward of the Lyman limit in composite QSO spectra (Prochaska et al. 2009; B21). One challenge with this approach is that the LyC transmission at $z > 5$ can be significantly affected by the QSO proximity effect. The ionizing flux from the QSO decreases the ionizing opacity in the proximity zone, which can bias the inferred $\lambda_{\rm mfp}$ high by a factor of two or more (D'Aloisio et al. 2018; B21). This is especially important when $\lambda_{\rm mfp}$ is smaller than the proximity zone size, which is true for bright QSOs at $z \gtrsim 5$.

To address this bias, we follow the methods and modeling presented in B21, which modified the Prochaska et al. (2009) method of measuring $\lambda_{\rm mfp}$ to explicitly include the proximity effect. Motivated by simulations, we account for the decrease in ionizing opacity near the QSO by scaling the opacity, $\kappa_{\rm LL}$, according to the local photoionization rate, Γ. This dependence is modeled as a power law such that $\kappa_{\rm LL} \propto \Gamma^{-\xi}$,

$$\kappa_{\rm LL}(r) = \kappa_{\rm LL}^{\rm bg} \left[1 + \frac{\Gamma_{\rm qso}(r)}{\Gamma_{\rm bg}}\right]^{-\xi} , \tag{4.1}$$

where $\Gamma_{\rm qso}(r)$ is the photoionization rate due to the QSO at a distance r from the QSO, and $\Gamma_{\rm bg}$ is the background photoionization rate. In order to calculate $\Gamma_{\rm qso}(r)$, a key parameter used to describe the proximity zone effect in B21 is $R_{\rm eq}$. It denotes the distance from the QSO where $\Gamma_{\rm qso}(r)$ and $\Gamma_{\rm bg}$ would be equal for purely geometric dilution in the absence of attenuation. Following Calverley et al. (2011), $R_{\rm eq}$ is related to $\Gamma_{\rm bg}$ and the ionizing

luminosity of the QSO, L_{912}, by

$$R_{\rm eq} = \left[\frac{L_{912}\sigma_0}{8\pi^2\hbar\Gamma_{\rm bg}(\alpha_\nu^{\rm ion}+2.75)}\right]^{1/2}, \tag{4.2}$$

where σ_0 and $\alpha_\nu^{\rm ion}$ are the H I ionization cross section at 912 Å and the power-law index of the QSO continuum at $\lambda < 912$ Å in the frequency domain, respectively, and $\hbar$ is the reduced Planck constant. L_{912} can be further related to the absolute magnitude corresponding the the mean luminosity at rest-frame 1450 Å, L_{1450}, by $L_{912} = L_{1450}(\nu_{912}/\nu_{1450})^{-\alpha_\nu^{\rm UV}}$. Here, ν_{912} and ν_{1450} are the photon frequencies at 912 Å and 1450 Å, respectively, and $\alpha_\nu^{\rm UV}$ is the power-law slope for the non-ionizing ($912\text{Å} < \lambda < 1450\text{Å}$) continuum of the QSO continuum. Following B21, we adopt $\alpha_\nu^{\rm ion} = 1.5 \pm 0.3$ [3] (Telfer et al. 2002; Stevans et al. 2014; Lusso et al. 2015) and $\alpha_\nu^{\rm UV} = 0.6 \pm 0.1$ (Lusso et al. 2015, see also Vanden Berk et al. 2001; Shull et al. 2012; Stevans et al. 2014).

Following B21, the observed flux, $f_\lambda^{\rm obs}$, is given by the mean intrinsic QSO continuum, $f_\lambda^{\rm cont} \propto (\lambda/912\text{Å})^{\alpha_\lambda^{\rm ion}}$, attenuated by the effective opacity of the Lyman series in the foreground IGM, and the LyC optical depth. The foreground Lyman series opacity is calculated by

$$\tau_{\rm eff}^{\rm Lyman}(\lambda_{\rm obs}) = \sum_{j={\rm Ly}\alpha,\,{\rm Ly}\beta,\,...,\,{\rm Ly40}} \tau_{\rm eff}^{j}(z_j)\,, \tag{4.3}$$

where $\tau_{\rm eff}^{j}(z_j)$ is the effective opacity of transition j at redshift z_j such that $(1+z_j)\lambda_j = \lambda_{\rm obs}$, and λ_j is the wavelength of transition j in the rest frame. To implement this, we utilized Sherwood simulations (Bolton et al. 2017) to determine the effective optical depth for each Lyman series line across a range of absorption redshifts and Γ values. We then include the proximity zone effect for each Lyman series line by matching the effective optical depth

[3] $\alpha_\lambda^{\rm ion} = -0.5 \pm 0.3$ in the wavelength domain.

to a Γ value that corresponds to $\Gamma_{\rm bg} + \Gamma_{\rm qso}$ as a function of distance from the QSO. We compute $\Gamma_{\rm qso}$ by dividing the line of sight into small steps of distance δr, and solving for $\Gamma_{\rm qso}(r)$ numerically using the method described in B21. For the first step we assume that $\Gamma_{\rm qso}$ decreases purely geometrically, i.e.

$$\Gamma_{\rm qso}(r = \delta r) = \Gamma_{\rm bg} \left(\frac{r}{R_{\rm eq}} \right)^{-2} . \tag{4.4}$$

We then solve for $\Gamma_{\rm qso}(r + \delta r)$ over subsequent steps as

$$\Gamma_{\rm qso}(r + \delta r) = \Gamma_{\rm qso}(r) \left(\frac{r + \delta r}{r} \right)^{-2} e^{-\kappa_{\rm LL}(r)\delta r} , \tag{4.5}$$

where $\kappa_{\rm LL}(r)$ is computed using Equation (4.1). Finally, $\tau_{\rm eff}^{j}(z_j)$ and $\tau_{\rm eff}^{\rm Lyman}(\lambda_{\rm obs})$ can be computed for each combination of ($\kappa_{\rm LL}^{\rm bg}$, ξ, $z_{\rm qso}$, $R_{\rm eq}$), when fitting to the composite spectra. We have also tested that it does not significantly change the measured $\lambda_{\rm mfp}$ by either stacking the foreground Lyman series transmission based on $z_{\rm qso}$ and $R_{\rm eq}$ of each individual QSO (as in B21), or computing a total foreground Lyman series transmission based on the averaged $z_{\rm qso}$ and $R_{\rm eq}$ in each redshift bin.

We use the same procedures for parameterizing the LyC transmission as outlined in B21. However, we make one modification by employing the recent measurements of $\Gamma_{\rm bg}$ from Gaikwad et al. (2023) that match multiple diagnostics of the IGM from observations to the Lyα forest. For reference, the new estimates are $\Gamma_{\rm bg} \simeq 5 \times 10^{-13}{\rm s}^{-1}$ and $1.5 \times 10^{-13}{\rm s}^{-1}$ at $z = 5.1$ and 6.0, respectively, in contrast to $7 \times 10^{-13}{\rm s}^{-1}$ and $3 \times 10^{-13}{\rm s}^{-1}$ utilized in B21. The new $\Gamma_{\rm bg}$ at $z = 5.1$ is also consistent with measurements in e.g., D'Aloisio et al. (2018). Moreover, instead of assuming a nominal ± 0.15 dex error in $\Gamma_{\rm bg}$, we propagate the uncertainties in the measurements of $\Gamma_{\rm bg}$ from Gaikwad et al. (2023) into $R_{\rm eq}$. We discuss the effect of $\Gamma_{\rm bg}$ on $\lambda_{\rm mfp}$ in Section 4.5.2.

We fit the transmission for each composite shown in Figure 4.3 over 820–912 Å in the rest frame. Following B21, uncertainties in $\lambda_{\rm mfp}$ are estimated using a bootstrap approach wherein we randomly draw QSO spectra with replacement in each redshift bin, and refit the new QSO composites for 10,000 realizations. To account for errors in redshift, we randomly shift the spectrum of each QSO that does not have a sub-mm $z_{\rm qso}$ in redshift following a Gaussian distribution with a standard deviation of 370 km s^{-1}(see Section 4.3). We include the zero-point as a free parameter while fitting models to the composite, to account for flux zero-point errors. We also treat the normalization of the LyC profile as a free parameter. We randomly vary ξ by assuming a flat prior over $[0.3, 1.0]$. The fits are shown in the right-hand panel of Figure 4.3.

4.5 Results and Discussion

4.5.1 $\lambda_{\rm mfp}$ over $5 < z < 6$

We measure $\lambda_{\rm mfp} = 9.33^{+2.06(3.76)}_{-1.80(2.82)}$, $5.40^{+1.47(2.39)}_{-1.40(2.39)}$, $3.31^{+2.74(4.02)}_{-1.34(2.06)}$, and $0.81^{+0.73(1.22)}_{-0.48(0.68)}$ pMpc at the averaged redshifts $\langle z \rangle = 5.08$, 5.31, 5.65, and 5.93, respectively. The errors give 68% (95%) confidence limits. Figure 4.4 plots our results along with previous direct $\lambda_{\rm mfp}$ measurements from the literature (Prochaska et al. 2009; Fumagalli et al. 2013; O'Meara et al. 2013; Worseck et al. 2014; Lusso et al. 2018, and B21). Our measurements at both redshift ends are highly consistent with those presented in B21. Our findings clearly indicate a rapid evolution in $\lambda_{\rm mfp}$ at $5 < z < 6$, particularly at $z \geq 5.3$.

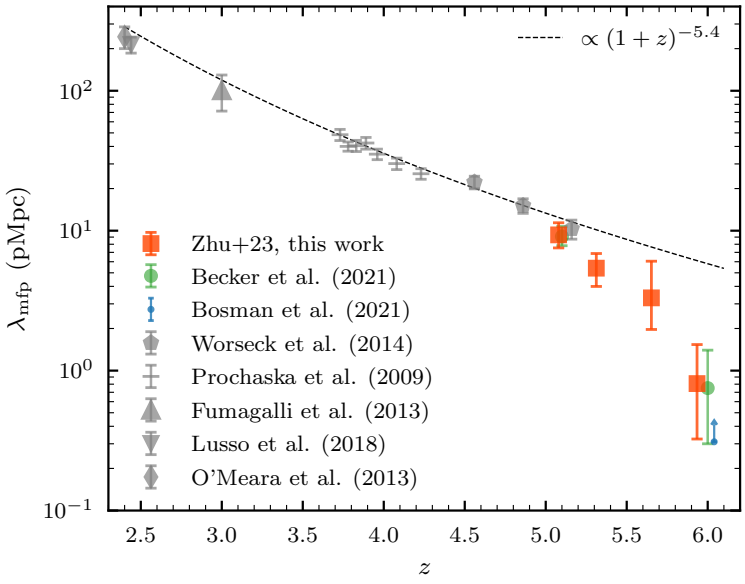
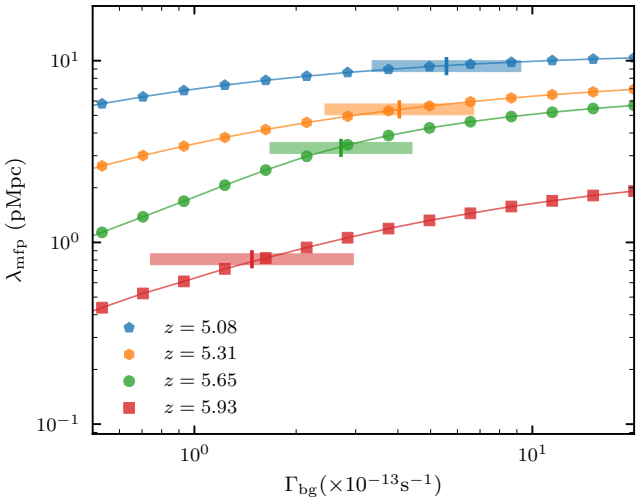
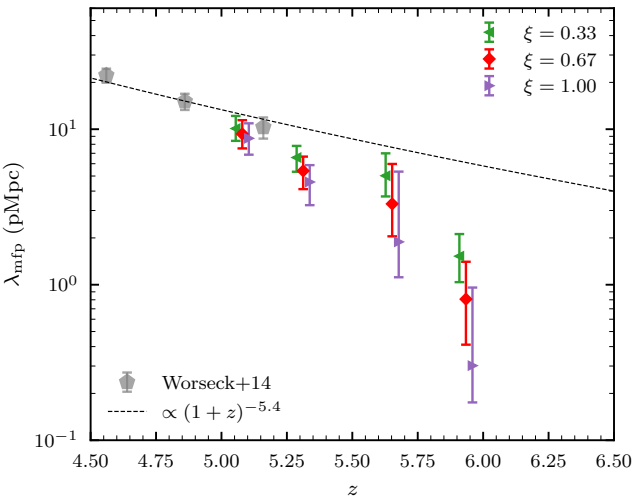


Figure 4.4: Direct measurements of $\lambda_{\rm mfp}$ from this work (orange-red squares) and from the literature (Becker et al. 2021; Worseck et al. 2014; Prochaska et al. 2009; Fumagalli et al. 2013; Lusso et al. 2018; O'Meara et al. 2013). Error bars show 68% limits. The dashed line shows the power-law extrapolation of $\lambda_{\rm mfp}$ from measurements at $z < 5.16$ (Worseck et al. 2014). The blue arrow shows the lower-limit constraint on the $\lambda_{\rm mfp}$ at $z = 6$ from Bosman (2021). Symbols are displaced along redshift for display.



(a)



(b)

Figure 4.5: **(a)** Dependence of the $\lambda_{\rm mfp}$ measurements on $\Gamma_{\rm bg}$ at different redshifts. Nominal values and uncertainties of $\Gamma_{\rm bg}$ we adopt from Gaikwad et al. (2023) are shown with horizontal error bars. **(b)** Measured $\lambda_{\rm mfp}$ based on fixed ξ values of 0.33, 0.67, and 1.00. Symbols are displaced along redshift for display.

We have confirmed that using different redshift binning does not significantly affect our results. As the composite spectrum at $\langle z \rangle = 5.65$ includes relatively fewer QSOs, we tested the robustness of our fitting using mock spectra. We created 1000 sets of $N = 9$ mock QSO spectra with similar redshifts and $R_{\rm eq}$ as our sample. The mock spectra were based on our modeling of the transmission at $\lambda < 912$ Å, with the mean free path randomly spanning a wide parameter space. We performed 1000 fitting realizations to each of these mock spectra sets, and found that the 68% limits at $z = 5.65$ could recover the simulated confidence level quite well. In addition, it is worth noting that none of the objects in our sample are identified as young QSOs with extremely small proximity zones (e.g., Eilers et al. 2020; Satyavolu et al. 2023b), given that the $\lambda_{\rm mfp}$ values are quite short near $z \sim 6$.

We note that our $\langle z \rangle = 5.65$ stack (Figure 4.3) includes a small amount of transmission near $\lambda_{\rm rest} \sim 880$ Å, even though the flux has been significantly attenuated at $\lambda_{\rm rest} > 890$ Å. We have inspected each individual spectrum in this redshift bin and have found potential transmission near this wavelength ($\lambda_{\rm obs} \sim 5800$ Å) only in the ESI spectra, and not in the X-Shooter spectra. This region is at the boundary of two amplifiers of the ESI CCD. It may also be contaminated by scattered light. This feature might therefore be an instrumental issue [4]; however, it is also possible that the transmission is real, in which case it may indicate a significant variation in the ionizing free paths towards individual spectra. We hope to explore this in a future work. For the uniform $\lambda_{\rm mfp}$ model used here, however, our tests based on mock spectra indicate that this transmission feature does not significantly impact our measurements.

[4] Prochaska et al. (2003) also report high flux near $\lambda_{\rm obs} \sim 5800$ Å and infer that this is due to an incorrect matching in the gain of the two amplifiers for ESI. We still observe this feature after attempting to carefully account for the gain ratio, however.

Table 4.2: Error analysis for the measured $\lambda_{\rm mfp}$

	$\langle z \rangle = 5.08$	$\langle z \rangle = 5.31$	$\langle z \rangle = 5.65$	$\langle z \rangle = 5.93$
Measured $\lambda_{\rm mfp}$	$9.33^{+2.06}_{-1.80}$	$5.40^{+1.47}_{-1.40}$	$3.31^{+2.74}_{-1.34}$	$0.81^{+0.73}_{-0.48}$
Fixed $\Gamma_{\rm bg}$ and ξ	$9.33^{+1.95}_{-1.83}$	$5.40^{+1.23}_{-1.14}$	$3.31^{+2.52}_{-0.89}$	$0.81^{+0.45}_{-0.34}$
Varying $\Gamma_{\rm bg}$ only	$9.33^{+0.43}_{-0.69}$	$5.40^{+0.62}_{-0.72}$	$3.31^{+0.92}_{-0.99}$	$0.81^{+0.21}_{-0.26}$
$\xi = 0.33$	$10.10^{+2.07}_{-1.69}$	$6.57^{+1.23}_{-1.25}$	$5.02^{+1.98}_{-1.32}$	$1.53^{+0.59}_{-0.49}$
$\xi = 0.67$	$9.33^{+2.10}_{-1.81}$	$5.40^{+1.27}_{-1.28}$	$3.31^{+2.67}_{-1.26}$	$0.81^{+0.60}_{-0.40}$
$\xi = 1.00$	$8.74^{+2.18}_{-1.88}$	$4.58^{+1.30}_{-1.33}$	$1.89^{+3.45}_{-0.77}$	$0.30^{+0.66}_{-0.13}$

Table Comments: $\lambda_{\rm mfp}$ values are reported in pMpc.
(1) "Measured $\lambda_{\rm mfp}$": our $\lambda_{\rm mfp}$ measurements at each redshift, including all sources of error;
(2) "Fixed $\Gamma_{\rm bg}$ and ξ": statistical error from bootstrapping the QSO lines of sight without changing the nominal $\Gamma_{\rm bg}$ or ξ;
(3) "Varying $\Gamma_{\rm bg}$ only": $\lambda_{\rm mfp}$ values from varying $\Gamma_{\rm bg}$ while keeping the QSO lines of sight and ξ fixed;
(4) Others: $\lambda_{\rm mfp}$ values based on different fixed ξ values.

