# Proximity Zones of Quasars in the Epoch of Reionization

A Thesis

Submitted to the

Tata Institute of Fundamental Research, Mumbai

Subject Board of Physics

for the degree of Doctor of Philosophy


by


## Sindhu Satyavolu


Department of Theoretical Physics

Tata Institute of Fundamental Research


July, 2024

Final Version Submitted in October, 2024

# Declaration

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions.

The work was done under the guidance of Professor Girish Kulkarni at the Tata Institute of Fundamental Research, Mumbai.

**Sindhu Satyavolu**

In my capacity as the formal supervisor of record of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

**Girish Kulkarni**

Date:

# Acknowledgements

# Abstract

The proximity zones of quasars with redshifts $z > 6$ are not only unique probes of the growth of their central supermassive black holes (SMBHs), but also the only parts of the Universe at these redshifts where the cosmological small-scale structure can be measured. In this thesis, we investigate SMBH growth and the last stages of reionization through new simulations and observations of quasar proximity zones around $z \sim 6$. We utilize high-dynamic range cosmological hydrodynamical radiative transfer simulations with a late reionization history consistent with several post-reionization observables. By post-processing this simulation with a one-dimensional radiative transfer model, we analyze quasar proximity zones within an inhomogeneously reionized intergalactic medium at redshift $z \sim 6$. Our results indicate that residual neutral hydrogen islands at $z \sim 6$ reduce quasar proximity zone sizes by up to 30%, with a more significant effect at higher redshifts. We also study the impact of quasar variability on proximity zone sizes using a toy model. Our comparison with observations suggests that quasars have a short duty cycle of $f_{\mathrm{duty}} \sim 0.1$ and an episodic lifetime of $t_{\mathrm{q}} \sim 1$ Myr. We further show that reconciling such variability with SMBH masses requires the black hole to continue to grow and accrete during its obscured phase. The consequent obscured fractions of $\gtrsim 0.7$ or higher are consistent with low-redshift measurements and models of black hole accretion, and will also be tested with our upcoming observations using JWST.

Further, we present 22 new measurements of proximity zones for quasars with redshifts between 5.8 and 6.6, utilizing the expanded XQR-30 sample of high-resolution, high-SNR quasar spectra. Our inferred proximity zone sizes range from 2 to 7 physical Mpc, with typical uncertainties below 0.5 physical Mpc, which, for the first time, also includes uncertainty in the quasar continuum. With this, we increase the total number of proximity zone measurements in the late stages of cosmic reionization to 87.

Finally, we use the above simulations of quasar proximity zones to model their Lyman

continuum (LyC) spectra, using which the mean free path (MFP) of hydrogen ionizing photons has been directly measured. We critically assess the analytic models used in previous measurements of the MFP for potential biases from quasar environments, their lifetimes and incomplete reionization around quasar proximity zones. We show that the bias from overdensities and patchiness in reionization around quasars is $< 25\%$, and the bias from quasar lifetimes is $< 10\%$ at $z \sim 6$. Further, we develop a new technique to directly measure the MFP by fitting our model stacks to the observed LyC stack. Our mean free path measurement is $\lambda_{\mathrm{mfp}} = 1.49^{+0.47}_{-0.52}$ pMpc at $z = 6$, reconciling both direct and indirect measurements in the literature.

Overall, this thesis provides insights into the interplay between quasars and the evolving IGM during the epoch of reionization, improving our understanding of SMBH growth and the reionization process.

# List of Publications

## Publications relevant to the thesis

1. *"The need for obscured supermassive black hole growth to explain quasar proximity zones in the epoch of reionization"*, **Sindhu Satyavolu**, Girish Kulkarni, Laura C. Keating, and Martin G. Haehnelt, MNRAS, 521, 3108 (2023).

2. *"New quasar proximity zone size measurements at z 6 using the enlarged XQR-30 sample"*, **Sindhu Satyavolu**, Anna-Christina Eilers, Girish Kulkarni, Emma Ryan-Weber, Rebecca L. Davies, George D. Becker, Sarah E. I. Bosman, Bradley Greig, Chiara Mazzucchelli, Eduardo Bañados, Manuela Bischetti, Valentina D'Odorico, Xiaohui Fan, Emanuele Paolo Farina, Martin G. Haehnelt, Laura C. Keating, Samuel Lai, Fabian Walter, MNRAS, 522, 4918 (2023).

3. *"Robustness of direct measurements of the mean free path of ionizing photons in the epoch of reionization"*, **Sindhu Satyavolu**, Girish Kulkarni, Laura C. Keating, and Martin G. Haehnelt, arXiv:2311.06344v2 **[astro-ph.CO]** (Accepted for publication in MNRAS).

## Other publications

1. *"Chronicling the reionization history at $6 \lesssim z \lesssim 7$ with emergent quasar damping wings"*, Dominika Ďurovčíková, Anna-Christina Eilers, Huanqing Chen, **Sindhu Satyavolu**, Girish Kulkarni, Robert A. Simcoe, Laura C. Keating, Martin G. Haehnelt, Eduardo Bañados, arXiv:2401.10328v1 **[astro-ph.CO]** (Accepted for publication in ApJ).