4.5.2 Error analysis & dependence on $\Gamma_{\rm bg}$ and ξ

As described in Section 4.4, we mitigate the bias on $\lambda_{\rm mfp}$ from the QSO proximity effect by modeling its impact on the ionizing opacity. The effect is parameterized by a nominal proximity zone size $R_{\rm eq}$, which specifies the proper distance at which the hydrogen ionization rate due to the QSO would be equal to the background rate in the absence of any attenuation. Therefore, the measured $\lambda_{\rm mfp}$ has some dependence on $R_{\rm eq}$. Notably, the uncertainties in $R_{\rm eq}$ listed in Table 4.1 primarily emanate from $\Gamma_{\rm bg}$, with the contribution of uncertainty from $\alpha^{\rm UV}$ and $\alpha^{\rm ion}$ being relatively small ($\lesssim 10\%$). The measurements also depend on ξ, as suggested by Equation 4.1.

Here, we have examined various sources of error in our $\lambda_{\rm mfp}$ measurements, including statistical error, error arising from $\Gamma_{\rm bg}$, and error stemming from ξ. Specifically, we evaluate the following: (1) the statistical error obtained by bootstrapping the QSO lines of

sight while keeping the nominal values of $\Gamma_{\rm bg}$ and ξ fixed, (2) $\lambda_{\rm mfp}$ values derived by varying $\Gamma_{\rm bg}$ while keeping the QSO composite and ξ constant, and (3) $\lambda_{\rm mfp}$ values obtained for different fixed ξ. Table 4.2 summarizes the results. We would like to emphasize that fixing any of the parameters to their nominal value can lead to a slightly altered distribution of the resulting $\lambda_{\rm mfp}$, and consequently, the corresponding 68% limits may not align precisely with those of the main results. In this case, our primary interest lies in understanding the magnitude of the error.

The "Fixed $\Gamma_{\rm bg}$ and ξ" row indicates that the magnitude of the statistical error is comparable to the total error across all redshifts. This suggests that the dominant factor contributing to the error in our measurements is statistical fluctuations, encompassing factors such as the limited number of QSO spectra, flux noise, uncertainties in redshifts, and uncertainties in $\alpha_\nu^{\rm UV}$ and $\alpha_\nu^{\rm ion}$, among others. As shown in the third row, the random fluctuation in $\Gamma_{\rm bg}$ alone have a minor impact on the overall error. Regarding ξ, the last three rows indicate that using a lower (upper) value of ξ shifts the $\lambda_{\rm mfp}$ estimates towards higher (lower) face values. This effect is comparable to the statistical fluctuations and is more pronounced at higher redshifts due to the relatively large $R_{\rm eq}$ in comparison to smaller $\lambda_{\rm mfp}$. Thus, uncertainties in ξ also make a significant contribution to our $\lambda_{\rm mfp}$ measurements.

We also explore how our $\lambda_{\rm mfp}$ measurements depend on $\Gamma_{\rm bg}$ and ξ systematically, such that the results can be easily adjusted for future constraints. Figure 4.5(a) illustrates the dependence of our best-fitting $\lambda_{\rm mfp}$ measurements on a wide range of $\Gamma_{\rm bg}$ values at each redshift. The figure also displays the nominal $\Gamma_{\rm bg}$ values and their uncertainties. The

dependence is minimal at $z \leq 5.3$, where $\lambda_{\rm mfp} \propto \Gamma_{\rm bg}^{\sim 0.2}$. At these redshifts, the proximity zone size is smaller or comparable to $\lambda_{\rm mfp}$, and hence, the measurements are not highly sensitive to $\Gamma_{\rm bg}$. At $z \geq 5.6$, however, $R_{\rm eq}$ is similar to or larger than $\lambda_{\rm mfp}$, leading to a stronger dependence, with $\lambda_{\rm mfp} \propto \Gamma_{\rm bg}^{\sim 0.6}$. Nevertheless, if we adopt the higher end of $\Gamma_{\rm bg} = 3 \times 10^{-13} {\rm s}^{-1}$ at $z = 5.93$, $\lambda_{\rm mfp}$ would only increase to ~ 1.0 pMpc. The results would remain consistent with a steady and rapid $\lambda_{\rm mfp}$ evolution with time.

While there might be an enhanced ionizing background due to galaxies clustering near QSOs, recent research suggests this effect is likely secondary to QSO ionization. Davies (2020) found that even the "ghost" proximity effect of QSOs — a large-scale bias in the ionizing photon mean free path caused by QSO radiation — could overwhelm the ionizing contribution from the clustering of nearby galaxies. In recent JWST observations, Kashino et al. (2023) also found that the impact of a QSO's ionizing radiation often dominates over local galactic sources near the QSO. These studies reinforce that, despite potential $\Gamma_{\rm bg}$ enhancements from galaxy clustering, the QSO's influence is typically predominant, as adopted in our modeling for the $\lambda_{\rm mfp}$ measurements.

For our main results, following B21, the mean free path is measured based on a uniform distribution of $\xi \in [0.33, 1.00]$, and the face value is measured for $\xi = 0.67$. As discussed in B21, however, the scaling of $\kappa_{\rm LL}$ with Γ is highly uncertain, especially for high redshifts. The scaling can be milder (smaller ξ) when neutral islands and/or self-shielding absorbers are present, and steeper (greater ξ) when the IGM shows a more uniform photoionization equilibrium (see e.g., D'Aloisio et al. 2020; Furlanetto & Oh 2005; McQuinn et al. 2011). For reference, Figure 4.5(b) shows how our measurements vary with

fixed ξ values of 0.33, 0.67, and 1.00. The face value and errors are also listed in Table 4.2. Similar to the dependence on $\Gamma_{\rm bg}$, the measured $\lambda_{\rm mfp}$ becomes more sensitive to ξ as redshift increases, and as the QSO's proximity effect becomes stronger relative to a smaller $\lambda_{\rm mfp}$. Even with the extreme ξ values discussed in B21, nevertheless, the measurements are still consistent with our main results with the 1σ error bars, given the current data. Reassuringly, radiative transfer simulations recently developed by Roth et al. (in prep) suggest that the inferred $\lambda_{\rm mfp}$ using our methods only modestly depends ($\lesssim 20-30\%$) on the QSO lifetimes and environments (see also Satyavolu et al. in prep). Future improved realistic reionization models would provide more insights into the scaling relation, especially when reionization is not fully concluded by $z = 6$.

4.5.3 Implication on when reionization ends

Our measurements not only confirm the $\lambda_{\rm mfp}$ values presented in B21 at $z = 5.08$ and 5.93, but also depict a clear evolutionary trend over $5 < z < 6$. The mean free path increases steadily and rapidly with time: $\lambda_{\rm mfp}$ increases by a factor of ~ 4 from $z \simeq 6.0$ to $z \simeq 5.6$, and by a factor of ~ 2 every $\Delta z = 0.3$ from $z \simeq 5.6$ to $z \simeq 5.0$. This evolution carries important implications for the end of reionization.

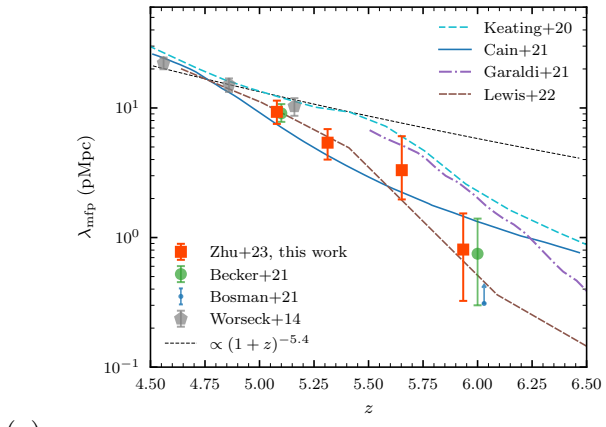
D'Aloisio et al. (2020) used radiative transfer hydrodynamic simulations to study the expected evolution of the mean free path following reionization. They found that if reionization had ended well before $z = 6$ and the IGM had sufficient time to relax hydrodynamically, then the evolution would expected to follow a trend of $\lambda_{\rm mfp} \propto (1+z)^{-5.4}$. This relation, based on a fully ionized IGM with a homogeneous ionizing UVB, is identical to the best-fitting redshift dependence for direct $\lambda_{\rm mfp}$ measurements at $z \leq 5$ (Worseck

et al. 2014). However, as shown in Figure 4.4, this relation significantly overpredicts the measurements by a factor of $\sim 2-10$ over $5.3 < z < 6.0$. By this comparison, the data disfavor a fully ionized and relaxed IGM with a homogeneous UVB at these redshifts.

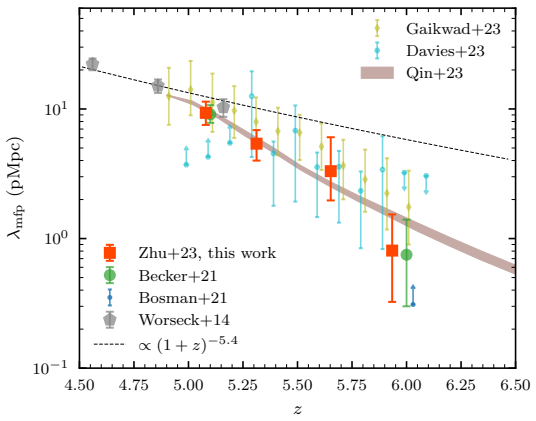
One possible explanation for the rapid evolution in $\lambda_{\rm mfp}$ is that reionization ends later than $z = 6$. Such a late-ending reionization scenario has recently been proposed to explain large-scale fluctuations in the observed Lyα effective optical depth ($\tau_{\rm eff}$) at $z > 5$ (e.g., Kulkarni et al. 2019a; Keating et al. 2020b; Nasir & D'Aloisio 2020; Choudhury et al. 2021; Qin et al. 2021). The rapid evolution in $\lambda_{\rm mfp}$ can be naturally explained by ongoing reionization when large ionized bubbles merge and dense, optically-thick ionization sinks are photo-evaporated (e.g., Furlanetto & Oh 2005; Sobacchi & Mesinger 2014; D'Aloisio et al. 2020). As mentioned above, the rapid evolution in $\lambda_{\rm mfp}$ persists as late as $z \lesssim 5.3$, and the significant discrepancy between measurements and predictions from the relaxed IGM model also appears as late as $z = 5.3$. Interestingly, the rapid evolution in $\lambda_{\rm mfp}$ appears to coincide in redshift with the appearance of large fluctuations in the observed Lyα $\tau_{\rm eff}$ at $z \gtrsim 5.3$ (e.g., Becker et al. 2015; Eilers et al. 2018; Bosman et al. 2018, 2022; Yang et al. 2020b). This may be due to the fact that a shorter $\lambda_{\rm mfp}$, along with any potential neutral component from incomplete reionization, will boost the fluctuations in the ionizing UV background, producing scatter in $\tau_{\rm eff}$ (e.g., Davies & Furlanetto 2016; Nasir & D'Aloisio 2020; Qin et al. 2021). This joint evolution in the mean free path and UV background is expected near the end of reionization (e.g., Kulkarni et al. 2019a; Keating et al. 2020b; Nasir & D'Aloisio 2020). Such a scenario is also consistent with long dark gaps observed in the

Lyα/Lyβ forest (Zhu et al. 2021, 2022) at $z < 6$, and the fraction of dark pixels measured in the forest (McGreer et al. 2015; Jin et al. 2023).

4.5.4 Comparison with reionization simulations



(a)



(b)

Figure 4.6: **(a)** Direct measurements of $\lambda_{\rm mfp}$ from this work (orange-red squares) compared to the predictions from recent models (Cain et al. 2021; Garaldi et al. 2022; Lewis et al. 2022). **(b)** Our measurements compared with indirect constraints based on Lyα opacities (Gaikwad et al. 2023, Davies et al. in prep., Qin et al. in prep.). Open circles correspond to marginal constraints and arrows correspond to 2σ limits for Davies et al. (in prep). The shaded region shows the posterior from Qin et al. (in prep.) at the 68% confidence level. In both panels, for comparison, we also show the direct $\lambda_{\rm mfp}$ measurements in Becker et al. (2021) and Worseck et al. (2014), as well as the lower-limit constraint in Bosman (2021). Symbols are displaced along redshift for display.

Figure 4.6(a) compares our $\lambda_{\rm mfp}$ measurements to recent numerical simulations of reionization, including the enhanced-sink simulation in Cain et al. (2021), THESAN in Garaldi et al. (2022), and CoDaIII in Lewis et al. (2022). These models use late-ending reionization histories and aim to explain the observed short $\lambda_{\rm mfp}$ at $z = 6$ and the rapid evolution measured in B21, which we have confirmed in finer detail here. We also include the ATON simulation ("low $\tau_{\rm CMB}$" model) in Keating et al. (2020a) for reference.

Cain et al. (2021) reproduce $\lambda_{\rm mfp}$ that is consistent with the $z = 6$ measurements in a late-ending reionization driven by faint galaxies. However, they also find that either a rapid drop in emissivity at $z < 6$ or extra sinks are required to reproduce the $\lambda_{\rm mfp}$ measurements at $z < 5.2$ in this scenario. Garaldi et al. (2022) use a radiative hydrodynamics simulation and generally reproduce the rapid evolution of $\lambda_{\rm mfp}$ although overshoot the $z = 6$ measurement. They find that all of their late-reionization simulations can reproduce a dramatic evolution in $\lambda_{\rm mfp}$ from $z = 5.5$ to 6, while one simulation wherein reionization ends by $z > 6$ cannot. Lewis et al. (2022) also find that reionization ending later than $z < 6$ is able to naturally explain the observations, although a drop in the emissivity is required near the end of reionization (Ocvirk et al. 2021).

We note that these simulations measure the $\lambda_{\rm mfp}$ in slightly different manners. Cain et al. (2021) generate mock LyC QSO spectra using randomly placed sightlines, and fit the stacked spectra using the model proposed by Prochaska et al. (2009). This procedure mimics the method used in Worseck et al. (2014). On the other hand, Garaldi et al. (2022) measure the distance at which the LyC transmission is attenuated by a factor of $1/e$ from individual sightlines, and take the average. Finally, Lewis et al. (2022) adopted multiple

measurement methods for the $\lambda_{\rm mfp}$, and the curve shown in Figure 4.6(a) represents the median distance to a $1/e$ attenuation in LyC among sightlines. We also note that the models shown here may not necessarily reproduce observations of the Lyα forest transmission (e.g., Garaldi et al. 2022). Nevertheless, the rapid evolution of $\lambda_{\rm mfp}$ that we measure over $5 < z < 6$ is broadly consistent with these models that align with a late conclusion to reionization.

4.5.5 Comparison with constraints on $\lambda_{\rm mfp}$ from Lyα opacities

Recently, Gaikwad et al. (2023) and Davies et al. (in prep) have used alternative methods to probe the evolution of $\lambda_{\rm mfp}$ at these redshifts. Instead of directly measuring $\lambda_{\rm mfp}$ from the LyC transmission, they use inference methods to jointly constrain $\lambda_{\rm mfp}$ and $\Gamma_{\rm bg}$ by modeling the observed Lyα effective optical depth distribution. Specifically, Gaikwad et al. (2023) post-processed hydrodynamic IGM simulations using a fluctuating UV background specified by a spatially-averaged photoionization rate and a mean-free path parameter, λ_0. These variables are constrained by comparing the simulated cumulative distribution function of $\tau_{\rm eff}$ with observations using a non-parametric two-sample Anderson-Darlington test. A value of the physical mean free path, $\lambda_{\rm mfp}$, is then inferred from the simulated neutral hydrogen distribution once the best-fitting UV background is applied. This step is particularly significant at $z \lesssim 5.3$, where the $\tau_{\rm eff}$ distribution is consistent with a uniform UVB (see also Becker et al. 2018; Bosman et al. 2022) and does not constrain $\lambda_{\rm mfp}$ directly. The fact that the constraints at these redshifts from Gaikwad et al. (2023) agree with the direct constrains presented here suggests that their simulations may be modeling much of the ionizing opacity.

Davies et al. (in prep) take a similar approach but use a combination of hydrodynamical and semi-numerical simulations to model the density field, and employ a likelihood-free inference technique of approximate Bayesian computation to constrain $\lambda_{\rm mfp}$ and $\Gamma_{\rm bg}$ based on the Lyα forest observations. Davies et al. (in prep) also constrain $\lambda_{\rm mfp}$ by treating it as an "input" sub-grid parameter for the UVB fluctuations rather than inferring it from a derived H I density distribution. This accounts for the fact that the Davies et al. (in prep) values at $z < 5.3$ are lower limits.

As shown in Figure 4.6(b), these indirect $\lambda_{\rm mfp}$ constraints are generally consistent ($\lesssim 1\sigma$) with the direct measurements presented here and in B21. This suggests that the rapidly evolving $\lambda_{\rm mfp}$ values needed for UV background fluctuations to drive the observed Lyα $\tau_{\rm eff}$ distribution are consistent with the attenuation of ionizing photons we observe directly. We note that Gaikwad et al. (2023) and Davies et al. (in prep) present somewhat different pictures of the IGM at these redshifts; the Gaikwad et al. (2023) models include neutral islands persisting to $z \simeq 5.2$ while Davies et al. (in prep) have the flexibility to vary the $\lambda_{\rm mfp}$ distribution within ionized regions although no neutral islands are explicitly included. These models are broadly consistent with one another in that the Lyα opacity fluctuations are mainly driven by UV background fluctuations, but this difference highlights the range of physical scenarios that are still formally consistent with the data.

We also include the inference based on multiple observations. Recently, Qin et al. (in prep) use the `21cmFAST` Epoch of Reionization (EoR) simulations (Mesinger et al. 2011; Murray et al. 2020) to constrain IGM properties including its $\lambda_{\rm mfp}$. Their input parameters represent galaxy properties such as the stellar-to-halo mass ratio, UV escape

fraction and duty cycles. These allow them to evaluate the UV ionizing photon budget and forward model the impact on the IGM. Within a Bayesian framework, they include not only the XQR-30 measurement of the forest fluctuations (Bosman et al. 2022) when computing the likelihood but also the CMB optical depth and galaxy UV luminosity functions. Therefore, the posterior they obtain for the IGM mean free path potentially represents a comprehensive constraint from all currently existing EoR observables. As Figure 4.6(b) displays, the posterior from Qin et al. (in prep) also show a rapid increase in the $\lambda_{\rm mfp}$ with time between $5 \lesssim z \lesssim 6$. Although the posterior does not follow the exact trace of our direct measurements, the general consistency may suggest that such a rapid evolution in $\lambda_{\rm mfp}$ is potentially favored by other EoR observables.

4.6 Conclusion

We have presented new measurements of the ionizing mean free path between $z = 5.0$ and 6.0. These are the first direct measurements in multiple redshift bins over this interval, allowing us to trace the evolution of $\lambda_{\rm mfp}$ near the end of reionization. Our measurements are based on new and archival data, including new Keck/ESI observations and spectra from the XQR-30 program. By fitting the LyC transmission in composite spectra, we report $\lambda_{\rm mfp}= 9.33^{+2.06}_{-1.80}$, $5.40^{+1.47}_{-1.40}$, $3.31^{+2.74}_{-1.34}$, and $0.81^{+0.73}_{-0.48}$ pMpc, at $\langle z \rangle = 5.08$, 5.31, 5.65, and 5.93, respectively. The results confirm the dramatic evolution in $\lambda_{\rm mfp}$ over $5 < z < 6$, as first reported in B21, and further show a steady and rapid evolution, with a factor of ~ 4 increase from $z \simeq 6.0$ to $z \simeq 5.6$, and a factor of ~ 2 increase every $\Delta z = 0.3$ from $z \simeq 5.6$ to $z \simeq 5.0$. Our $\lambda_{\rm mfp}$ measurements disfavor a fully ionized and relaxed IGM

with a homogeneous UVB at $\gtrsim 95\%$ confidence level down to at least $z \sim 5.3$ and are coincident with the onset of the fluctuations in observed $\tau_{\rm eff}$ at $z \sim 5.3$.

Recent indirect $\lambda_{\rm mfp}$ constraints based on IGM Lyα opacity from Gaikwad et al. (2023) and Davies et al. (in prep) agree well with our measurements and those in B21. Our results are also broadly consistent with a range of late-ending reionization models (Cain et al. 2021; Garaldi et al. 2022; Lewis et al. 2022; Gaikwad et al. 2023). Along with other probes from the Lyα and Lyβ forests, our results suggest that islands of neutral gas and/or large fluctuations in the UV background may persist in the IGM well below redshift six.

Chapter 5

Damping Wing-Like Features in the Stacked Lyα Forest: Potential Neutral Hydrogen Islands at $z < 6$

5.1 Abstract

Recent QSO absorption line observations suggest that reionization may end as late as $z \approx 5.3$. As a means to search for large neutral hydrogen islands at $z < 6$, we revisit long dark gaps in the Lyβ forest in in VLT/X-Shooter Keck/ESI QSO spectra. We stack the Lyα forest corresponding to the red edge of these dark gaps and identify a damping wing-like extended absorption profile. The average redshift of the stacked forest is $z = 5.82$. By comparing these observations with reionization simulations, we infer that such a damping wing-like feature can be naturally explained by the presence of neutral hydrogen within

these Lyβ dark gaps. Conversely, simulated dark gaps lacking neutral hydrogen struggle to replicate the observed damping wing features. Furthermore, this damping wing-like profile implies that the volume-averaged neutral hydrogen fraction must be $\langle x_{\rm HI} \rangle \geq 6.1 \pm 3.9\%$ at $z \approx 5.8$. Our results offer robust evidence supporting the extension of reionization below $z = 6$. [1]

5.2 Introduction

Hydrogen reionization carries key implications for the formation and evolution of the first stars, galaxies, and supermassive black holes in the early Universe (see, e.g., Fan et al. 2023 for a review). Cosmic microwave background (CMB) observations indicate a midpoint of reionization at $z \sim 7-8$ (Planck Collaboration et al. 2020, see also de Belsunce et al. 2021). Meanwhile, transmitted flux observed in the Lyα forest at $z \sim 6$ towards high-redshift QSOs (e.g., Fan et al. 2006) has long been interpreted as an indicator of the end of reionization. However, if reionization indeed ends by $z \sim 6$, galaxies would have to produce an extremely large amount of ionizing photons to complete reionization within a short timeframe. This poses a significant challenge to our understanding, particularly in terms of the inferred star formation intensity, escape fraction of ionizing photons, and the spectral energy distribution (SED) of early galaxies (e.g., Robertson et al. 2015; Bouwens et al. 2015; Finkelstein et al. 2019; Stark 2016).

A later end to reionization, which would help resolve such tension, is gaining support from recent observations. Supporting evidence includes large-scale fluctuations in

[1] A version of this chapter is to be submitted for publication and a preprint can be found at https://ydzhuastro.github.io/10202617.

the Lyα effective optical depth[2] measured in QSO spectra (e.g., Fan et al. 2006; Becker et al. 2015; Eilers et al. 2018; Bosman et al. 2018, 2022; Yang et al. 2020b); long troughs extending down to or below $z \simeq 5.5$ in the Lyα and Lyβ forests (e.g., Becker et al. 2015; Zhu et al. 2021, 2022), potentially indicating the existence of large neutral intergalactic medium (IGM) islands (e.g., Kulkarni et al. 2019a; Keating et al. 2020b; Nasir & D'Aloisio 2020; Qin et al. 2021); observed underdensities around long dark gaps traced by Lyα emitting galaxies (LAEs, Becker et al. 2018; Kashino et al. 2020; Christenson et al. 2021); the evolution of metal-enriched absorbers at $z \sim 6$ (e.g., Becker et al. 2019; Davies et al. 2023b,a); and the dramatic evolution in the measured mean free path of ionizing photons over $5 < z < 6$ (Becker et al. 2021; Zhu et al. 2023, see also Bosman 2021; Gaikwad et al. 2023; Satyavolu et al. 2023a; Roth et al. 2023; F.B. Davies et al. in prep.), which provides the most unambiguous evidence to date of an ongoing reionization at $z < 6$. A late reionization scenario is also consistent with numerical models that reproduce a variety of observations (e.g., Weinberger et al. 2019; Choudhury et al. 2021; Qin et al. 2021; Gaikwad et al. 2023, Y. Qin et al. in prep.).

A pressing question is whether we can detect neutral hydrogen islands at $z < 6$ if they exist. In terms of resonant Lyα and Lyβ absorption, the observational signatures of neutral islands may be difficult to distinguish from those of large-scale fluctuations in the ionizing ultraviolet background; either one may produce extended regions of nearly zero transmission. At higher redshifts, where the IGM is more neutral, damping wing features over the Lyα emission of QSOs and galaxies have been observed and used as a powerful probe of the volume-averaged neutral hydrogen fraction ($x_{\rm HI}$) of the Universe (e.g., Bañados

[2]Defined as $\tau_{\rm eff} = -\ln \langle F \rangle$, where F is the continuum-normalized transmission flux.

et al. 2018; Davies et al. 2018b; Wang et al. 2020; Yang et al. 2020a; Greig et al. 2022; Umeda et al. 2023). Any signature of damping wing absorption at lower redshifts might similarly indicate the presence of neutral islands. A detection of damping wings would also enable constrain $x_{\rm HI}$ that would complement the upper limits from the fraction of dark gaps and dark pixels in the forest (McGreer et al. 2013; Zhu et al. 2022; Jin et al. 2023).

Malloy & Lidz (2015) propose a test for neutral islands at $z \sim 5.5$ by observing the damping wing over stacked Lyα forests corresponding to highly absorbed regions. This method has not been implemented in observations because it requires a large sample of high-quality spectra of high-z QSOs and a relatively complete catalog of intervening metal absorbers to exclude contamination from damped Lyα systems (DLAs) due to nearby galaxies. Thanks to a large sample of QSO spectra provided in the (Extended) VLT XQR-30 large program (D'Odorico et al. 2023) and the metal absorber catalog of Davies et al. (2023b), we are now able to carry out the experiment proposed in Malloy & Lidz (2015). Long dark gaps have been detected in the Lyβ forest (Zhu et al. 2022) and they indicate regions of high IGM opacity that may potentially host neutral islands (e.g., Kulkarni et al. 2019a; Keating et al. 2020b; Nasir & D'Aloisio 2020; Qin et al. 2021). Therefore, stacking the Lyα forest at the redshifts of these long dark gaps is a powerful way to implement the (Malloy & Lidz 2015) test and search for IGM damping wing features at $z < 6$.

This chapter is organized as follows. In Section 5.3 we describe the data and Lyβ dark gaps used in this work. Section 5.4 presents the stacked Lyα profile. Sections 5.5-5.7 compares our results to model predictions and infers lower-limit constraints on $x_{\rm HI}$. Finally, we conclude our findings in Section 5.8. Throughout this paper, we quote distances

in comoving units unless otherwise noted and assume a ΛCDM cosmology with $\Omega_{\rm m} = 0.308$, $\Omega_\Lambda = 0.692$, and $h = 0.678$ (Planck Collaboration et al. 2014).

5.3 Data

To create a stacked spectrum of the Lyα forest over potentially neutral regions at $z < 6$, we revisit dark gaps detected in the Lyβ forest in Zhu et al. (2022). We refer the readers to Zhu et al. (2021, 2022) for details of the data and dark gap detection. Briefly, Lyβ dark gaps are detected in 42 spectra of QSOs at $5.77 \lesssim z_{\rm qso} \lesssim 6.31$. The spectra are taken with the Echellette Spectrograph and Imager (ESI) on Keck (Sheinis et al. 2002) and the X-Shooter spectrograph on the Very Large Telescope (VLT; Vernet et al. 2011). Among these, 19 X-Shooter spectra are from the XQR-30 VLT large program (D'Odorico et al. 2023). A dark gap in the Lyβ forest is defined as a continuous spectral region in which all pixels binned to 1 h^{-1} Mpc have an observed normalized flux $F = F_{\rm obs}/F_{\rm c} < 0.02$, where $F_{\rm obs}$ is the observed flux and $F_{\rm c}$ is the continuum flux predicted from Principal Component Analysis (PCA; see Bosman et al. 2021, 2022; Zhu et al. 2021). The Lyα forest at the corresponding redshift of a Lyβ dark gap also needs to be opaque ($F_{\rm Ly\alpha} = F_{\rm obs}/F_{\rm c} < 0.05$).

In this work we use only the Zhu et al. (2022) dark gaps at $z < 6$. Dark gap statistics based on mock spectra generated from simulations suggest that longer dark gaps are more likely to have a higher covering fraction of neutral hydrogen (Nasir & D'Aloisio 2020; Zhu et al. 2022). The stack based on longer dark gaps might also create a stronger damping wing profile in the corresponding Lyα forest (Malloy & Lidz 2015). To balance the sample size and dark gap length, here we use Lyβ dark gaps with $L \geq 7$ h^{-1} Mpc.

Considering the low transmission in the Lyα forest at these redshifts, we only include QSO spectra with a continuum-normalized flux error < 0.05 per pixel in the Lyα and Lyβ forest. Therefore, dark gaps toward QSOs PSO J308-21 and VIK J2318-3029 are excluded due to the relatively low S/N of the spectra.

We have carefully checked these dark gaps to ensure they do not contain intervening metal absorbers within 3000 km s^{-1} near the edge of dark gaps, based on the metal line catalog presented in Davies et al. (2023b) and through visual inspection. This requirement rejects the dark gap toward QSO SDSS J0818+1722 spanning $z = 5.761 - 5.794$ with $L = 10\ h^{-1}\,\mathrm{Mpc}$. This dark gap contains absorption lines of O I, C II, Si II, C IV, etc., near the red edge at $z \simeq 5.79$. We note that, however, including this dark gap produces no substantial difference in the results. A summary of Lyβ dark gaps used in this work is provided in Table 5.1. Our final sample includes 24 dark gaps, for which the average redshift at the red edge is $z = 5.82$.

5.4 Stacked Lyα Forest

Inspired by Malloy & Lidz (2015), we create a stacked Lyα forest after aligning the spectra based on the redshift at the red edge of Lyβ dark gaps. We firstly bin the continuum-normalized Lyα forest to 80 km s^{-1} intervals, corresponding to $\sim 0.5\ h^{-1}\,\mathrm{Mpc}$ at $z = 5.8$, which is half of the bin size used by Zhu et al. (2022) when searching for dark gaps. To remove the redshift evolution in the mean Lyα transmission, we re-scale the sightlines with a factor equal to the mean transmission at $z = 5.82$ divided by the mean transmission at the redshift of each pixel according to the measurements in Bosman et al.

Table 5.1: Lyβ Dark Gaps Used in this Work

Index	QSO	$z_{\rm blue}$	$z_{\rm red}$	$L_{\rm gap}$
1	ULASJ1319+0950	5.876	5.903	8
2	PSOJ060+24	5.833[a]	5.856	7
3	PSOJ108+08	5.674	5.751	24
4	SDSSJ0842+1218	5.784	5.830	14
5	SDSSJ2315-0023	5.790	5.883	28
6	SDSSJ2315-0023	5.897	5.937[b]	$\geq$ 12
7	CFHQSJ1509-1749	5.800	5.863	19
8	CFHQSJ1509-1749	5.870	5.910	12
9	SDSSJ2054-0005	5.751	5.790	12
10	SDSSJ0840+5624	5.585	5.607	7
11	PSOJ340-18	5.774	5.810	11
12	PSOJ065-26	5.954	5.988	10
13	PSOJ007+04	5.741	5.780	12
14	ULASJ0148+0600	5.654	5.735	25
15	ULASJ0148+0600	5.741	5.803	19
16	PSOJ217-16	5.807	5.873	20
17	PSOJ217-16	5.920	5.961[b]	$\geq$ 12
18	J0408-5632	5.715	5.741	8
19	PSOJ359-06	5.886	5.917	9
20	PSOJ025-11	5.526	5.613	28
21	PSOJ025-11	5.632	5.661	9
22	PSOJ158-14	5.764	5.836	22
23	PSOJ158-14	5.843	5.880	11
24	SDSSJ1250+3130	5.836	5.870	10

Table Comments: Columns: (1) Index of dark gaps; (2) QSO spectra used in Zhu et al. (2021); (3) redshift at the blue end of the gap; (4) redshift at the red end of the gap; (5) dark gap length in the units of h^{-1}Mpc.
[a] Dark gap starting at the blue edge of the Lyβ forest.
[b] Dark gap ending at the red edge of the Lyβ forest as defined in Zhu et al. (2022).

(2021). The stacked spectrum is then created by taking a mean value of the flux in each bin. We exclude the QSO proximity zone effect by conservatively disregarding the Lyα forest within 11 proper-Mpc towards the QSO following Zhu et al. (2022). We also mask pixels that are affected by sky-line subtraction residuals or telluric absorption correction as indicated by peaks in the flux error array.

Figure 5.1 shows the stacking result along with the Lyα forest corresponding to each individual Lyβ dark gap. We see an extended damping wing-like absorption feature redward of the edge of the dark gaps, with the transmission gradually recovering to the mean value at a velocity offset of $\Delta v \sim 1400$ km s^{-1}. Notably, such a damping wing-like feature is not obvious in any of the individual spectra.

Following Malloy & Lidz (2015), the absorption due to the presence of a neutral hydrogen island will be extended, with optical depth far from the line center described by a damping wing. The optical depth at a velocity offset Δv from the edge of a neutral island will be

$$\tau_{\mathrm{Ly}\alpha}^{\mathrm{DW}}(\Delta v) \approx \frac{\tau_{\mathrm{GP}} R_\alpha c}{\pi} \left[\frac{1}{\Delta v} - \frac{1}{\Delta v + v_{\mathrm{ext}}} \right] , \tag{5.1}$$

where τ_{GP} is the Gunn-Peterson optical depth, for which we take a nominal value of 2.5×10^5 at $z \sim 5.8$, $R_\alpha \equiv \Gamma_\alpha \lambda_\alpha / 4\pi c$, and $\Gamma_\alpha = 6.265 \times 10^8$ s^{-1} is the Lyα decay constant.

We fit the damping wing profile in Equation 5.1 to the upper envelope of the stacked spectrum in the interval between $\Delta v = 0$ km s^{-1} and 1300 km s^{-1}. Specifically, we perform a least-squares fitting over the stacked flux, and require that the fitted curve cannot be significantly lower than the stacked flux in any pixel (e.g., Prochaska et al. 2003). Here, we set the threshold to be 80% of the stacked flux and we have tested that using a

different threshold does not change our conclusions. As Figure 5.1 shows, the fitted curve well describes the observed extended absorption feature and yields $v_{\rm ext} = 550$ km s^{-1}. We note that this damping wing-like profile is comparable with that expected for neutral islands with $L = 5.34\ h^{-1}$ Mpc in Malloy & Lidz (2015).

We emphasize that our approach does not represent a physical fit to a damping wing per se, but rather a fit to a mean transmission profile using a *damping wing-like* function. The observed mean profile potentially includes damping wing absorption arising from genuine neutral islands, but will also include resonant Lyα absorption, which may be modified by on the scales of interest by UVB fluctuations. As described in Section 5.6, we assess the evidence for damping wing absorption by fitting the same function to mock stacks drawn from simulations.

We note that there are large dips in the stacked profile over $, 1500{\rm km\,s^{-1}} \lesssim \Delta v \lesssim 2,400{\rm km\,s^{-1}}$. These are likely statistical fluctuation resulting from the relatively small number of spectra included in the stack. A clustering of neutral islands on the scales of $\sim 10\ h^{-1}$ Mpc may also create absorption features redward the damping wing-like profile in the stack, although this hypothesis would require more data to test.