2. *"XQR-30: The ultimate XSHOOTER quasar sample at the reionization epoch"*, Valentina

D'Odorico, E. Banados, G. D. Becker, M. Bischetti, S. E. I. Bosman, G. Cupani, R. Davies, E. P. Farina, A. Ferrara, C. Feruglio, C. Mazzucchelli, E. Ryan-Weber, J.-T. Schindler, A. Sodini, B. P. Venemans, F. Walter, H. Chen, S. Lai, Y. Zhu, F. Bian, S. Campo, S. Carniani, S. Cristiani, F. Davies, R. Decarli, A. Drake, A.-C. Eilers, X. Fan, P. Gaikwad, S. Gallerani, B. Greig, M. G. Haehnelt, J. Hennawi, L. Keating, G. Kulkarni, A. Mesinger, R. A. Meyer, M. Neeleman, M. Onoue, A. Pallottini, Y. Qin, S. Rojas-Ruiz, **S. Satyavolu**, A. Sebastian, R. Tripodi, F. Wang, M. Wolfson, J. Yang, M. V. Zanchettin, MNRAS, 523, 1399 (2023).

3. *"XQR-30: Black Hole Masses and Accretion Rates of 42 z > 6 Quasars"*, C. Mazzucchelli, M. Bischetti, V. D'Odorico, C. Feruglio, J.-T. Schindler, M. Onoue, E. Bañados, G. D. Becker, F. Bian, S. Carniani, R. Decarli, A.-C. Eilers, E. P. Farina, S. Gallerani, S. Lai, R. A. Meyer, S. Rojas-Ruiz, **S. Satyavolu**, B. P. Venemans, F. Wang, J. Yang, Y. Zhu, A&A 676, A71 (2023).

4. *"Probing Ultra-late Reionization: Direct Measurements of the Mean Free Path over 5 < z < 6"*, Yongda Zhu, George D. Becker, Holly M. Christenson, Anson D'Aloisio, Sarah E. I. Bosman, Tom Bakx, Valentina D'Odorico, Manuela Bischetti, Christopher Cain, Frederick B. Davies, Rebecca L. Davies, Anna-Christina Eilers, Xiaohui Fan, Prakash Gaikwad, Martin G. Haehnelt, Laura C. Keating, Girish Kulkarni, Samuel Lai, Hai-Xia Ma, Andrei Mesinger, Yuxiang Qin, **Sindhu Satyavolu**, Tsutomu T. Takeuchi, Hideki Umehata, Jinyi Yang, ApJ 955, 115 (2023).

5. *"Constraints on the Evolution of the Ionizing Background and Ionizing Photon Mean Free Path at the End of Reionization"*, Frederick B. Davies, Sarah E. I. Bosman, Prakash Gaikwad, Fahad Nasir, Joseph F. Hennawi, George D. Becker, Martin G. Haehnelt, Valentina D'Odorico, Manuela Bischetti, Anna-Christina Eilers, Laura C. Keating, Girish Kulkarni, Samuel Lai, Chiara Mazzucchelli, Yuxiang Qin, **Sindhu Satyavolu**, Feige Wang, Jinyi Yang, Yongda Zhu, ApJ 965, 134 (2024).

6. *"IGM damping wing constraints on the tail end of reionisation from the enlarged XQR-30 sample"*, Bradley Greig, Andrei Mesinger, Eduardo Bañados, George D. Becker, Sarah E. I. Bosman, Huanqing Chen, Frederick B. Davies, Valentina D'Odorico, Anna-Christina

Eilers, Simona Gallerani, Martin G. Haehnelt, Laura Keating, Samuel Lai, Yuxiang Qin, Emma Ryan-Weber, **Sindhu Satyavolu**, Feige Wang, Jinyi Yang, Yongda Zhu, MNRAS, 530, 3208 (2024).

7. *"Damping wing-like features in the stacked Lyα forest: Potential neutral hydrogen islands at z < 6"*, Yongda Zhu, George D. Becker, Sarah E. I. Bosman, Christopher Cain, Laura C. Keating, Fahad Nasir, Valentina D'Odorico, Eduardo Bañados, Fuyan Bian, Manuela Bischetti, James S. Bolton, Huanqing Chen, Anson D'Aloisio, Frederick B. Davies, Rebecca L. Davies, Anna-Christina Eilers, Xiaohui Fan, Prakash Gaikwad, Bradley Greig, Martin G. Haehnelt, Girish Kulkarni, Samuel Lai, Ewald Puchwein, Yuxiang Qin, Emma V. Ryan-Weber, **Sindhu Satyavolu**, Benedetta Spina, Fabian Walter, Feige Wang, Molly Wolfson, Jinyi Yang, arXiv:2405.12275v2 **[astro-ph.CO]** (Accepted for publication in MNRAS).

# List of Figures

# List of Tables

# Contents

# Chapter 1

# Cosmic History in the First Billion Years

The Standard Model of Cosmology, or the $\Lambda$CDM model, has been the most successful framework to explain the evolution of the Universe as we observe it (Peebles, 1993; Planck Collaboration, 2020). It postulates that the Universe is flat, homogeneous, and isotropic on large scales ($\gtrsim 100$ Mpc). The energy content of the Universe drives its expansion and is composed of dark energy (approximately 70%), cold or non-relativistic dark matter (CDM; about 25%), and baryonic matter (approximately 5%), which includes all elementary particles, atoms, molecules, and radiation (Tanabashi et al., 2018).