5.5 Lyβ Dark Gaps in Simulations

A critical question is whether the observed damping wing-like profile in the stacked Lyα forest indicates the presence of genuine neutral islands at $z < 6$. To explore the conditions under which such a profile might arise, we generate mock spectra from simulations,

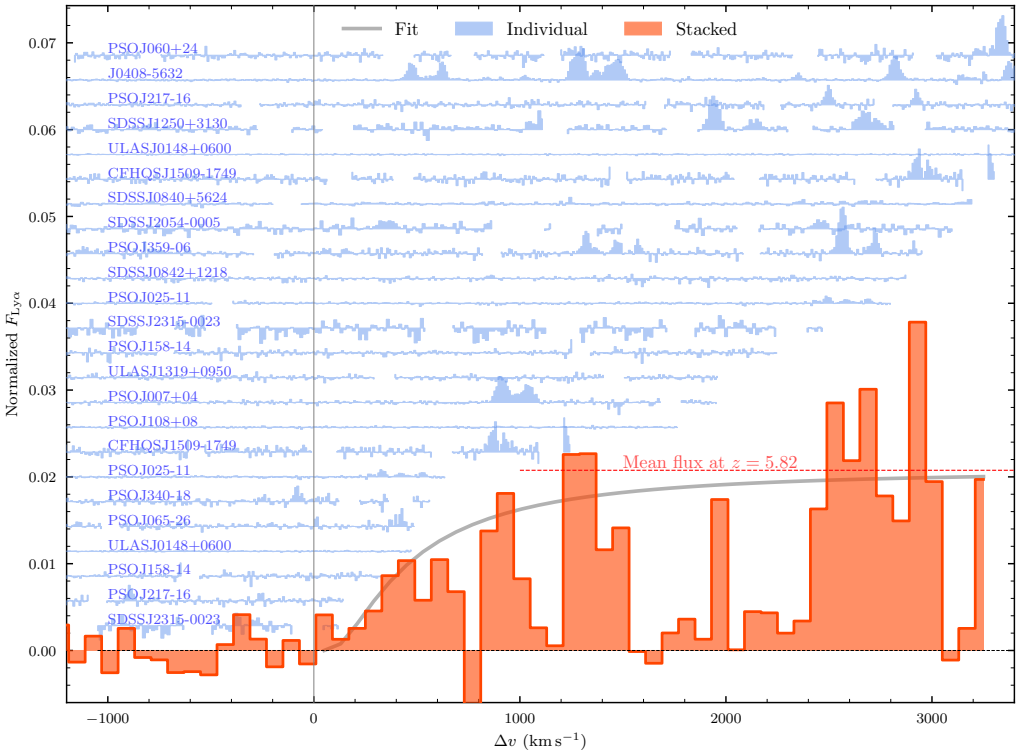


Figure 5.1: Stacked Lyα forest corresponding to the red edges of Lyβ dark gaps (displayed in orange). The grey curve represents the fitted damping wing profile, as described by Equation 5.1, applied to the upper envelope of the stacked flux between $\Delta v = 0$ km s^{-1} and 1300 km s^{-1}. The individual spectra used in this stack are shown in blue, with the flux scaled by a factor of 0.01 for display purposes.

create stacks based on the Lyβ gaps in these spectra, and compare the properties of the mock stacks to the observations.

The sightlines are drawn from a radiative transfer (RT) simulation conducted in a $(200h^{-1}\mathrm{Mpc})^3$ volume box with $N = 200^3$ RT cells, using the adaptive ray-tracing code described in Cain et al. (2021, 2022). Detailed methodology will be presented in C. Cain et al. (in prep.). In the simulation, halos and galactic sources are populated based on abundance matching according to the luminosity function from Finkelstein et al. (2019), and the hydrogen ionizing emissivity of the sources is assumed to be proportional to their UV luminosity. This simulation is calibrated to match the observed mean transmission in the Lyα forest as reported in Bosman et al. (2022). We account for the attenuation from the foreground Lyα forest when computing the Lyβ forest flux. In this section, we use synthesis spectra without noise.

Dark gaps in the simulated sightlines are identified following the same criteria as in Zhu et al. (2022); specifically, we select gaps that are dark in both the Lyβ forest and the Lyα forest. We then stack the Lyα forest at the red edge of Lyβ dark gaps with lengths $L \geq 7\ h^{-1}\,\mathrm{Mpc}$. For our fiducial case we only include dark gaps with a neutral hydrogen covering fraction of ≥ 0.7 (denoted as "$x_{\mathrm{HI}} \geq 0.7$"). We then explore different variations. First, we construct flux arrays in the Lyα and Lyβ forests using a Gaussian profile instead of a Voigt profile to assess the impact of the IGM damping wing and test whether the observed absorption profile can be reproduced without a genuine damping wing absorption. The stack generated in this manner is labeled as "no damping wing". Second, we use dark gaps that contain no neutral hydrogen (denoted as "ionized dark gap"). Finally, we

eliminate the attenuation from the foreground Lyα forest when building the Lyβ flux array for all the aforementioned cases. This last scenario is obviously not physically realistic, but will be helpful in assessing how foreground Lyα impacts the results.

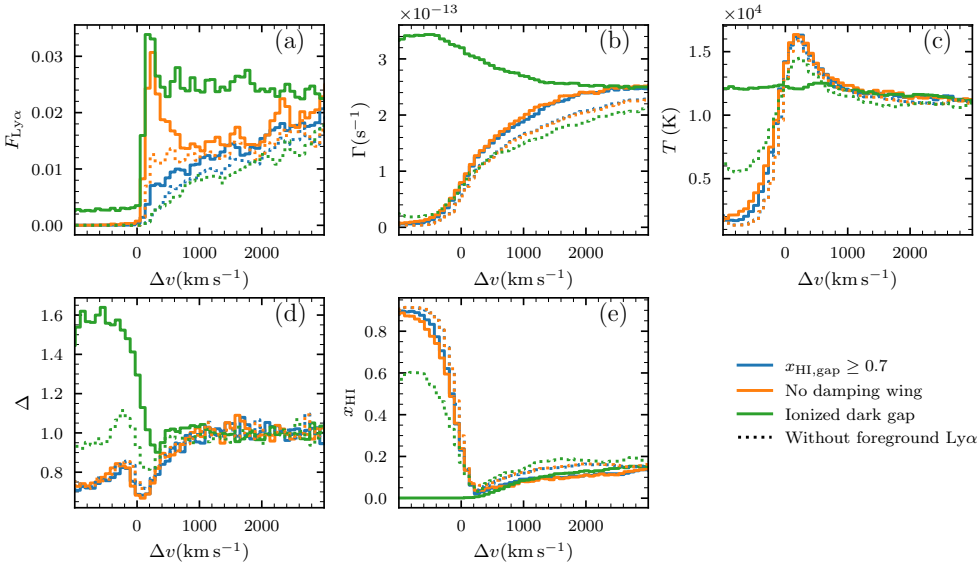


Figure 5.2: Stacked (mean) Lyα transmission profiles at the redshift of the red edge of Lyβ dark gaps detected in ideal sightlines from the simulation. Panels show a) the Lyα transmission, b) the ionization rate, c) the IGM temperature, d) the overdensity, and e) the volume filling factor of neutral hydrogen. The solid blue line represents the fiducial case, where dark gaps are at least 70% neutral. The solid orange line illustrates the case with neutral hydrogen where the absorption is calculated without a Lorentzian damping wing. The solid green line depicts the case of ionized dark gaps. Dotted lines indicate the stacked profiles when Lyβ dark gaps are selected without contamination from foreground Lyα absorption.

Figure 5.2 displays the stacked properties of the dark gaps at their red edge under various scenarios. In the fiducial case, dark gaps with substantial neutral hydrogen produce an extended absorption profile in the adjacent Lyα forest. These dark gaps are typically found in regions characterized by a low ionizing UVB, low temperature, and low densities. Their properties align with findings in studies such as Gnedin (2022). When the damping

wing feature, enabled by the Voigt profile, is removed, as illustrated by the orange curve, there is an immediate increase in the stacked Lyα flux after $\Delta z = 0$ km s^{-1}, indicating that a genuine damping wing absorption is needed to reproduce an extended absorption profile in this case. A similar rapid increase in the stacked Lyα flux is observed in the case of dark gaps that arise from ionize gas. These dark gaps are likely situated in high-density regions.

Nevertheless, we note that if Lyβ dark gaps are selected in the absence of contamination from foreground Lyα absorption, then even the "ionized dark gap" scenario exhibits extended Lyα absorption. Physically, this is because the Lyα transmission should recover gradually as moving away from ionized regions of high Lyα opacity. Contamination from "pseudo" Lyβ gaps that are created by foreground Lyα absorption, however, obscures this effect. As for the "$x_{\rm HI} \geq 0.7$" scenario, the selection criteria ensure that the dark gaps contain genuine neutral islands and they are not contaminated. As discussed in following sections, the observed extended absorption profile may indicate that the long dark gaps in (Zhu et al. 2022) are similar to those selected dark gaps that contain genuine neutral islands.

5.6 Comparison of the Stacked Lyα Forest

In the previous section we have investigated "ideal" cases based on stacks of large numbers of dark gaps without noise in the flux. In this section, we create more realistic mock spectra and use bootstrap to test the likelihood that a profile similar to the data will emerge in small samples. Building upon the ideal sightlines from Section 5.5, we adjust the modeled spectra to have the same redshift coverage as the observed sample and add noise

according to the flux error array from the observations. To mitigate the effect of variations in dark gap length distributions across different cases, in each bootstrap realization, we randomly draw dark gaps from the simulations, following exactly the same distribution of dark gap lengths as in the observed data to create the stack. Nevertheless, the shape of the stacked Lyα forest does not change significantly even without the correction on the dark gaps length distribution.

In addition to the scenarios mentioned in Section 5.5, we add model predictions from a fluctuating UVB model, "ND20-early-shortmfp", wherein the UVB fluctuations are boosted by a short mean free path and neutral islands are not explicitly included (refer to Nasir & D'Aloisio 2020; Zhu et al. 2021 for details). It is not included in Section 5.5 because we do not have access to the complete gas information. In contrast to the radiative transfer simulations of our fiducial model, Nasir & D'Aloisio (2020) employ approximate "semi-numeric" methods to model reionization's effects on the forest. For this model, we include all Lyβ dark gaps with $L \geq 7\ h^{-1}$ Mpc, except for the fully neutral ones. Here, this model serves as an independent reference for dark gaps that are not dominated by neutral hydrogen islands.

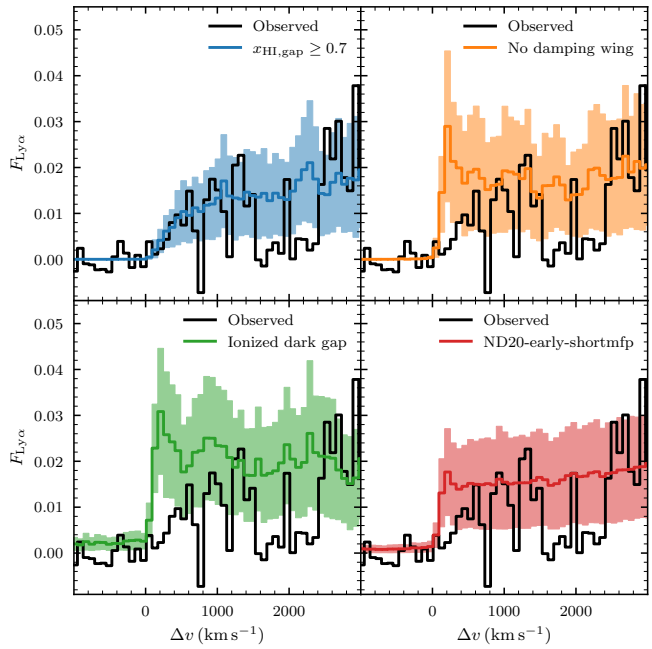
In Figure 5.3 (a), we compare the stacked Lyα forest in models with the observed stack. As with the ideal mocks, the stack of dark gaps containing a significant fraction of neutral gas exhibits an extended and damping wing-like profile in the Lyα forest. The observed curve broadly aligns with the model profile and falls within the 68% interval. As expected, the "no damping wing" and "ionized dark gap" cases show a rapid increase in the stacked Lyα transmission, inconsistent with the observed damping wing-like profile. The

"ND20-early-shortmfp" case also displays a steep increase in the stacked Lyα transmission, albeit with a slightly lower maximum flux compared to the "no damping wing" and "ionized dark gap" cases.

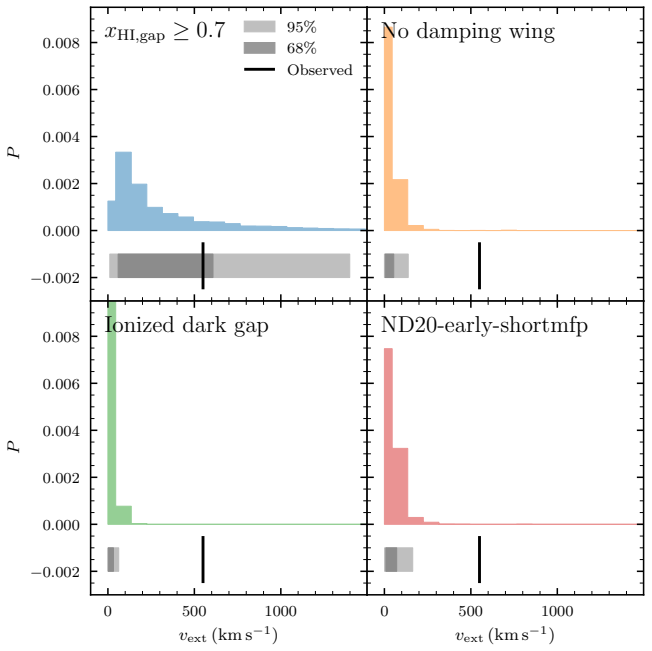
For a quantitative comparison, we fit the upper envelope of the model-predicted stacked Lyα transmission in each bootstrap realization to Equation 5.1 and plot the distribution of the damping wing parameter $v_{\rm ext}$ in Figure 5.3 (b). Only the scenario where dark gaps contain a large fraction of neutral gas predicts a broad distribution of $v_{\rm ext}$, spanning the observed value within the 68% interval. Conversely, scenarios that fail to reproduce an extended absorption profile predict a very narrow $v_{\rm ext}$ distribution, confined to $v_{\rm ext} \lesssim 200$ km s^{-1} , with the observed $v_{\rm ext}$ lying beyond their 95% confidence limits. These results suggest that the observed damping wing-like profile in the stacked Lyα transmission likely originates from genuine neutral IGM absorption at $z < 6$, as opposed to fully ionized gas that is highly opaque to due to a fluctuating UVB or high densities.

5.7 Lower Limit Constraint on Neutral Hydrogen Fraction

We can use the observed damping wing-like profile to constrain the hydrogen neutral fraction in unbiased locations within the IGM. To do this, we generate dark gaps following the observed gap length distribution for different ranges of $x_{\rm HI,gap}$, specifically $0.1 \leq x_{\rm HI,gap} < 0.2$, $0.2 \leq x_{\rm HI,gap} < 0.3$, ..., up to $0.9 \leq x_{\rm HI,gap} < 1.0$, and create mock Ly$\alpha$ forest stacks for each range. We find that only $x_{\rm HI,gap}$ values greater than 0.3 can predict $v_{\rm ext}$ consistent with the observed value within the 95% limit. After convolving $x_{\rm HI,gap} \geq 0.4$



(a)



(b)

Figure 5.3: (a) Comparison of the stacked Lyα forest at the red edge of Lyβ dark gaps in mock spectra to the observations. The shaded regions show the 68% intervals of model predictions from bootstrap trials. (b) Distributions of the damping wing parameter, $v_{\rm ext}$, in the models. The light and dark shaded horizontal bars in each subplot show the 68% and 90% intervals, respectively. The vertical black line denotes $v_{\rm ext}$ acquired from the observation.

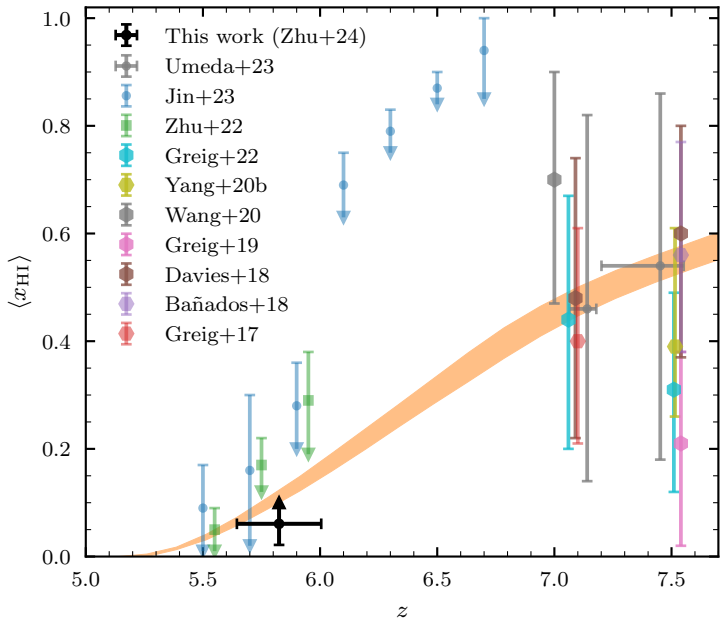


Figure 5.4: Lower limit constraint on the volume-averaged neutral hydrogen fraction from the damping wing-like profile in the stacked Lyα forest based on Lyβ dark gaps over $5.6 < z < 6.0$. We also plot $\langle x_{\rm HI} \rangle$ constraints from recent observations for reference: upper limit constraints from Lyα + Lyβ dark gap statistics (Zhu et al. 2022), upper limit constraints from the fraction of Lyα + Lyβ dark pixels (Jin et al. 2023), measurements from high-z QSO damping wings (Greig et al. 2017, 2019, 2022; Bañados et al. 2018; Davies et al. 2018b; Wang et al. 2020; Yang et al. 2020a), and measurements from the Lyα damping wing of high-z galaxies (Umeda et al. 2023). The orange shade plot the posterior from inference based on multiple reionization observables (Y. Qin et al. in prep.).

with the fraction of the Lyα forest subtended by dark gaps longer than the mean length of dark gaps used in this work, $f_{L\geq 15\ h^{-1}\,\mathrm{Mpc}} = 0.152 \pm 0.098$, (Zhu et al. 2022) we derive a lower limit on the volume-averaged neutral hydrogen fraction of $\langle x_{\mathrm{HI}} \rangle \geq 0.061 \pm 0.039$ at $z = 5.82$. This constraint is conservative as it excludes potential neutral hydrogen in dark gaps with $L < 7\ h^{-1}\,\mathrm{Mpc}$, and our model incorporates significant UVB fluctuations associated with the neutral islands.