Figure 1.1 shows a schematic of the chronology of the Universe. After the Big Bang, the Universe underwent a brief phase of rapid accelerated expansion known as inflation, which ended with slight inhomogeneities in the distribution of matter (Ellis & Wands, 2023). Within the first few minutes, nucleosynthesis occurred, forming deuterium, helium, and trace amounts of other light nuclei, including lithium (Fields & Sarkar, 2006; Cyburt et al., 2016). It took another 380,000 years until the Universe cooled enough for electrons and protons to combine and form neutral hydrogen atoms, marking the era known as recombination. This era also saw the decoupling of photons from baryons, as the rate of Compton scattering between photons and electrons became slower than the expansion rate of the Universe. These decoupled photons are observed today as the Cosmic Microwave Background (CMB), which has a nearly perfect black body spectrum (Penzias & Wilson, 1965). Following recombination were the Dark Ages, during which time the Universe was largely devoid of luminous sources. During this period, dark matter inhomogeneities grew under gravity, and the gas of atoms, now primarily hydrogen and helium, cooled due to the expansion of the Universe. Eventually, this gas began to accumulate

**Figure 1.1:** Chronology of the Universe according to the ΛCDM model. The epoch marked as Reionization marks the period during which the intergalactic gas was ionized by the radiation from the first stars and galaxies. Also shown are some of the current (in yellow) and upcoming or recently launched (in the last five years, highlighted in gold) ground-based and space-based telescopes/interferometers placed approximately at the farthest epochs they probe. Figure adapted. Original figure credit: NAOJ.

in the overdense regions created by dark matter, leading to the formation of the first star-forming regions. The Cosmic Dawn era followed, characterized by the birth of the first luminous sources, such as stars, and later galaxies (Bromm & Yoshida, 2011; Bromm, 2013). Around the same time, supermassive black holes (SMBHs) also began to form. The Intergalactic Medium (IGM), which filled the space between galaxies, was initially composed mostly of neutral hydrogen (H I) and helium (He I). As these early luminous sources emitted energetic photons, they began to ionize the surrounding hydrogen in a process known as reionization. By the end of the first billion years, nearly all of the hydrogen in the IGM had been reionized (Fan et al., 2006b). The reionization of helium, which has a higher ionization energy than hydrogen, occurred later in two stages: the first helium reionization (where neutral helium was ionized to singly-ionized helium) likely happened concurrently with hydrogen reionization, and the second helium reionization (where singly-ionized helium was ionized to doubly-ionized helium) occurred about one billion years later, further reheating the IGM (McQuinn, 2016). The growth of structures and the formation and evolution of galaxies continued through processes such as mergers, ultimately leading to the intricate Cosmic Web observed in the current Universe.

In this concordant picture of the Universe, the cosmic dawn and reionization epoch mark a significant period that transformed the intergalactic medium (IGM) and set the stage for cosmological structure formation. However, many details about the reionization epoch remain unanswered. Questions such as the nature of the first luminous sources, their contribution to reionization, and the duration of the reionization process are still under investigation (Mesinger, 2016). Given the theoretical complexity of modeling reionization, high-redshift observations

2

are crucial to constrain theoretical models.

For a long time, quasars, being amongst the brightest astrophysical objects, provided an essential window into this era. Quasar absorption spectroscopy has been invaluable in inferring properties of the IGM, such as its ionization state, temperature at mean density, the photoionizing background it experiences, amongst other factors that inform reionization models (Becker et al., 2015a; Fan et al., 2023). Quasars in the reionization era are also noteworthy for hosting SMBHs with masses billions of times that of the Sun, prompting further questions about the mechanisms behind the early formation of these SMBHs (Inayoshi et al., 2020). In the last decade, the number of high-redshift quasars discovered rose from few tens to few hundreds (see Figure 1.2). Presently, powerful multi-wavelength efforts with telescopes such as JWST, Subaru, and DESI are expanding our view of high-redshift quasars (e.g., Matsuoka et al., 2022; Yang et al., 2023; Matthee et al., 2024). In the future, the Rubin-LSST is poised to discover around $\sim 200,000$ quasars at redshifts $z > 5$ (LSST Science Collaboration et al., 2009). As more observations become available, there is a pressing need to enhance theoretical models to fully leverage these data. The aim of this thesis is to investigate and improve theoretical modeling, while also expanding observations of the so-called quasar proximity zones in the epoch of reionization, to advance our understanding of the IGM during reionization and SMBH growth.
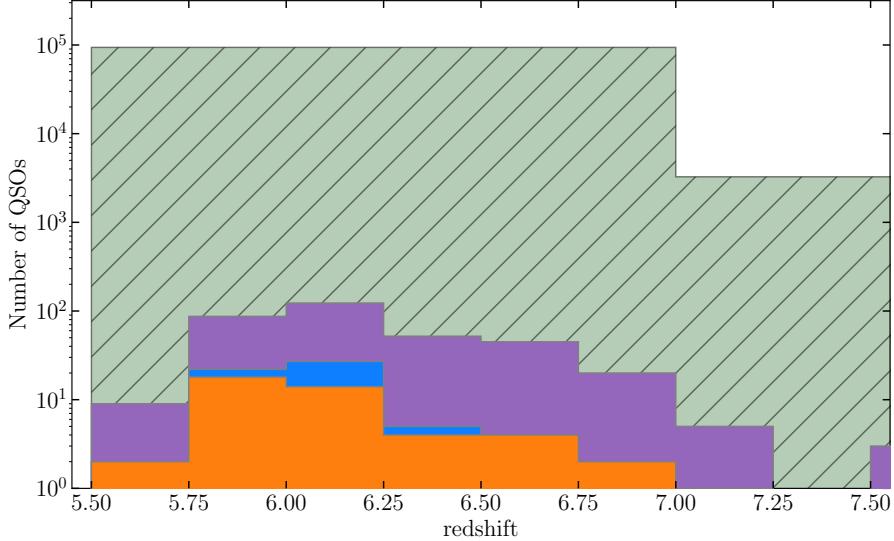
This chapter provides a brief mathematical description of the $\Lambda$CDM model, followed by an overview of the EoR and SMBH growth, which are the key topics of focus in this thesis.

## 1.1 Background Cosmology

The Friedmann-Lemaître-Robertson-Walker (FLRW) metric, describing the homogeneous and isotropic Universe is

$$\mathrm{d}s^2 = \mathrm{c}^2\mathrm{d}t^2 - a^2(t)\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2)\right), \qquad (1.1.1)$$

where $a(t)$ is the scale factor and $k$ is the curvature scalar, which is equal to 0, 1 and $-1$ for a flat, closed and open universe, respectively. The current best constraints on the curvature indicate that the Universe is flat ($k = 0$) (Masi, 2002). The Friedmann equation, which describes

**Figure 1.2:** High-$z$ QSO sample across the years. The blue, orange and purple histograms show the distribution of quasars discovered between 2000 and 2022, through large scale surveys such as SDSS, PANSTARRS1, HSC. The expected number of quasars to be discovered by the flagship large scale survey of this decade, Rubin-LSST, is shown in the green hatched region. Data from Fan et al. (2019) and Bosman (2021a).

the dynamical behavior of the scale factor, can be derived from the Einstein field equations of General Relativity. The rate of change of the scale factor depends on the total energy content of the Universe, including dark matter, baryons, radiation, and dark energy, as follows

$$\left(\frac{\dot{a}}{a}\right)^2 = H^2(t) = H^2(t_0)\left(\Omega_{\mathrm{m}}a^{-3} + \Omega_{\mathrm{r}}a^{-4} + \Omega_{\Lambda}\right), \tag{1.1.2}$$

where $H(t)$ is the Hubble parameter, and the scale factor today, $a(t_0)$, is assumed to be equal to 1 by convention. The density parameter $\Omega_i$ is defined as ratio of energy density of the $i$th species to the critical density today, where the critical density is defined as $\rho_{\mathrm{cr}} = 3H^2/8\pi G = 1.87 \times 10^{-26}\ h^2\ \mathrm{kg\,m^{-3}}$ . The Hubble constant is often expressed in terms of $h = H_0/100\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$. The matter density parameter $\Omega_{\mathrm{m}}$ is a sum of $\Omega_{\mathrm{DM}}$ and $\Omega_{\mathrm{b}}$, corresponding to dark matter and baryonic matter, respectively. The parameters $\Omega_{\mathrm{m}}, \Omega_{\mathrm{r}}$ are among the six minimum parameters that provide an agreeable fit to the $\Lambda$CDM model (the other parameters are the scalar index $n_S$, reionization optical depth, age of the Universe and amplitude of curvature fluctuations). Precise measurements of these parameters have been made possible by the Planck telescope which measured the CMB spectrum, temperature and polarisation signals (e.g., Planck Collaboration XVI, 2014).

As a result of the geometry and expansion of the Universe, the momentum of photons as they travel scales as $p \propto 1/a$. Hence, in the absence of any absorption or scattering, photons

emitted from a source $S$ at $t_e$ at wavelength $\lambda_e$ will have a wavelength $\lambda_0$ today, expressed as follows

$$\frac{\lambda_e}{\lambda_0} = \frac{a(t_e)}{a(t_0)} = \frac{1}{1+z_e}, \tag{1.1.3}$$

where $z_e$ is the cosmological redshift experienced by the photons from the source, such that $z_e = 1/a(t_e) - 1$ and $z = 0$ today. It is useful to express the scaling of physical quantities in terms of redshift. The proper or physical distance between a source and observer in expanding Universe increases with the scale factor as

$$r = ax = \frac{x}{(1+z)}, \tag{1.1.4}$$

where $x$ is the comoving distance, representing the distance between them as measured in the comoving frame. The comoving frame is the frame of reference in which the Universe appears static, despite its ongoing expansion.
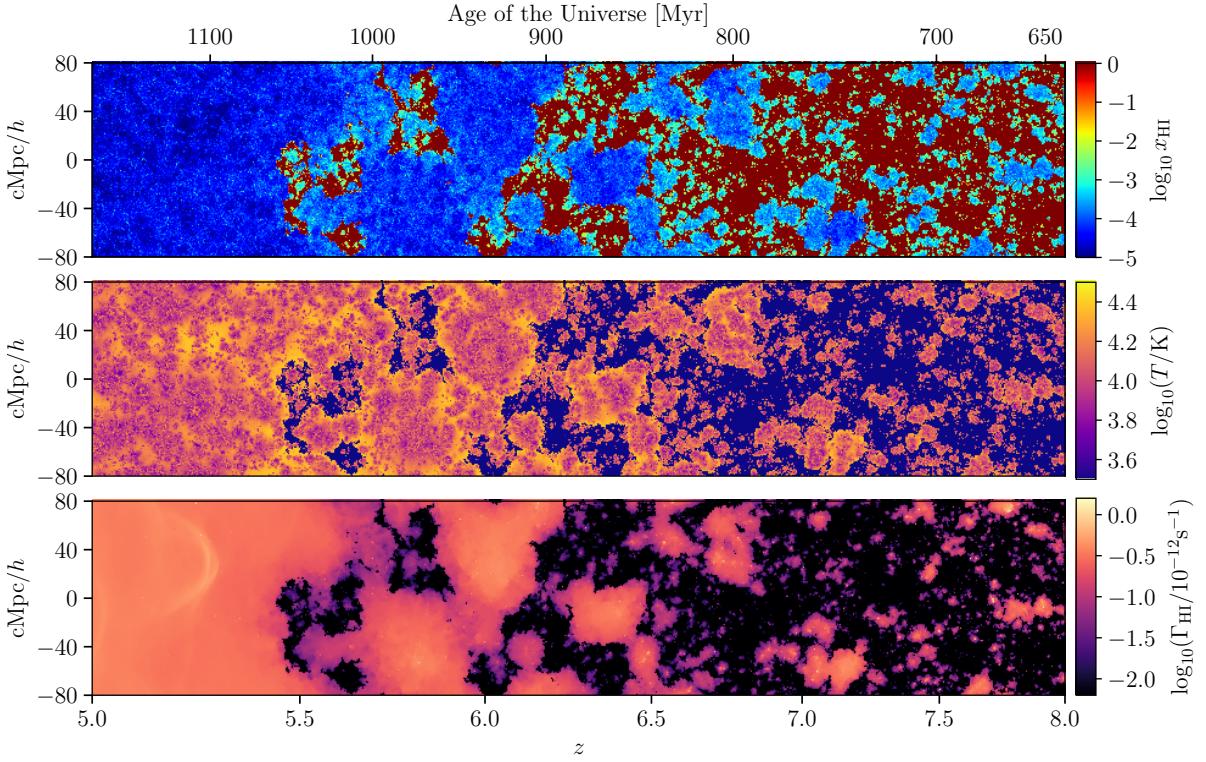
The line-of-sight comoving distance from the observer to the source, which is the distance covered by a photon traveling in the radial direction in the comoving frame, is given by