Figure 5.4 displays our constraint alongside other recent $\langle x_{\mathrm{HI}} \rangle$ measurements. Our constraint on $\langle x_{\mathrm{HI}} \rangle$ suggests that the IGM must be at least 6.1% neutral even down to $z = 5.8$. Alongside upper limit constraints from dark gaps and dark pixels (Zhu et al. 2022; Jin et al. 2023), these observations accommodate a wide range of late-ending reionization histories, with $\langle x_{\mathrm{HI}} \rangle$ ranging from approximately 6% to 20% near $z = 5.8$. We also note that the inference based on multiple recent reionization observables (Y. Qin et al. in prep.), which includes the mean transmission in the Lyα forest, galaxy UV luminosity functions, and CMB observations, aligns well with these $\langle x_{\mathrm{HI}} \rangle$ constraints.

5.8 Summary

We have identified a damping wing-like profile in the stacked Lyα forest at the redshift of the red edge of Lyβ dark gaps at $z < 6$, using data from the XQR-30 and archival spectra. A comparison with mock spectra generated from simulations suggests that this damping wing-like profile most likely arises from dark gaps containing genuine neutral hydrogen islands, as opposed to fully ionized regions, whose high opacities are the result of high densities and/or a locally low UVB. We further ascertain that such a profile

necessitates a neutral hydrogen fraction of at least $6.1 \pm 3.9\%$ at $z = 5.8$. Combined with other recent observations, our results provide robust evidence that reionization extends to $z < 6$. Looking into the future, a larger sample of long Lyβ dark gaps will set even more precise constraints on the timeline of reionization, and this can be achieved with the help of upcoming Extremely Large Telescopes (ELTs) given their unprecedented sensitivity and efficiency in acquiring high-z QSO spectra.

Chapter 6

Conclusions

The IGM *observed* everything happened during the epoch of reionization. The timing of and sources responsible for reionization carry important implications of how early galaxies form and evolve. Determining **when and how reionization occurred**, therefore, has been a frontier astrophysics topic, to which efforts are devoted by current and future facilities, e.g., JWST, SPHEREx, HERA, and SKA. My PhD research studied the timeline of reionization with absorption lines in QSO spectra using observations from Keck and VLT as well as novel forward-modeling methods. This dissertation reports robust observational constraints on reionization:

- In Chapter 2, we searched 55 high-quality spectra of QSOs for dark gaps ($> 95\%$ Lyα photons absorbed), and discovered that long ($L \geq 30h^{-1}$Mpc) dark gaps extending down to $z \simeq 5.4$. To distinguish the competing models, we built mock spectra from simulations that use different reionization histories. By comparing the dark gap distribution in the observed spectra with mock data, our results suggest that the

scenario wherein reionization fully completes by $z = 6$ with a homogeneous UV background is disfavored, but reionization models with neutral islands and/or large-scale fluctuations in the UVB persisting down to $z \lesssim 5.5$ are consistent with the data.

- In the work presented in Chapter 3, we took advantage of new $z \sim 6$ QSO spectra from our XQR-30 VLT large program. The large sample size and the high quality of the spectra made measuring Lyβ dark gaps with a enough path-length possible. We report the discovery of an extremely long dark gap in the Lyβ forest persisting down to $z \sim 5.5$. We also derived the upper limits on IGM neutral fraction from dark gap length distributions based on a conservative model. Along with the high fraction of QSO spectra showing long Lyβ dark gaps, our results suggest that the IGM can still be considerably neutral ($\sim 15\%$) by $z = 5.8$.

- Chapter 4 discusses the direct measurements of ionizing photon mean free path. We conducted Keck/ESI observations to measure $\lambda_{\rm mfp}$ at $z \sim 5.6$ for the first time using recently discovered QSOs in Yang et al. (2017, 2019b). Following Becker et al. (2021), we take the potential bias on the measurements from QSO proximity zone effect into account in the modeling of LyC transmission. We convincingly demonstrated a rapid evolution of $\lambda_{\rm mfp}$ over $5 < z < 6$ and provided the strongest evidence to date that reionization ends significantly below $z = 6$.

- Chapter 5 reports the discovery of damping wing-like features in the stacked Lyα forest at the red edge of these Lyβ dark gaps. These features cannot be explained by a fully ionized IGM at $z < 6$ and require a neutral hydrogen fraction of $\gtrsim 5\%$ even down to $z = 5.8$ in the unbiased locations. The new constraints on neutral fractions

are consistent with later-ending reionization scenarios, and potentially help relax the tension between the galaxy emissivity and the late midpoint of reionization.

In conclusion, the findings of this work suggest that reionization may have concluded significantly later than the previously believed redshift of $z = 6$, necessitating a revision of the chronology of the early Universe. Such an ultra-late reionization scenario calls for a reevaluation of early Universe models, particularly concerning the emergence and influence of the first luminous sources. The extended timeline of reionization implies that the first stars and galaxies might have formed and evolved more gradually (e.g., Nakane et al. 2023).

The exploration of the IGM presented in this dissertation not only advances our understanding of the reionization epoch but also opens avenues for broader astrophysical inquiries. For example, a better understanding of the connection between IGM opacity and galaxy densities is of great interest in the ultra-late reionization framework (e.g., Christenson et al. 2021, 2023). From a cosmology and particle physics perspective, measuring the reionization history may help probe small-scale perturbations (e.g., Minoda et al. 2023) and constrain the properties of dark matter particles (e.g., Cain et al. 2022). The methodologies used in this work also inspired lower-redshift explorations, e.g., using dark gaps and transmission peaks in the He II Lyα forest to constrain Helium reionization at $2 \lesssim z \lesssim 3$ (Makan et al. 2022).

Our efforts in constraining reionization using QSO spectra will complement and enable upcoming studies of reionization, including those using JWST, ALMA, and 21-cm experiments. Looking into the future, with the upcoming SPHEREx mission, the QSO

sample at $z > 6$ will grow significantly, and approximately 1100 of the most luminous QSOs will be discovered (Doré et al. 2018). Combined with the unprecedented sensitivity of Extremely Large Telescopes (ELTs), hydrogen absorption lines in high-redshift QSOs will greatly aid in answering important questions posed in the *Astro2020 Decadal Survey*, such as *"how did the intergalactic medium and the first sources of radiation evolve from cosmic dawn through the epoch of reionization?"*

Bibliography

Aubert, D., & Teyssier, R. 2008, Monthly Notices of the Royal Astronomical Society, 387, 295, doi: 10.1111/j.1365-2966.2008.13223.x

Aubert, D., & Teyssier, R. 2010, The Astrophysical Journal, 724, 244, doi: 10.1088/0004-637X/724/1/244

Bañados, E., Decarli, R., Walter, F., et al. 2015, The Astrophysical Journal Letters, 805, L8, doi: 10.1088/2041-8205/805/1/L8

Bañados, E., Venemans, B. P., Decarli, R., et al. 2016, The Astrophysical Journal Supplement Series, 227, 11, doi: 10.3847/0067-0049/227/1/11

Bañados, E., Venemans, B. P., Mazzucchelli, C., et al. 2018, Nature, 553, 473, doi: 10.1038/nature25180

Bañados, E., Schindler, J.-T., Venemans, B. P., et al. 2023, The Astrophysical Journal Supplement Series, 265, 29, doi: 10.3847/1538-4365/acb3c7

Barnett, R., Warren, S. J., Becker, G. D., et al. 2017, Astronomy and Astrophysics, 601, A16, doi: 10.1051/0004-6361/201630258

Becker, G. D., Bolton, J. S., Madau, P., et al. 2015, Monthly Notices of the Royal Astronomical Society, 447, 3402, doi: 10.1093/mnras/stu2646

Becker, G. D., D'Aloisio, A., Christenson, H. M., et al. 2021, Monthly Notices of the Royal Astronomical Society, 508, 1853, doi: 10.1093/mnras/stab2696

Becker, G. D., Davies, F. B., Furlanetto, S. R., et al. 2018, The Astrophysical Journal, 863, 92, doi: 10.3847/1538-4357/aacc73

Becker, G. D., Pettini, M., Rafelski, M., et al. 2019, The Astrophysical Journal, 883, 163, doi: 10.3847/1538-4357/ab3eb5

Bischetti, M., Feruglio, C., D'Odorico, V., et al. 2022, Nature, 605, 244, doi: 10.1038/s41586-022-04608-1

Boera, E., Becker, G. D., Bolton, J. S., & Nasir, F. 2019, The Astrophysical Journal, 872, 101, doi: 10.3847/1538-4357/aafee4

Bolton, J. S. 2007, The Observatory, 127, 262

Bolton, J. S., Puchwein, E., Sijacki, D., et al. 2017, Monthly Notices of the Royal Astronomical Society, 464, 897, doi: 10.1093/mnras/stw2397

Bosman, S. E. I. 2021, Constraints on the Mean Free Path of Ionising Photons at $z\sim6$ Using Limits on Individual Free Paths, doi: 10.48550/arXiv.2108.12446

Bosman, S. E. I., Ďurovčíková, D., Davies, F. B., & Eilers, A. C. 2021, Monthly Notices of the Royal Astronomical Society, 503, 2077, doi: 10.1093/mnras/stab572

Bosman, S. E. I., Fan, X., Jiang, L., et al. 2018, Monthly Notices of the Royal Astronomical Society, 479, 1055, doi: 10.1093/mnras/sty1344

Bosman, S. E. I., Davies, F. B., Becker, G. D., et al. 2022, Monthly Notices of the Royal Astronomical Society, 514, 55, doi: 10.1093/mnras/stac1046

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, The Astrophysical Journal, 811, 140, doi: 10.1088/0004-637X/811/2/140

Bromm, V., & Larson, R. B. 2004, Annual Review of Astronomy and Astrophysics, 42, 79, doi: 10.1146/annurev.astro.42.053102.134034

Cain, C., D'Aloisio, A., Gangolli, N., & Becker, G. D. 2021, The Astrophysical Journal, 917, L37, doi: 10.3847/2041-8213/ac1ace

Cain, C., D'Aloisio, A., Gangolli, N., & McQuinn, M. 2023, Monthly Notices of the Royal Astronomical Society, 522, 2047, doi: 10.1093/mnras/stad1057

Cain, C., D'Aloisio, A., Irsic, V., Gangolli, N., & Dhami, S. 2022, Small-Scale Clumping of Dark Matter and the Mean Free Path of Ionizing Photons at $z=6$, arXiv. http://ascl.net/2207.02876

Calverley, A. P., Becker, G. D., Haehnelt, M. G., & Bolton, J. S. 2011, Monthly Notices of the Royal Astronomical Society, 412, 2543, doi: 10.1111/j.1365-2966.2010.18072.x

Carilli, C. L., Neri, R., Wang, R., et al. 2007, The Astrophysical Journal, 666, L9, doi: 10.1086/521648

CASA Team, Bean, B., Bhatnagar, S., et al. 2022, Publications of the Astronomical Society of the Pacific, 134, 114501, doi: 10.1088/1538-3873/ac9642

Chardin, J., Puchwein, E., & Haehnelt, M. G. 2017, Monthly Notices of the Royal Astronomical Society, 465, 3429, doi: 10.1093/mnras/stw2943

Chen, H., Eilers, A.-C., Bosman, S. E. I., et al. 2022, The Astrophysical Journal, 931, 29, doi: 10.3847/1538-4357/ac658d

Chen, S.-F. S., Simcoe, R. A., Torrey, P., et al. 2017, The Astrophysical Journal, 850, 188, doi: 10.3847/1538-4357/aa9707

Choudhury, T. R., Paranjape, A., & Bosman, S. E. I. 2021, Monthly Notices of the Royal Astronomical Society, 501, 5782, doi: 10.1093/mnras/stab045

Christenson, H. M., Becker, G. D., Furlanetto, S. R., et al. 2021, The Astrophysical Journal, 923, 87, doi: 10.3847/1538-4357/ac2a34

Christenson, H. M., Becker, G. D., D'Aloisio, A., et al. 2023, The Astrophysical Journal, 955, 138, doi: 10.3847/1538-4357/acf450

D'Aloisio, A., McQuinn, M., Davies, F. B., & Furlanetto, S. R. 2018, Monthly Notices of the Royal Astronomical Society, 473, 560, doi: 10.1093/mnras/stx2341

D'Aloisio, A., McQuinn, M., & Trac, H. 2015, The Astrophysical Journal Letters, 813, L38, doi: 10.1088/2041-8205/813/2/L38

D'Aloisio, A., McQuinn, M., Trac, H., Cain, C., & Mesinger, A. 2020, The Astrophysical Journal, 898, 149, doi: 10.3847/1538-4357/ab9f2f

Davies, F. B. 2020, Monthly Notices of the Royal Astronomical Society, 494, 2937, doi: 10.1093/mnras/staa528

Davies, F. B., Becker, G. D., & Furlanetto, S. R. 2018a, The Astrophysical Journal, 860, 155, doi: 10.3847/1538-4357/aac2d6

Davies, F. B., Bosman, S. E. I., Furlanetto, S. R., Becker, G. D., & D'Aloisio, A. 2021, The Astrophysical Journal, 918, L35, doi: 10.3847/2041-8213/ac1ffb

Davies, F. B., & Furlanetto, S. R. 2016, Monthly Notices of the Royal Astronomical Society, 460, 1328, doi: 10.1093/mnras/stw931

Davies, F. B., Hennawi, J. F., Bañados, E., et al. 2018b, The Astrophysical Journal, 864, 142, doi: 10.3847/1538-4357/aad6dc

Davies, F. B., Hennawi, J. F., Bañados, E., et al. 2018c, The Astrophysical Journal, 864, 143, doi: 10.3847/1538-4357/aad7f8

Davies, R. L., Ryan-Weber, E., D'Odorico, V., et al. 2023a, Monthly Notices of the Royal Astronomical Society, 521, 314, doi: 10.1093/mnras/stad294

Davies, R. L., Ryan-Weber, E., D'Odorico, V., et al. 2023b, Monthly Notices of the Royal Astronomical Society, 521, 289, doi: 10.1093/mnras/stac3662

Dayal, P., & Ferrara, A. 2018, Physics Reports, 780, 1, doi: 10.1016/j.physrep.2018.10.002

de Belsunce, R., Gratton, S., Coulton, W., & Efstathiou, G. 2021, Monthly Notices of the Royal Astronomical Society, 507, 1072, doi: 10.1093/mnras/stab2215

Decarli, R., Walter, F., Venemans, B. P., et al. 2018, The Astrophysical Journal, 854, 97, doi: 10.3847/1538-4357/aaa5aa

D'Odorico, V., Banados, E., Becker, G. D., et al. 2023, XQR-30: The Ultimate XSHOOTER Quasar Sample at the Reionization Epoch, doi: 10.48550/arXiv.2305.05053

Doré, O., Werner, M. W., Ashby, M. L. N., et al. 2018, Science Impacts of the SPHEREx All-Sky Optical to Near-Infrared Spectral Survey II: Report of a Community Workshop on the Scientific Synergies Between the SPHEREx Survey and Other Astronomy Observatories, doi: 10.48550/arXiv.1805.05489

Eilers, A.-C., Davies, F. B., & Hennawi, J. F. 2018, The Astrophysical Journal, 864, 53, doi: 10.3847/1538-4357/aad4fd

Eilers, A.-C., Davies, F. B., Hennawi, J. F., et al. 2017, The Astrophysical Journal, 840, 24, doi: 10.3847/1538-4357/aa6c60

Eilers, A.-C., Hennawi, J. F., Davies, F. B., & Oñorbe, J. 2019, The Astrophysical Journal, 881, 23, doi: 10.3847/1538-4357/ab2b3f

Eilers, A.-C., Hennawi, J. F., Decarli, R., et al. 2020, The Astrophysical Journal, 900, 37, doi: 10.3847/1538-4357/aba52e

Fan, X., Bañados, E., & Simcoe, R. A. 2023, Annual Review of Astronomy and Astrophysics, 61, 373, doi: 10.1146/annurev-astro-052920-102455

Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, The Astronomical Journal, 132, 117, doi: 10.1086/504836

Farina, E. P., Arrigoni-Battaia, F., Costa, T., et al. 2019, The Astrophysical Journal, 887, 196, doi: 10.3847/1538-4357/ab5847

Finkelstein, S. L., D'Aloisio, A., Paardekooper, J.-P., et al. 2019, The Astrophysical Journal, 879, 36, doi: 10.3847/1538-4357/ab1ea8

Fumagalli, M., O'Meara, J. M., Prochaska, J. X., & Worseck, G. 2013, The Astrophysical Journal, 775, 78, doi: 10.1088/0004-637X/775/1/78

Furlanetto, S. R., Hernquist, L., & Zaldarriaga, M. 2004, Monthly Notices of the Royal Astronomical Society, 354, 695, doi: 10.1111/j.1365-2966.2004.08225.x

Furlanetto, S. R., & Oh, S. P. 2005, Monthly Notices of the Royal Astronomical Society, 363, 1031, doi: 10.1111/j.1365-2966.2005.09505.x

Gaikwad, P., Srianand, R., Haehnelt, M. G., & Choudhury, T. R. 2021, Monthly Notices of the Royal Astronomical Society, 506, 4389, doi: 10.1093/mnras/stab2017

Gaikwad, P., Rauch, M., Haehnelt, M. G., et al. 2020, Monthly Notices of the Royal Astronomical Society, 494, 5091, doi: 10.1093/mnras/staa907

Gaikwad, P., Haehnelt, M. G., Davies, F. B., et al. 2023, Measuring the Photo-Ionization Rate, Neutral Fraction and Mean Free Path of HI Ionizing Photons at $4.9 \leq z \leq 6.0$ from a Large Sample of XShooter and ESI Spectra, arXiv. http://ascl.net/2304.02038