$$\mathrm{d}s^2 = 0 \implies D_c = \int \frac{c\mathrm{d}t}{a(t)} = \int_0^z \frac{c\mathrm{d}z}{H}. \tag{1.1.5}$$

Other cosmological distance measures useful from an observational point of view include luminosity distance and angular diameter distance, which can be derived from observable properties of the source, such as its apparent brightness and angular size.

## 1.2 The Epoch of Reionization

This section provides a summary of our current understanding of the fundamental questions surrounding reionization, through an investigative approach, using the 5W1H format. It is important to note that these questions are interrelated, and discussing each involves considering insights from the others.

**Figure 1.3:** The evolution of hydrogen neutral fraction (top), gas temperature (middle) and photoionization rate (bottom) between $z = 5$ and $8$ from a patchy reionization simulation (Kulkarni et al., 2019).

### *What is the Epoch of Reionization?*

The Epoch of Reionization (EoR) marks the period in the evolution of the Universe during which the intergalactic gas, predominantly in the form of Hydrogen, was ionized. In mathematical terms, let the volume averaged ionized hydrogen fraction in the IGM be denoted as $Q_{\mathrm{H\,II}}$. As reionization proceeds, $Q_{\mathrm{H\,II}}$ goes from 0 to 1. The rate of change of $Q_{\mathrm{H\,II}}$ will depend on the total number of ionizing photons being produced and emitted into the IGM, and the total number of recombinations with free electrons. The 'reionization equation' is a balance of the photoionizations (ignoring collisional ionizations) and the photorecombination processes, as follows (Madau et al., 1999)

$$\frac{\mathrm{d}Q_{\mathrm{H\,II}}}{\mathrm{d}t} = \frac{\langle \dot{n}_{\mathrm{ion}} \rangle}{\langle n_{\mathrm{H}} \rangle} - \frac{Q_{\mathrm{H\,II}}}{\bar{t}_{\mathrm{rec}}}, \qquad (1.2.1)$$

where $\langle \dot{n}_{\mathrm{ion}} \rangle$ is the volume average of the number of ionizing photons per unit physical volume of the IGM per unit time, and $\langle n_{\mathrm{H}} \rangle$ is the average cosmological hydrogen physical number density in the IGM, which evolves as $\propto (1+z)^3$. The total number of ionizing photons depends on the nature and abundance of the sources of reionization. Assuming galaxies are the dominant

6

contributors to the UV photons responsible for reionization, one can express $\langle \dot{n}_{\rm ion} \rangle$ as

$$\langle \dot{n}_{\rm ion} \rangle = \xi_{\rm ion}^{\rm SFR} \rho_{\rm SFR} f_{\rm esc}, \tag{1.2.2}$$

where $\rho_{\rm SFR}(t)$ is the cosmic star-formation rate density, $\xi_{\rm ion}^{\rm SFR}$ is the average number of ionizing photons per unit SFR, and $f_{\rm esc}$ is the fraction of the total number of the ionizing photons expected to escape into the IGM, without being absorbed within the Interstellar Medium (ISM) or the Circumgalactic Medium (CGM) of the galaxies where they originate from. Accurately estimating these astrophysical parameters is crucial for understanding reionization. They are typically determined from both simulations and observational data, as will be discussed later in this section.

The effective recombination timescale $\bar{t}_{\rm rec}$ is defined as

$$\frac{1}{\bar{t}_{\rm rec}} = \frac{\langle n_{\rm HII} n_{\rm e} \alpha_{\rm H\,II}^{\rm A} \rangle}{\langle n_{\rm HII} \rangle} = \langle \alpha_{\rm H\,II}^{\rm A} \rangle \langle n_{\rm e} \rangle C_{\rm H\,II}, \tag{1.2.3}$$

where $C_{\rm H\,II}$ is the time-dependent clumping factor,

$$C_{\rm H\,II} = \frac{\langle n_{\rm HII} n_{\rm e} \alpha_{\rm H\,II}^{\rm A} \rangle}{\langle \alpha_{\rm H\,II}^{\rm A} \rangle \langle n_{\rm e} \rangle \langle n_{\rm HII} \rangle}. \tag{1.2.4}$$

Ignoring the temperature dependence of the recombination coefficient and assuming $n_{\rm e} \propto n_{\rm HII}$, the clumping factor becomes $C_{\rm H\,II} = \langle n_{\rm HII}^2 \rangle / \langle n_{\rm HII} \rangle^2$. This parameter is dependent on the co-evolution of dark matter and baryons and is generally inferred from cosmological simulations (Kaurov & Gnedin, 2015). Following Gnedin & Madau (2022), for a quick intuition about how reionization proceeds, we can solve Equation 1.2.1 for a fixed photon production rate and redshift-dependent recombination rate. Figure 1.4 shows the evolution of the global volume averaged hydrogen ionized fraction as determined by several observations and simulations. Also shown is the solution to the reionization equation assuming reionization starts at a redshift of $z = 11$, with a constant $\langle \dot{n}_{\rm ion} \rangle / \langle n_{\rm H} \rangle = 2.7 \ {\rm Gyr}^{-1}$ and the recombination time scale as follows

$$\bar{t}_{\rm rec} = 2.3 \left( \frac{1+z}{6} \right)^{-4.35} \ {\rm Gyr}. \tag{1.2.5}$$

7

This rather simple model is consistent with several of the observations of $x_{\rm HI}$, as seen in Figure 1.4. Also shown in the Figure is a model where $\langle \dot{n}_{\rm ion} \rangle / \langle n_{\rm H} \rangle \propto (1+z)^{-0.1}$, as an example to illustrate a model that aligns with most observations.

Effectively, reionization is complete when $Q_{\rm H\,II} = 1$ and the total number of ionizations balance the total number of recombinations in the IGM. Equation 1.2.1 is not suitable for the post-reionization IGM, as it can in principle lead to $Q_{\rm H\,II}$ greater than unity, which is not physical. In the standard paradigm, reionization begins around ionizing sources, leading to ionized bubbles that later coalesce (referred to as 'inside-out' topology). Post reionization, the mean free path of ionizing photons is dependent on absorption systems such as the Lyman limit systems, which are dense non-linear collapsed structures. Madau (2017) modified Equation 1.2.1 in the presence of LLSs that allows us to study the post-overlap phase of the reionization.

While Equation 1.2.1 works well as a one-zone model for predicting the average properties of the Universe during reionization, it cannot capture the detailed aspects of the reionization process. To understand the details of reionization, it is essential to study the transport of the Lyman-continuum photons within a realistical universe evolving according to the cosmological theory of structure formation (Haardt & Madau, 1996). This necessitates the use of cosmological simulations. Reionization simulations can be broadly classified into three categories (For reviews, see Vogelsberger et al., 2020; Gnedin & Madau, 2022). All of them rely on N-body simulations coupled with hydrodynamical simulations (Springel, 2005) to establish the cosmological initial conditions of dark matter and baryons prior to reionization. The most efficient approach involves post-processing the cosmological hydrodynamical simulation with a radiative transfer (RT) simulation that solves the RT equation (Peebles, 1971) in the 3D Universe (Ciardi et al., 2003; Iliev et al., 2006a; Kulkarni et al., 2019; Cain et al., 2021). However, this method decouples radiation from gas, potentially missing out on feedback physics. A more advanced approach is to perform hybrid simulations, where radiative transfer is approximated but not fully decoupled, resulting in greater accuracy than the post-processed RT simulations (Feng et al., 2016; Bird et al., 2022; Puchwein et al., 2023). The third, most accurate method involves fully coupled cosmological hydrodynamical radiative transfer simulations to study reionization (Gnedin et al., 2017; Garaldi et al., 2022; Lewis et al., 2022). These are computationally very expensive and not feasible for multiple runs.

Although the full numerical simulations offer the most accurate depiction of the Universe,

they are challenging to adjust for a wide parameter space that can be scanned to match observations, and often require compromises between box size and resolution. This is where semi-numerical or semi-analytic simulations become useful. In typical semi-numerical models, the path of photons from generation to propagation is not directly followed. Instead, techniques such as the excursion set formalism or abundance matching are employed to determine ionized regions (Mesinger et al., 2011; Trac et al., 2022; Maity & Choudhury, 2023). Another category of semi-numerical simulations combines N-body simulations with semi-analytic models for galaxies and approximate RT (Mutch et al., 2016; Hutter et al., 2021).

Figure 1.3 shows a snaphsot of reionization simulation (post-processed RT; Kulkarni et al., 2019) which captures the reionization process from redshift $z \sim 8$ to $z \sim 5$, till the end of the first billion years. It can be seen that the reionization process is inhomogeneous, with pockets of ionized regions (in blue) expanding into the neutral regions (in red) until they merge to form a uniformly ionized medium (top panel). This is concomitant with an increase in the temperature in the ionized regions (middle panel), in the regions with large photoionization rates (bottom panel).