Gallerani, S., Ferrara, A., Fan, X., & Choudhury, T. R. 2008, Monthly Notices of the Royal Astronomical Society, 386, 359, doi: 10.1111/j.1365-2966.2008.13029.x

Garaldi, E., Kannan, R., Smith, A., et al. 2022, Monthly Notices of the Royal Astronomical Society, 512, 4909, doi: 10.1093/mnras/stac257

Gnedin, N. Y. 2022, The Astrophysical Journal, 937, 17, doi: 10.3847/1538-4357/ac8d0a

Gnedin, N. Y., Becker, G. D., & Fan, X. 2017, The Astrophysical Journal, 841, 26, doi: 10.3847/1538-4357/aa6c24

Greig, B., Mesinger, A., & Bañados, E. 2019, Monthly Notices of the Royal Astronomical Society, 484, 5094, doi: 10.1093/mnras/stz230

Greig, B., Mesinger, A., Davies, F. B., et al. 2022, Monthly Notices of the Royal Astronomical Society, 512, 5390, doi: 10.1093/mnras/stac825

Greig, B., Mesinger, A., Haiman, Z., & Simcoe, R. A. 2017, Monthly Notices of the Royal Astronomical Society, 466, 4239, doi: 10.1093/mnras/stw3351

Gunn, J. E., & Peterson, B. A. 1965, The Astrophysical Journal, 142, 1633, doi: 10.1086/148444

Haardt, F., & Madau, P. 2012, The Astrophysical Journal, 746, 125, doi: 10.1088/0004-637X/746/2/125

Hoag, A., Bradač, M., Huang, K., et al. 2019, The Astrophysical Journal, 878, 12, doi: 10.3847/1538-4357/ab1de7

Horne, K. 1986, Publications of the Astronomical Society of the Pacific, 98, 609, doi: 10.1086/131801

Hu, W., Wang, J., Zheng, Z.-Y., et al. 2019, The Astrophysical Journal, 886, 90, doi: 10.3847/1538-4357/ab4cf4

Jiang, L., Fan, X., Vestergaard, M., et al. 2007, The Astronomical Journal, 134, 1150, doi: 10.1086/520811

Jin, X., Yang, J., Fan, X., et al. 2023, The Astrophysical Journal, 942, 59, doi: 10.3847/1538-4357/aca678

Jones, A., Noll, S., Kausch, W., Szyszka, C., & Kimeswenger, S. 2013, Astronomy and Astrophysics, 560, A91, doi: 10.1051/0004-6361/201322433

Jung, I., Finkelstein, S. L., Dickinson, M., et al. 2020, The Astrophysical Journal, 904, 144, doi: 10.3847/1538-4357/abbd44

Kashino, D., Lilly, S. J., Matthee, J., et al. 2023, The Astrophysical Journal, 950, 66, doi: 10.3847/1538-4357/acc588

Kashino, D., Lilly, S. J., Shibuya, T., Ouchi, M., & Kashikawa, N. 2020, The Astrophysical Journal, 888, 6, doi: 10.3847/1538-4357/ab5a7d

Keating, L. C., Kulkarni, G., Haehnelt, M. G., Chardin, J., & Aubert, D. 2020a, Monthly Notices of the Royal Astronomical Society, 497, 906, doi: 10.1093/mnras/staa1909

Keating, L. C., Weinberger, L. H., Kulkarni, G., et al. 2020b, Monthly Notices of the Royal Astronomical Society, 491, 1736, doi: 10.1093/mnras/stz3083

Kelson, D. D. 2003, Publications of the Astronomical Society of the Pacific, 115, 688, doi: 10.1086/375502

Kogut, A., Spergel, D. N., Barnes, C., et al. 2003, The Astrophysical Journal Supplement Series, 148, 161, doi: 10.1086/377219

Kulkarni, G., Keating, L. C., Haehnelt, M. G., et al. 2019a, Monthly Notices of the Royal Astronomical Society, 485, L24, doi: 10.1093/mnrasl/slz025

Kulkarni, G., Worseck, G., & Hennawi, J. F. 2019b, Monthly Notices of the Royal Astronomical Society, 488, 1035, doi: 10.1093/mnras/stz1493

Kurk, J. D., Walter, F., Fan, X., et al. 2007, The Astrophysical Journal, 669, 32, doi: 10.1086/521596

Lai, S., Bian, F., Onken, C. A., et al. 2022, Monthly Notices of the Royal Astronomical Society, 513, 1801, doi: 10.1093/mnras/stac1001

Lewis, J. S. W., Ocvirk, P., Sorce, J. G., et al. 2022, Monthly Notices of the Royal Astronomical Society, 516, 3389, doi: 10.1093/mnras/stac2383

Lusso, E., Fumagalli, M., Rafelski, M., et al. 2018, The Astrophysical Journal, 860, 41, doi: 10.3847/1538-4357/aac514

Lusso, E., Worseck, G., Hennawi, J. F., et al. 2015, Monthly Notices of the Royal Astronomical Society, 449, 4204, doi: 10.1093/mnras/stv516

Maiolino, R., Cox, P., Caselli, P., et al. 2005, Astronomy and Astrophysics, 440, L51, doi: 10.1051/0004-6361:200500165

Makan, K., Worseck, G., Davies, F. B., et al. 2022, The Astrophysical Journal, 927, 175, doi: 10.3847/1538-4357/ac524a

Malloy, M., & Lidz, A. 2015, The Astrophysical Journal, 799, 179, doi: 10.1088/0004-637X/799/2/179

Mann, H. B., & Whitney, D. R. 1947, The Annals of Mathematical Statistics, 18, 50, doi: 10.1214/aoms/1177730491

Mason, C. A., Treu, T., Dijkstra, M., et al. 2018, The Astrophysical Journal, 856, 2, doi: 10.3847/1538-4357/aab0a7

Mason, C. A., Fontana, A., Treu, T., et al. 2019, Monthly Notices of the Royal Astronomical Society, 485, 3947, doi: 10.1093/mnras/stz632

Mazzucchelli, C., Bañados, E., Venemans, B. P., et al. 2017, The Astrophysical Journal, 849, 91, doi: 10.3847/1538-4357/aa9185

McGreer, I. D., Fan, X., Jiang, L., & Cai, Z. 2018, The Astronomical Journal, 155, 131, doi: 10.3847/1538-3881/aaaab4

McGreer, I. D., Mesinger, A., & D'Odorico, V. 2015, Monthly Notices of the Royal Astronomical Society, 447, 499, doi: 10.1093/mnras/stu2449

McGreer, I. D., Mesinger, A., & Fan, X. 2011, Monthly Notices of the Royal Astronomical Society, 415, 3237, doi: 10.1111/j.1365-2966.2011.18935.x

McGreer, I. D., Jiang, L., Fan, X., et al. 2013, The Astrophysical Journal, 768, 105, doi: 10.1088/0004-637X/768/2/105

McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, 376, 127

McQuinn, M. 2016, Annual Review of Astronomy and Astrophysics, 54, 313, doi: 10.1146/annurev-astro-082214-122355

McQuinn, M., Oh, S. P., & Faucher-Giguère, C.-A. 2011, The Astrophysical Journal, 743, 82, doi: 10.1088/0004-637X/743/1/82

Meiksin, A. 2020, Monthly Notices of the Royal Astronomical Society, 491, 4884, doi: 10.1093/mnras/stz3395

Mesinger, A., Furlanetto, S., & Cen, R. 2011, Monthly Notices of the Royal Astronomical Society, 411, 955, doi: 10.1111/j.1365-2966.2010.17731.x

Minoda, T., Yoshiura, S., & Takahashi, T. 2023, Physical Review D, 108, 123542, doi: 10.1103/PhysRevD.108.123542

Muñoz, J. B., Qin, Y., Mesinger, A., et al. 2022, Monthly Notices of the Royal Astronomical Society, 511, 3657, doi: 10.1093/mnras/stac185

Murray, S., Greig, B., Mesinger, A., et al. 2020, The Journal of Open Source Software, 5, 2582, doi: 10.21105/joss.02582

Naidu, R. P., Tacchella, S., Mason, C. A., et al. 2020, The Astrophysical Journal, 892, 109, doi: 10.3847/1538-4357/ab7cc9

Nakane, M., Ouchi, M., Nakajima, K., et al. 2023, Lyα Emission at $z=7-13$: Clear Lyα Equivalent Width Evolution Indicating the Late Cosmic Reionization History, doi: 10.48550/arXiv.2312.06804

Nasir, F., & D'Aloisio, A. 2020, Monthly Notices of the Royal Astronomical Society, 494, 3080, doi: 10.1093/mnras/staa894

Noll, S., Kausch, W., Barden, M., et al. 2012, Astronomy and Astrophysics, 543, A92, doi: 10.1051/0004-6361/201219040

Ocvirk, P., Lewis, J. S. W., Gillet, N., et al. 2021, Monthly Notices of the Royal Astronomical Society, 507, 6108, doi: 10.1093/mnras/stab2502

O'Meara, J. M., Prochaska, J. X., Worseck, G., Chen, H.-W., & Madau, P. 2013, The Astrophysical Journal, 765, 137, doi: 10.1088/0004-637X/765/2/137

Ostriker, J. P., & Steinhardt, P. 2003, Science, 300, 1909, doi: 10.1126/science.1085976

Park, H., Shapiro, P. R., Choi, J.-h., et al. 2016, The Astrophysical Journal, 831, 86, doi: 10.3847/0004-637X/831/1/86

Parsa, S., Dunlop, J. S., & McLure, R. J. 2018, Monthly Notices of the Royal Astronomical Society, 474, 2904, doi: 10.1093/mnras/stx2887

Paschos, P., & Norman, M. L. 2005, The Astrophysical Journal, 631, 59, doi: 10.1086/431787

Planck Collaboration, Ade, P. a. R., Aghanim, N., et al. 2014, Astronomy and Astrophysics, 571, A16, doi: 10.1051/0004-6361/201321591

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, Astronomy and Astrophysics, 641, A6, doi: 10.1051/0004-6361/201833910

Prochaska, J. X., Gawiser, E., Wolfe, A. M., Cooke, J., & Gelino, D. 2003, The Astrophysical Journal Supplement Series, 147, 227, doi: 10.1086/375839

Prochaska, J. X., Worseck, G., & O'Meara, J. M. 2009, The Astrophysical Journal Letters, 705, L113, doi: 10.1088/0004-637X/705/2/L113

Puchwein, E., Haardt, F., Haehnelt, M. G., & Madau, P. 2019, Monthly Notices of the Royal Astronomical Society, 485, 47, doi: 10.1093/mnras/stz222

Qin, Y., Mesinger, A., Bosman, S. E. I., & Viel, M. 2021, arXiv e-prints, 2101, arXiv:2101.09033. http://ascl.net/2101.09033

Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, The Astrophysical Journal, 802, L19, doi: 10.1088/2041-8205/802/2/L19

Roth, J. T., D'Aloisio, A., Cain, C., et al. 2023, The Effect of Reionization on Direct Measurements of the Mean Free Path, arXiv. http://ascl.net/2311.06348

Rutkowski, M. J., Scarlata, C., Haardt, F., et al. 2016, The Astrophysical Journal, 819, 81, doi: 10.3847/0004-637X/819/1/81

Satyavolu, S., Kulkarni, G., Keating, L. C., & Haehnelt, M. G. 2023a, Robustness of Direct Measurements of the Mean Free Path of Ionizing Photons in the Epoch of Reionization, arXiv. http://ascl.net/2311.06344

Satyavolu, S., Eilers, A.-C., Kulkarni, G., et al. 2023b, Monthly Notices of the Royal Astronomical Society, 522, 4918, doi: 10.1093/mnras/stad1326

Schindler, J.-T., Farina, E. P., Bañados, E., et al. 2020, The Astrophysical Journal, 905, 51, doi: 10.3847/1538-4357/abc2d7

Shapiro, P. R., Iliev, I. T., & Raga, A. C. 2004, Monthly Notices of the Royal Astronomical Society, 348, 753, doi: 10.1111/j.1365-2966.2004.07364.x

Sharma, M., Theuns, T., Frenk, C., et al. 2016, Monthly Notices of the Royal Astronomical Society, 458, L94, doi: 10.1093/mnrasl/slw021

Sheinis, A. I., Bolte, M., Epps, H. W., et al. 2002, Publications of the Astronomical Society of the Pacific, 114, 851, doi: 10.1086/341706

Shen, Y., Wu, J., Jiang, L., et al. 2019, The Astrophysical Journal, 873, 35, doi: 10.3847/1538-4357/ab03d9

Shull, J. M., Stevans, M., & Danforth, C. W. 2012, The Astrophysical Journal, 752, 162, doi: 10.1088/0004-637X/752/2/162

Sobacchi, E., & Mesinger, A. 2014, Monthly Notices of the Royal Astronomical Society, 440, 1662, doi: 10.1093/mnras/stu377

Songaila, A., & Cowie, L. L. 2002, The Astronomical Journal, 123, 2183, doi: 10.1086/340079

Springel, V. 2005, Monthly Notices of the Royal Astronomical Society, 364, 1105, doi: 10.1111/j.1365-2966.2005.09655.x

Stark, D. P. 2016, Annual Review of Astronomy and Astrophysics, 54, 761, doi: 10.1146/annurev-astro-081915-023417

Stevans, M. L., Shull, J. M., Danforth, C. W., & Tilton, E. M. 2014, The Astrophysical Journal, 794, 75, doi: 10.1088/0004-637X/794/1/75

Telfer, R. C., Zheng, W., Kriss, G. A., & Davidsen, A. F. 2002, The Astrophysical Journal, 565, 773, doi: 10.1086/324689

Trac, H., & Pen, U.-L. 2004, New Astronomy, 9, 443, doi: 10.1016/j.newast.2004.02.002

Umeda, H., Ouchi, M., Nakajima, K., et al. 2023, JWST Measurements of Neutral Hydrogen Fractions and Ionized Bubble Sizes at $z=7-12$ Obtained with Lyα Damping Wing Absorptions in 26 Bright Continuum Galaxies, arXiv. http://ascl.net/2306.00487

Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, The Astronomical Journal, 122, 549, doi: 10.1086/321167

Venemans, B. P., Walter, F., Zschaechner, L., et al. 2016, The Astrophysical Journal, 816, 37, doi: 10.3847/0004-637X/816/1/37

Venemans, B. P., Walter, F., Neeleman, M., et al. 2020, The Astrophysical Journal, 904, 130, doi: 10.3847/1538-4357/abc563

Vernet, J., Dekker, H., D'Odorico, S., et al. 2011, Astronomy and Astrophysics, 536, A105, doi: 10.1051/0004-6361/201117752

Villasenor, B., Robertson, B., Madau, P., & Schneider, E. 2021, arXiv:2111.00019 [astro-ph]. http://ascl.net/2111.00019

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: 10.1038/s41592-019-0686-2

Walther, M., Oñorbe, J., Hennawi, J. F., & Lukić, Z. 2019, The Astrophysical Journal, 872, 13, doi: 10.3847/1538-4357/aafad1

Wang, F., Wang, R., Fan, X., et al. 2019, The Astrophysical Journal, 880, 2, doi: 10.3847/1538-4357/ab2717

Wang, F., Wu, X.-B., Fan, X., et al. 2016, The Astrophysical Journal, 819, 24, doi: 10.3847/0004-637X/819/1/24

Wang, F., Davies, F. B., Yang, J., et al. 2020, The Astrophysical Journal, 896, 23, doi: 10.3847/1538-4357/ab8c45

Wang, F., Fan, X., Yang, J., et al. 2021, The Astrophysical Journal, 908, 53, doi: 10.3847/1538-4357/abcc5e

Wang, R., Carilli, C. L., Neri, R., et al. 2010, The Astrophysical Journal, 714, 699, doi: 10.1088/0004-637X/714/1/699

Wang, R., Wagg, J., Carilli, C. L., et al. 2013, The Astrophysical Journal, 773, 44, doi: 10.1088/0004-637X/773/1/44

Weinberger, L. H., Haehnelt, M. G., & Kulkarni, G. 2019, Monthly Notices of the Royal Astronomical Society, 485, 1350, doi: 10.1093/mnras/stz481

Wold, I. G. B., Malhotra, S., Rhoads, J., et al. 2021, arXiv e-prints, 2105, arXiv:2105.12191

Worseck, G., Prochaska, J. X., O'Meara, J. M., et al. 2014, Monthly Notices of the Royal Astronomical Society, 445, 1745, doi: 10.1093/mnras/stu1827

Yang, J., Fan, X., Wu, X.-B., et al. 2017, The Astronomical Journal, 153, 184, doi: 10.3847/1538-3881/aa6577

Yang, J., Venemans, B., Wang, F., et al. 2019a, The Astrophysical Journal, 880, 153, doi: 10.3847/1538-4357/ab2a02

Yang, J., Wang, F., Fan, X., et al. 2019b, The Astrophysical Journal, 871, 199, doi: 10.3847/1538-4357/aaf858

Yang, J., Wang, F., Fan, X., et al. 2020a, The Astrophysical Journal Letters, 897, L14, doi: 10.3847/2041-8213/ab9c26

Yang, J., Wang, F., Fan, X., et al. 2020b, The Astrophysical Journal, 904, 26, doi: 10.3847/1538-4357/abbc1b

Zhu, Y., Becker, G. D., Bosman, S. E. I., et al. 2021, The Astrophysical Journal, 923, 223, doi: 10.3847/1538-4357/ac26c2

Zhu, Y., Becker, G. D., Bosman, S. E. I., et al. 2022, The Astrophysical Journal, 932, 76, doi: 10.3847/1538-4357/ac6e60

Zhu, Y., Becker, G. D., Christenson, H. M., et al. 2023, The Astrophysical Journal, 955, 115, doi: 10.3847/1538-4357/aceef4

Appendix A

Numerical Convergence

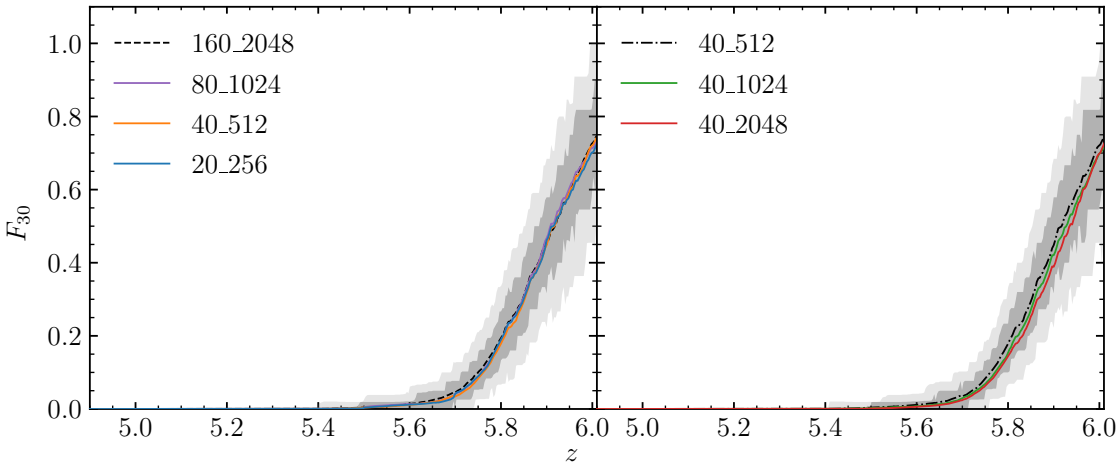


Figure A.1: Fractions of QSO spectra exhibiting long ($L \geq 30\ h^{-1}\,\mathrm{Mpc}$) dark gaps as a function of redshift with different simulation configurations. The left panel compares results for varying box sizes but fixed mass resolution. The dark gray and light gray shaded regions are the 68% and 95% limits on the expected scatter for the present sample size from the `160_2048` simulation, which are the same as the shaded regions in Figure 2.12. The right panel compares results for varying mass resolutions but fixed box size. The dash dotted line is the mean, and shaded regions are 68% and 95% limits of the prediction from the `40_512` simulation.