### *Why is it interesting?*

The EoR is interesting for several reasons. It directly follows the period of Cosmic Dawn, when the first stars and galaxies formed. These early sources emitted hydrogen-ionizing radiation, driving the reionization process, thus linking reionization closely to the formation and assembly of the first galaxies. On the other hand, reionization is a transformative event for the IGM. The reionization process not only altered the ionization state of the intergalactic medium (IGM) but also its thermal state, due to the heat associated with photoionization. The small-scale structure formation depends on both the nature of dark matter, which provides the gravitational potential well for gas to collapse, and the thermal state of the gas in the IGM, which provides thermal pressure. Understanding reionization is therefore important for constraining cosmological structure formation and dark matter properties at high-redshifts. Finally, understanding the post-reionization IGM is critical for isolating its contribution when inferring cosmological parameters from the CMB or the Ly$\alpha$ forest (Seljak et al., 2005; Planck Collaboration, 2020).
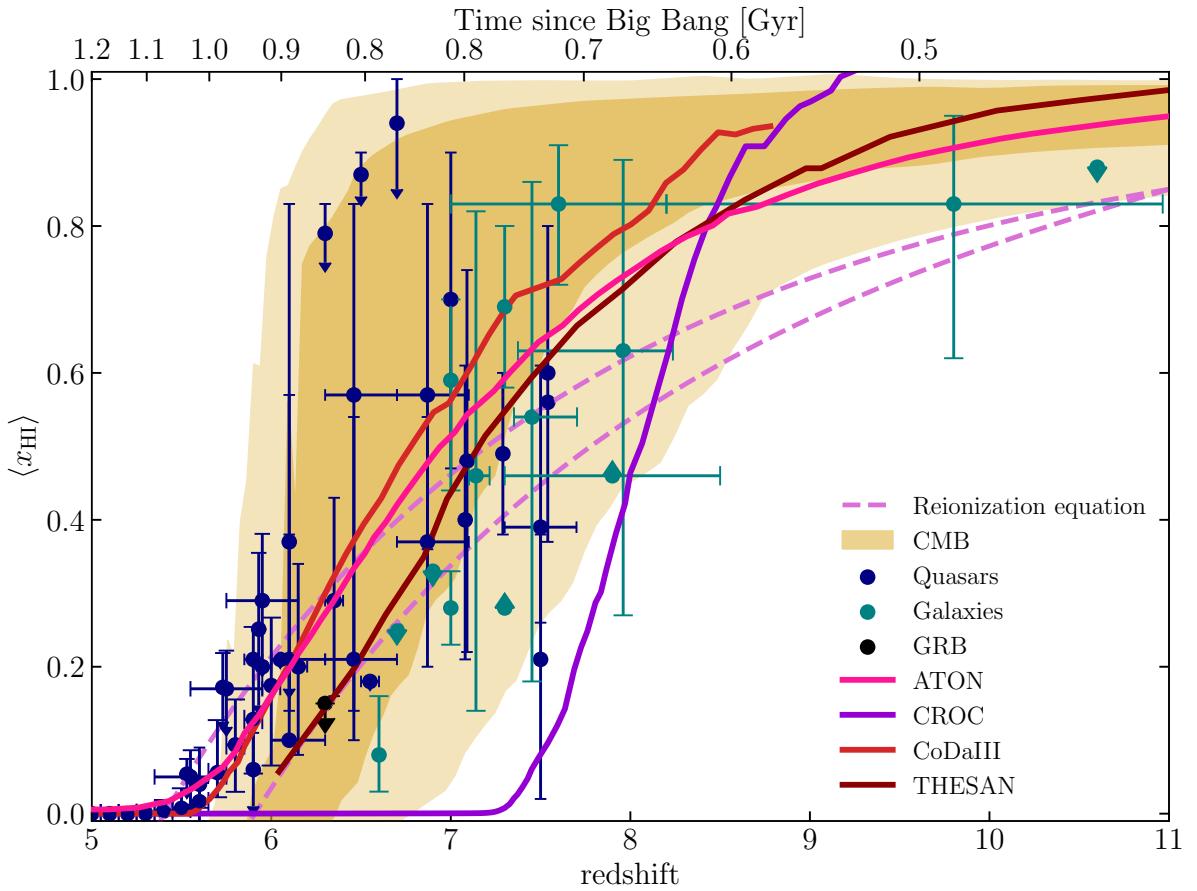
### *Where did reionization first occur? Voids or overdense regions?*

Reionization is expected to be a patchy process. In the heirarchical structure formation, dark matter halos form first and subsequently merge to grow into larger halos. Baryons are coupled to dark matter through gravity, and hence are pulled into the potential wells of dark matter halos. Once baryons are able to cool sufficiently within the halos, they begin to start forming stars via collapse of gas clouds. Eventually, mergers of dark matter halos and galaxies follow, allowing the formation of the first massive galaxies and thereafter galaxy clusters. Reionization will then proceed depending on the nature of the spectra or the ionizing output of the early stars and galaxies. Prior to reionization by the UV photons, X-rays could have partially ionized and heated the IGM to a few 100–1000 K , mainly from secondary collisional ionization with photoelectrons from photoionization (Chen & Miralda-Escudé, 2004). Given the lower photoionization cross-section ($\sigma_{H\,I} \propto \nu^{-3}$) and hence longer mean free path of X-ray photons, this ionization is expected to be more uniform. Conversely, the degree of patchiness in the ionization state of the IGM can be a probe of the spectral hardness of reionizing sources.

The UV radiation, which has a shorter mean free path, must be able to escape into the IGM to contribute to reionization. Since the densest structures are where the ionizing radiation originates, we expect them to be the first ionized regions despite the relatively large recombination rate, leading to a Swiss cheese topology with 'ionized bubbles' (see Figure 1.3). These ionized bubbles grow and merge eventually as a homogeneous UV background emerges and the IGM is homogeneously reionized. This is dubbed as 'inside out' reionization (Furlanetto et al., 2004) and is consistent with several reionization simulations (e.g., Garaldi et al., 2022). An 'outside-in' reionization is when the converse happens, where low density regions, owing to their lower optical depth and lower recombination rate are ionized earlier than the high density regions (Miralda-Escudé et al., 2000). These scenario is more likely in the post-overlap IGM that is under a uniform UV background (McQuinn, 2016).

Observational evidences for an inside-out reionization scenario in the early stages of reionization include measurements of higher fraction of Ly$\alpha$ emitters (LAEs) in brighter Lyman-break galaxies (LBGs) at $z \sim 7$ (e.g., Ono et al., 2012), measurements of ionizing bubble sizes and overdensities around $z \sim$8–10 galaxies with Ly$\alpha$ emission (e.g., Larson et al., 2022). Thanks to JWST, efforts to observe individual reionization bubbles at high redshifts above $z \sim 7$ have

accelerated through the measurements of the Lyα equivalent widths and damping wings of galaxies at these redshifts (e.g., Umeda et al., 2023; Hayes & Scarlata, 2023). Further evidence for a patchy reionization topology, associated with a fluctuating UV background, has been identified through the observed variations in the effective Lyα optical depth along multiple sightlines to high-redshift quasars (e.g., Fan et al., 2006a; Becker et al., 2015b; Bosman et al., 2022), and by cross-correlating the Lyα opacity with overdensities linked to galaxies (LAEs) (e.g., Christenson et al., 2021).