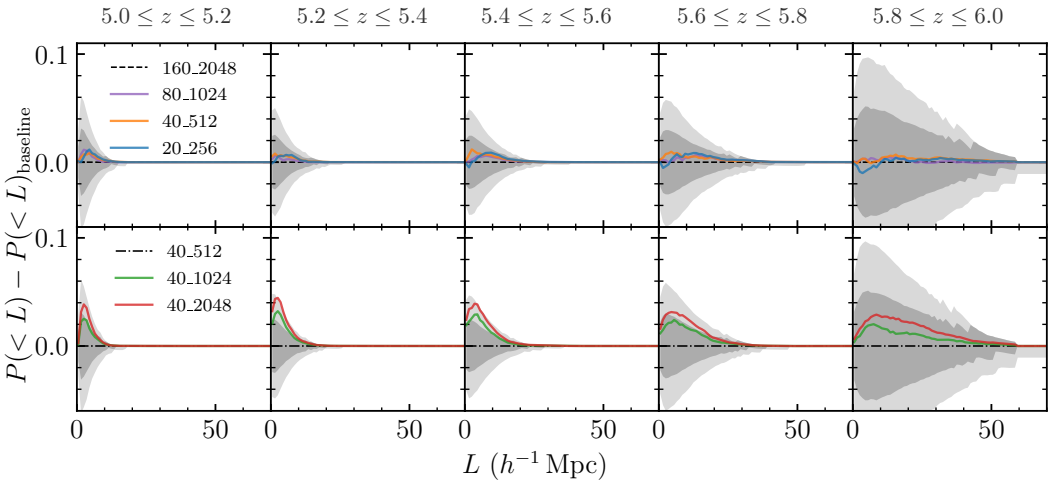


Figure A.2: Difference on dark gap length distributions for different box sizes and mass resolutions compared to the baseline configurations. The top and bottom rows compare results for varying box sizes and varying mass resolutions, respectively. The dark gray and light gray shaded regions are the 68% and 95% limits on the expected scatter for the present sample size from the `160_2048` simulation (Top panel) and from the `40_512` simulation (Bottom panel).

Here we test the convergence of our results for the `homogeneous-UVB` model with different box sizes and mass resolutions. We denote different simulation configurations as `X_Y`, where X is the box size in comoving Mpc h^{-1} and Y = [(number of dark matter + baryon particles)/2]$^{1/3}$. The fiducial configuration used in Section 2.4 is `160_2048`.

To calculate F_{30} with smaller boxes, we firstly stitch the short skewers to form 160 Mpc/h skewers, and then create mock spectra following the method described in Section 2.5.4. In Figure A.1, we compare F_{30} for mock data generated from different simulation configurations. We find little dependence on box size. F_{30} decreases slightly with increasing mass resolution, though the differences are within the expected 68% scatter for the present sample size. We compare $P(< L)$ with different configurations to the baseline configurations by calculating $P(< L) - P(< L)_{\rm baseline}$ in Figure A.2. For a fixed mass resolution and varying box size, we compare simulations to our fiducial `160_2048` simulation. For a fixed box size but varying mass resolution, we compare simulations to the `40_512` configuration, which has the same mass resolution of `160_2048`. Similar to F_{30}, the variations in $P(< L)$ with box size are relatively minor. The impact of mass resolution is more significant, especially for smaller L.

Figures A.1 and A.2 suggest that mass resolution has a larger impact than box size on our statistics, in the sense that simulations with lower mass resolutions tend to produce more long gaps and fewer short gaps. This is because weak, narrow Lyα transmission peaks tend to be suppressed at lower resolution. This effect may need to be considered for larger samples. We emphasize that the `homogeneous-UVB` models that we are using for these convergence tests contain significantly fewer long gaps than the late reionization and

fluctuating UVB models. It is therefore unclear how well the trends we see for large L would apply to these models, although it is likely that the effects of mass resolution we see at smaller gap lengths would generally be present in SPH simulations.

Appendix B

Simulation Predictions without Masking

For consistency with the observations, we mask out small wavelength regions in the mock spectra that coincide with peaks in the observed flux error arrays, as described in Section 2.4.1. Figure B.1 shows how the simulation results change without masking. The model predictions for F_{30} decrease because the masks sometimes fall on transmission peaks. The overall impact is minor; however, we emphasize that the observations should be compared to the simulation results with masking included.

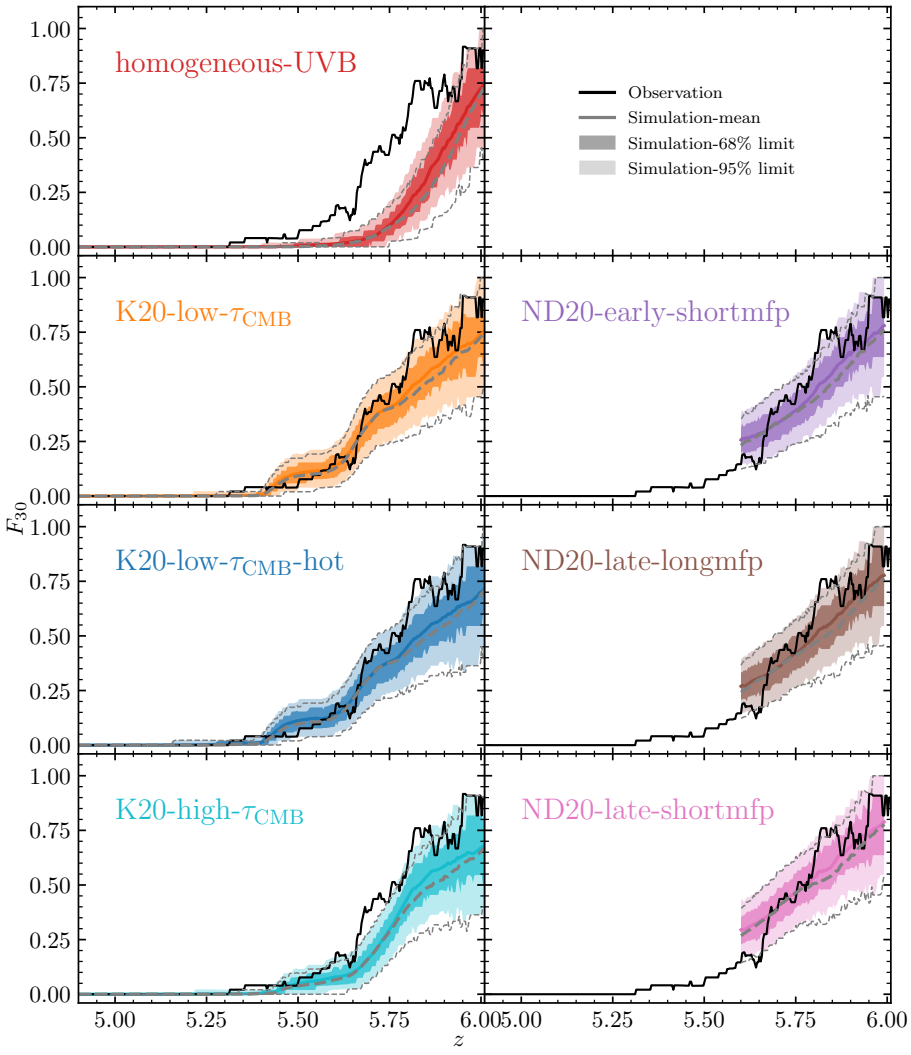


Figure B.1: The fraction of sightlines located in dark gaps with $L \geq 30\ h^{-1}\,\mathrm{Mpc}$ as a function of redshift. The dashed thick (thin) lines show the mean values (95% range) of simulation predictions without masking regions in the mock spectra that coincide with peaks in the flux error array. Other lines and shaded regions are as described in Figure 2.12.

Appendix C

F_{30} based on the power law continuum fitting

For reference, in this section, we calculate F_{30} based on the power law continuum fitting. The power-law continua are in the form of $a\lambda^{-b}$, with a and b being free parameters. We generally estimate the power-law continua over ∼1285-1350 Å in the rest frame, which is relatively free of emission lines, and we extend the fitting range out to ∼ 2000 Å when possible. Figure C.1 compares the results. The consistency (within 1σ) between F_{30} based on the PCA continuum and power law continuum suggests that our results are insensitive to continuum fitting methods.

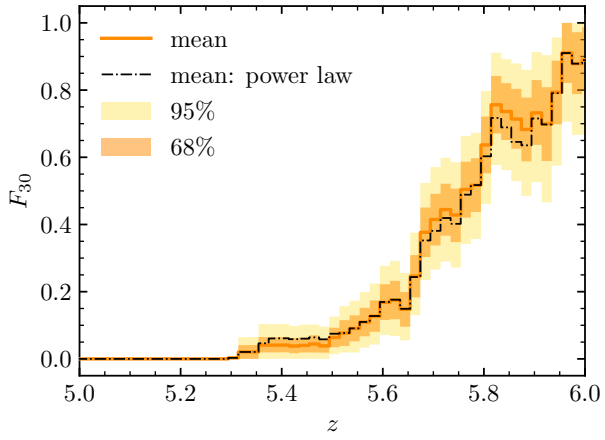


Figure C.1: Measured fraction of QSO spectra exhibiting long ($L \geq 30\ h^{-1}\,\mathrm{Mpc}$) dark gaps as a function of redshift. The dot dashed line shows the result based on the power law continuum fitting. Other lines and regions are as described in Figure 2.5.

Appendix D

Effect metal absorbers on F_{30}

The strong H I absorption typically associated with metal systems may potentially impact the observed F_{30} by connecting otherwise shorter gaps. We test whether this effect could be significant by dividing dark gaps at the redshifts of DLAs and other metal systems. We also exclude a 3000 km s^{-1} region surrounding the redshift of the metal absorber in order to allow for extended DLA absorption and/or strong absorbers clustered around the metal system. As shown in Figure D.1, the impact on F_{30} is relatively minor, with a maximum decrease of $\sim$0.1 at $z \sim 5.8$. We caution that list of metal absorbers used here may be incomplete; however, we have verified that the three long dark gaps at $z \lesssim 5.5$ in particular do not contain metals to within the sensitivity of our data. In summary, we find that the impact of metal systems on F_{30} in this regard is minor, and that the `homogenous-UVB` model is strongly ruled out regardless of how these systems are treated.

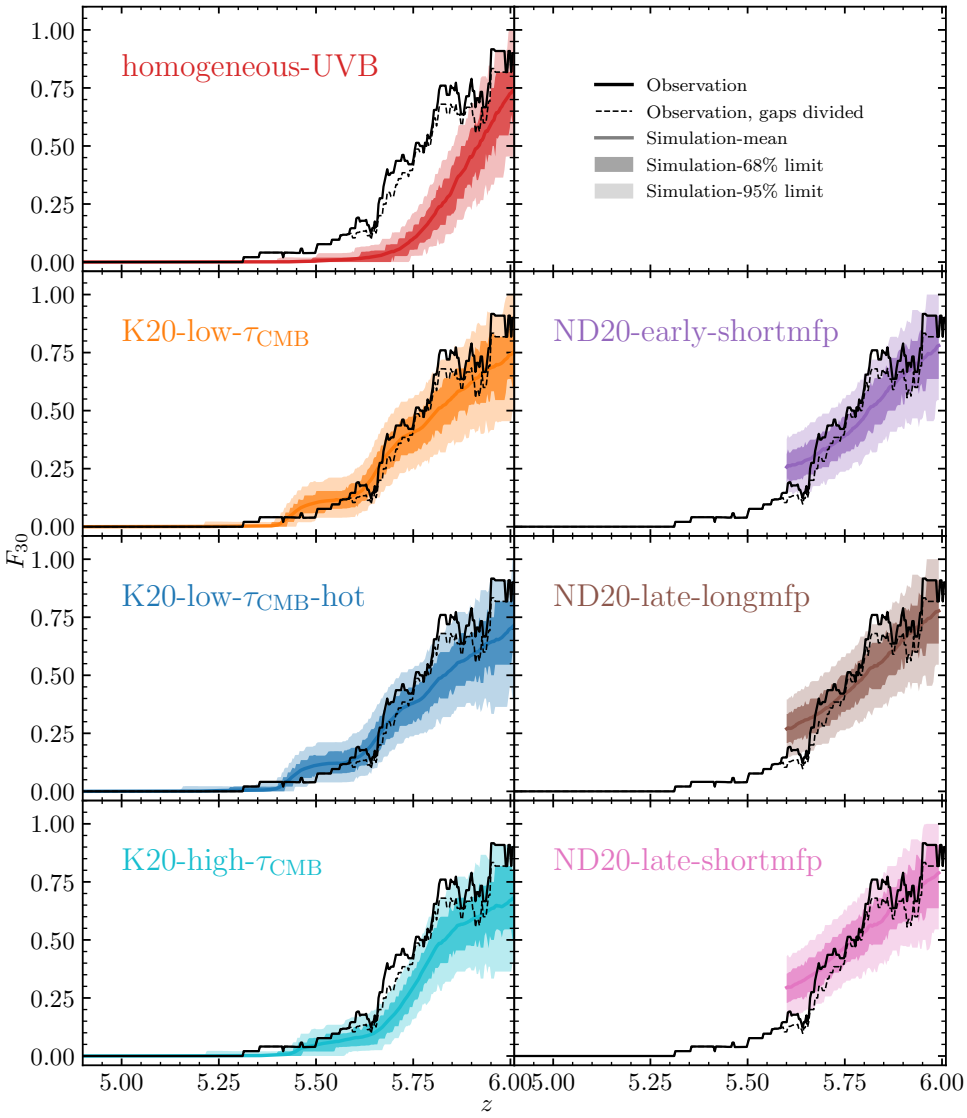


Figure D.1: The fraction of sightlines located in dark gaps with $L \geq 30\ h^{-1}\,\mathrm{Mpc}$ as a function of redshift. The dashed lines show F_{30} from observations with dark gaps divided at the redshifts of DLAs or metal systems with a 3000 $\mathrm{km\,s^{-1}}$ interval. Other lines and shaded regions are as described in Figure 2.12.

Appendix E

Relationship Between Lyβ Dark Gaps and Lyα Dark Gaps

To illustrate the effects of requiring Lyβ gaps to also be dark in the Lyα forest, here we explore the relationship between Lyβ dark gaps and Lyα dark gaps. In Figure E.1 we over-plot Lyβ-opaque regions ($\mathrm{Flux}_{\mathrm{Ly}\beta} < 0.02$ per 1 h^{-1} Mpc bin) on Lyα dark gaps as defined in Paper I. Although Lyβ-opaque regions overlap strongly with Lyα dark gaps, there do exist regions that are dark only in the Lyβ forest, e.g., the long Lyβ-opaque region toward CFHQS J1509−1749 that bridges two Lyα dark gaps, as shown in the figure. These cases are due to foregorund Lyα absorption in the Lyβ forest Requiring Lyβ gaps to also be dark in the Lyα forest partially avoids this kind of foreground contamination.

We further plot the length of Lyα dark gaps versus the length of corresponding Lyβ-opaque regions in Figure E.2. Most of long ($\geq$ 10 h^{-1} Mpc) Lyβ dark gaps appear

in $L \geq 30\ h^{-1}$ Mpc Lyα dark gaps. Only one out of 23 long Lyβ-opaque regions contains transmission in Lyα and is split into two Lyβ dark gaps.

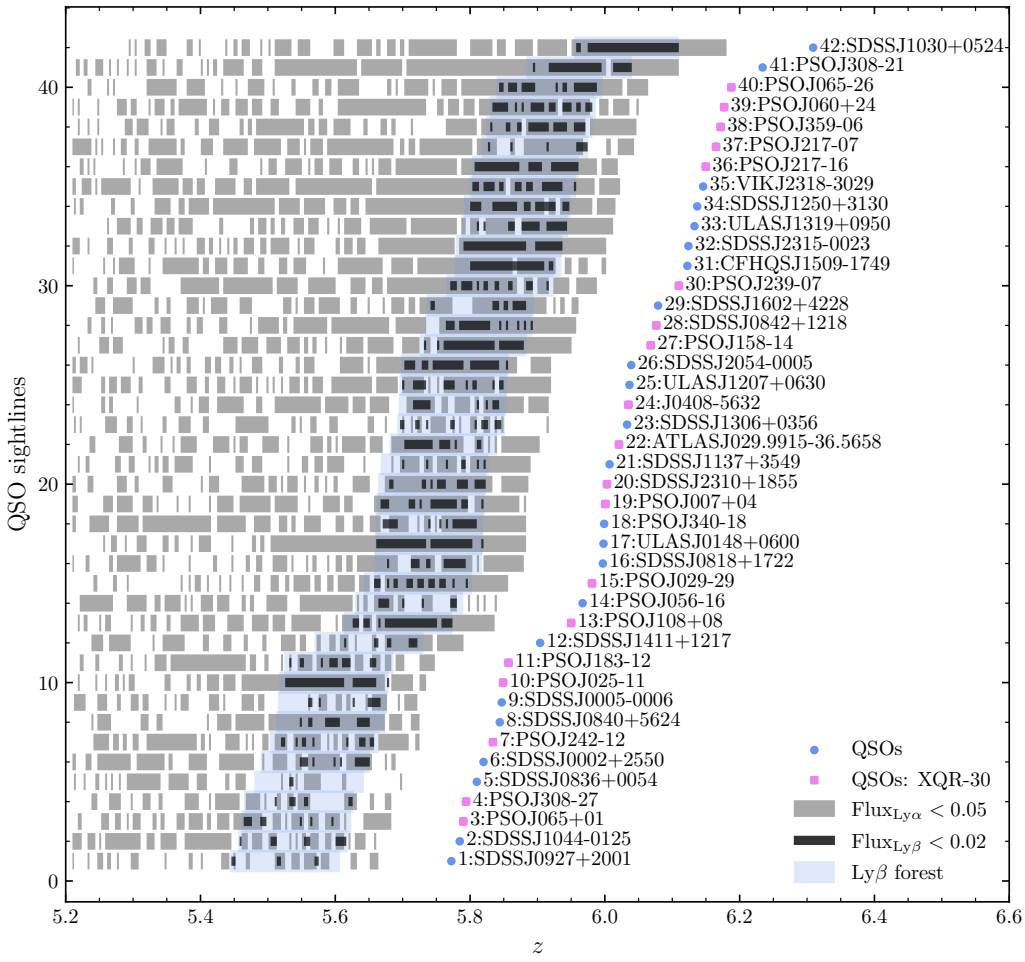


Figure E.1: Overview of Lyβ-opaque regions and Lyα dark gaps from our sample of 42 QSO lines of sight. Black bars show Lyβ-opaque regions, where normalized flux in the Lyβ forest Flux$_{\mathrm{Ly}\beta} < 0.02$ per 1 h^{-1} Mpc bin. Gray bars show Lyα dark gaps as defined in Paper I. Light blue shades highlight the redshift ranges of the Lyβ forest. The overlap between the gray bars and black bars yields Lyβ dark gaps as defined in this work.

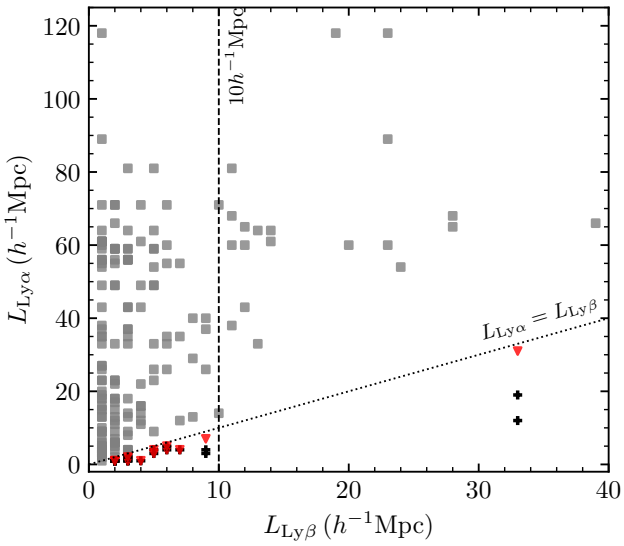


Figure E.2: Length of Lyα dark gaps versus length of Lyβ-opaque regions. For Lyβ dark gaps that are entirely within the redshift range of the Lyα dark gap, we plot the length of the Lyα gap versus the length of the Lyβ gap with a gray square. Red triangles denote situations where not all 1 h^{-1} Mpc pixels in a Lyβ-opaque region have $\mathrm{Flux}_{\mathrm{Ly}\alpha} < 0.05$ (Lyα dark gaps). The path length of Lyα dark gaps inside these Lyβ-opaque regions are marked with black crosses.

Appendix F

Metal-Enriched Systems

In Figure F.1 we display an overview of dark gaps with metal-enriched systems over-plotted for the 27 QSO sightlines in our sample where the identification of metals is relatively complete and consistent. We label metal systems with redshifts in the Lyβ forest and in the foreground Lyα forest separately. These systems are included in a metal absorber catalog that will be presented in Davies et al. in prep. Briefly, the `Python` application `Astrocook` was used to perform an automated search for Mg II, Fe II, C IV, Si IV, and N V absorbers, and DLA-like systems probed by C II and other low-ionization species. Candidate absorbers were identified using a cross-correlation algorithm within `Astrocook` that searches for redshifts where significant absorption is present in all relevant transitions. Custom filtering algorithms and visual inspection were then used to remove false positives and produce the final absorber list.

We then investigate the correlation between long ($L \geq 10\ h^{-1}$ Mpc) dark gaps and metal systems. We find that the probability for a metal system in the Lyβ forest to lie

in a long dark gap is $15 \pm 9\%$, where the 68% confidence limit comes from bootstraping these 27 sightlines 10,000 times. This probability is $31 \pm 9\%$ in the case of a system in the corresponding foreground Lyα forest. In these calculations we count clustered metal absorbers with a separation of $< 1\ h^{-1}$ Mpc as one system. By comparison, the probability that a randomly chosen point lies in a long dark gap is $22 \pm 9\%$. Our results suggest that the correlation between long dark gaps and (foreground) metal systems is not highly significant, at least for this sub-sample. The relatively lower probability of finding metal absorbers within the redshifts of long dark gaps nevertheless potentially favors the association between high IGM Lyα opacities and galaxy underdensities (see also Becker et al. 2018; Kashino et al. 2020; Christenson et al. 2021).

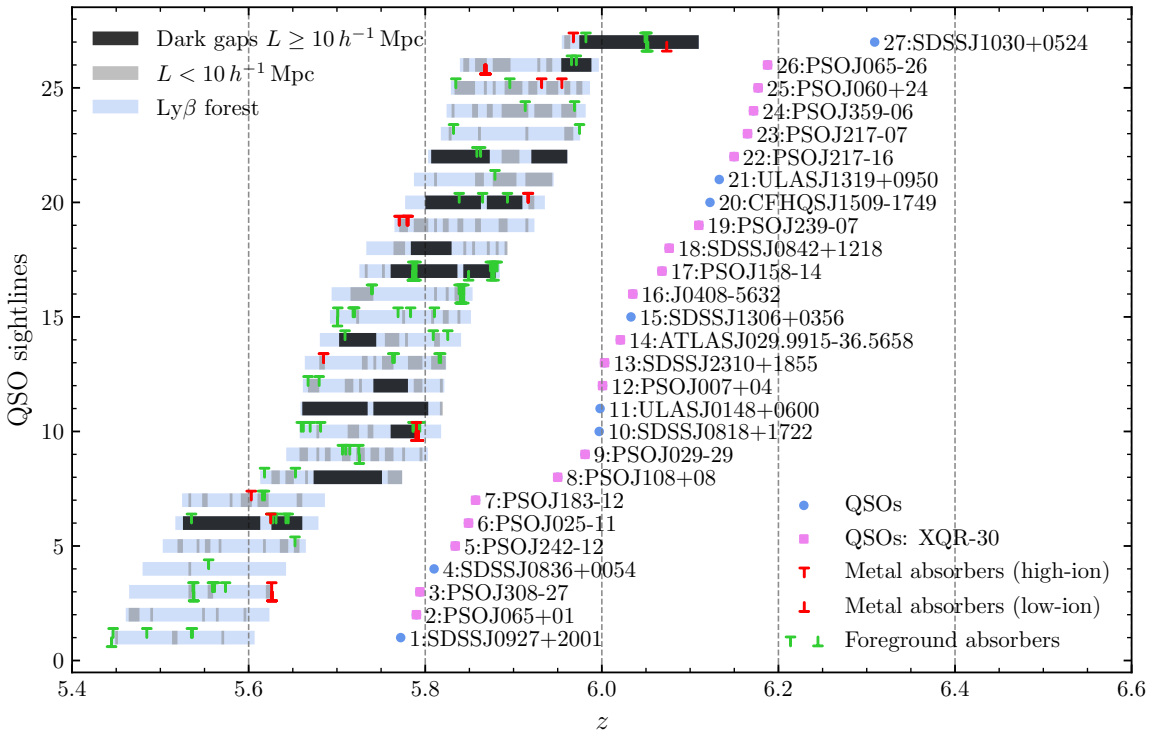


Figure F.1: Similar to Figure 2.4, but with metal-enriched absorbers over-plotted for 27 sightlines that have a relatively complete and consistent identification of metal-enriched absorbers (Davies et al., in prep). We label high-ionization and low-ionization systems within the redshift of the Lyβ forest with red "$\top$" and "$\perp$" symbols, respectively. Foreground metal-enriched absorbers whose Lyα absorption would fall within the Lyβ forest are labeled with green symbols at the corresponding Lyβ redshifts.

Appendix G

Uncertainties in the fraction of QSO spectra showing dark gaps

The evolution in F_{10} shown in Figure 3.5 shows a large drop near $z = 5.9$. To estimate the statistical fluctuations in F_{10}, we treat the "hit rate" of long dark gaps at individual redshifts as a binomial experiment defined by the number of hits (number of long dark gaps, $n_{\rm dark}$) inside a different number of trials (number of QSO sightlines, $n_{\rm qso}$). At a certain redshift, the posterior probability distribution function for the true "hit rate", x, can be expressed as a Beta distribution, $f(x;\alpha,\beta) \propto x^{\alpha-1}(1-x)^{\beta-1}$, with $\alpha = n_{\rm dark}+0.5$ and $\beta = n_{\rm qso}-n_{\rm dark}+0.5$, assuming a Jeffreys' prior. As shown in Figure G (a), the evolution of F_{10} is consistent with a monotonic increase with z within the 95% confidence intervals. We caution that the analysis here assumes that the "hit rates" at different redshifts are independent from each other.

While the dip could be due to statistical fluctuations, we nevertheless wish to check whether it may relate to possible biases in the data related to Lyβ absorption near that redshift. To check for possible systematic effects, we calculate the fraction of QSO spectra showing dark gaps of any length ($L \geq 1\ h^{-1}\,\mathrm{Mpc}$) as a function of redshift, F_{01}. As shown in Figure G, the drop in F_{10} at $z \sim 5.9$ is not present in F_{01}. Instead, the evolution of Lyβ-opaque regions with redshift appears relatively smooth. We thus find no evidence of systematic effects in the data that would suggest lower absorption overall near $z = 5.9$.

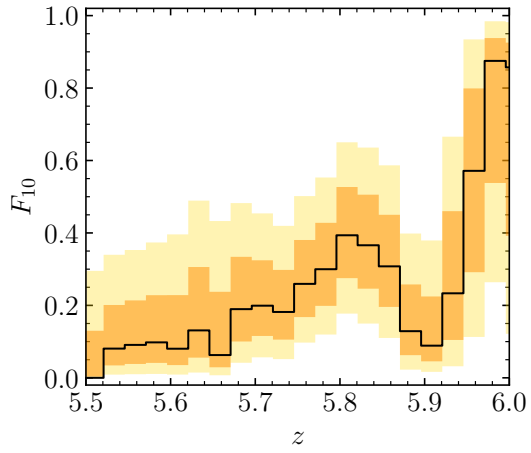


Figure G.1: Statistical uncertainty estimation for F_{10} shown in Figure 3.4 (c). Dark and light shaded regions mark the 68% and 95% limits of F_{10} based on Beta distribution.

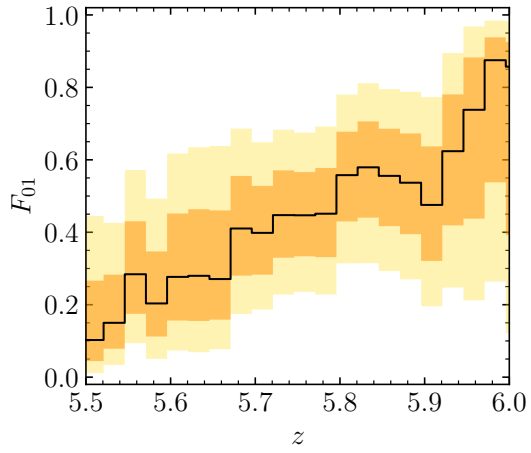


Figure G.2: Similar to our main results, but showing fraction of QSO spectra showing dark gaps with $L \geq 1\ h^{-1}\,\mathrm{Mpc}$.

Appendix H

Dark Gaps in a Lower-redshift Sample

Here we examine the extent to which strong, clustered absorbers associated with galaxies may be able to produce long dark gaps in the Lyβ forest. These (typically metal-enriched) absorbers may produce discrete absorption in either Lyβ over the redshift over the trough or Lyα at the corresponding foreground redshifts. They may also connect otherwise short dark gaps to form longer gaps. Of particular interest are very long gaps analogous to the $L = 28\ h^{-1}\,\mathrm{Mpc}$ gap toward PSO J025−11. To tests whether such gaps may be due to (circum-)galactic absorbers rather than the IGM, we search for dark gaps at $z \lesssim 5.5$ in a sample of QSO lines of sight that lie at somewhat lower redshifts than our main sample. Because the IGM becomes increasingly transparent toward lower redshifts, any long dark gaps in this sample might signal a significant contribution from discrete systems associated with galaxies.

Our lower-redshift sample includes 27 ESI and X-Shooter spectra of QSOs over $5.0 < z_{\rm em} < 5.7$ from the Keck and VLT archives. The selection of targets is based on their redshift and is independent from foreknowledge of dark gaps. QSO spectra in this lower-redshift sample have S/N greater than 20 per pixel in the Lyβ forest. In order to account for the increased mean transmission at low redshifts, we conservatively use a higher flux threshold of 0.08 when searching for dark gaps. The ratio of mean flux in the Lyβ forest at $z = 4.8$ and 5.6 is about 3.2 (e.g., Fan et al. 2006; Eilers et al. 2019; Bosman et al. 2021), thus a flux threshold of 4 times the high-redshift value is used.

Figure H.1 shows dark gaps detected in this lower-redshift sample. No dark gaps longer than 10 h^{-1} Mpc are detected. The lack of any long gaps in this sample suggests that extended gaps created largely by strong, discrete absorbers are rare, at least over $5 \lesssim z \lesssim 5.5$, which is reasonably close in redshift to our main sample. This increases our confidence that the $L = 28\ h^{-1}$ Mpc dark gap toward PSO J025−11 is likely to mainly arise from IGM absorption.

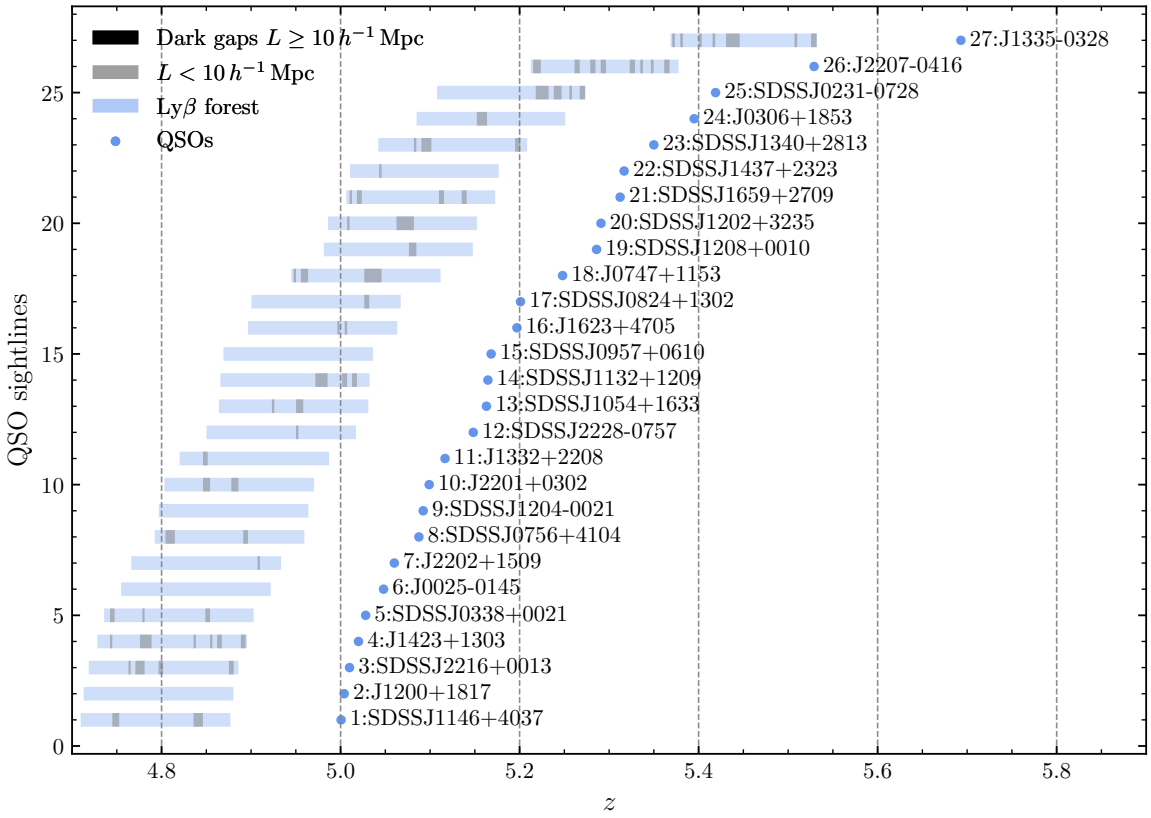


Figure H.1: Similar to Figure 2.4, but showing dark gaps identified in the Lyβ forest from a lower-redshift sample.