**Figure 1.4:** The evolution of the volume averaged hydrogen neutral fraction between $z = 11$ and 5. Shaded regions show the 1 $\sigma$ and 2 $\sigma$ constraints from CMB (Planck Collaboration, 2020). Data points show constraints from observations of quasars in blue (Damping wings (Greig et al., 2017b; Bañados et al., 2018; Davies et al., 2018b; Greig et al., 2019; Wang et al., 2020; Yang et al., 2020a; Greig et al., 2022; Ďurovčíková et al., 2024; Greig et al., 2024), Dark pixels (McGreer et al., 2015; Jin et al., 2023), Dark gap statistics (Zhu et al., 2021, 2022), Lyα opacity CDF (Gaikwad et al., 2023)), galaxies in green (LAEs (Mason et al., 2019; Goto et al., 2021; Morales et al., 2021; Wold et al., 2022; Bruton et al., 2023), LBGs (Bolan et al., 2022), damping wing (Umeda et al., 2023)) and gamma-ray burst damping wing in brown (Fausey et al., 2024). The solid lines show the predictions from the reionization simulation ATON in pink (Kulkarni et al., 2019), CROC in purple (Gnedin & Kaurov, 2014), CoDaIII in red (Lewis et al., 2022) and THESAN in maroon (Garaldi et al., 2022). The dashed pink curves shows the theoretical predictions based on the reionization equation for a fixed and redshift dependent ionizing photon production rate and an effective recombination timescale (see Equation 1.2.1).

### *Who or which sources/sinks were responsible?*

The nature of the sources and sinks of reionization remains incompletely understood. As previously discussed, the standard assumption is that the IGM was preheated before UV reionization occurred. However, the X-ray sources responsible for this preheating are still unconfirmed. They are largely attributed to X-ray binaries in star-forming galaxies or Active Galactic Nuclei (AGN) (e.g., Haardt & Madau, 2012; Xu et al., 2014). Directly observing these sources is incredibly challenging due to their high redshifts. On the other hand, galaxies are assumed to be the dominant contributors of UV radiation for reionization (Robertson et al., 2015). Following the formation and evolution of each galaxy together with their contribution to reionization requires resolving small scales ($O(\text{kpc})$) in large volume simulations ($O(100 \text{ Mpc})$), which is a mammoth task. For this reason, reionization simulations often rely on observations and subgrid prescriptions for modeling sources. Direct observation of the ionizing flux from galaxies responsible for reionization is only feasible if this flux is not entirely absorbed by the Lyman series, meaning it can be observed once the IGM is already nearly fully ionized. More commonly, post-reionization low-redshift galaxies are surveyed to estimate their ionizing output, which is then extrapolated to high-redshifts. The total ionizing output from galaxies is often measured as (Mesinger, 2016)

$$\dot{n}_\gamma = \rho_{\mathrm{UV}} \xi_{\mathrm{ion}} f_{\mathrm{esc}}, \tag{1.2.6}$$

where the parameters $\rho_{\mathrm{UV}}$, $\xi_{\mathrm{ion}}$ and, $f_{\mathrm{esc}}$ are independently measured. $\rho_{\mathrm{UV}}$ represents the observed UV luminosity density of galaxies, measured by integrating the UV luminosity function of galaxies within a specific magnitude range. A key limitation in measuring the $\rho_{\mathrm{UV}}$ arises from the faint-end of the luminosity function, which is often not captured in observational data. The measured UV luminosities often show a Schechter shape with a steep-rise towards the faint-end, suggesting that the fainter galaxies might indeed have a larger UV photon output and implying they might have a non-negligible contribution to reionization (e.g., Bouwens et al., 2012). The second term is the efficiency, which is the conversion factor for producing ionizing photons given the total UV output of the galaxies. This parameter is estimated for galaxies at different redshifts and at low mass and high mass ends by calculating the ratio of ionizing photon production rate (e.g., measured from H$\alpha$ luminosities or through the use of measurements of the UV continuum slope in stellar population synthesis models of the intrinsic continuum

spectrum) to the UV luminosity at 1500Å (Robertson et al., 2013; Dunlop et al., 2013; Chen et al., 2024). Typical values of $\xi_{\rm ion}$ are around $10^{25.3}$ s$^{-1}$/ergs s$^{-1}$ Hz$^{-1}$(Tang et al., 2019; Naidu et al., 2022) . The third parameter is the escape fraction $f_{\rm esc}$, which is the fraction of ionizing photons that escape into the IGM through the optically thick ISM and CGM. Constraints on $f_{\rm esc}$ have been rather difficult to obtain and are usually quoted by their upper limits, ranging upto $f_{\rm esc} \lesssim 20\%$ (e.g., Siana et al., 2007; Mostardi et al., 2015). Indirect measurements of $f_{\rm esc}$ at high redshifts $z \gtrsim 5$, based on analogues of reionization-era galaxies at lower redshifts, generally yield higher values of $f_{\rm esc} \lesssim 50\%$ (e.g., Mascia et al., 2023; Endsley et al., 2023), with some exceptions where $f_{\rm esc} \sim 2\%$ (e.g., Ma et al., 2024).

An independent way to measure the total ionizing emissivity (number of ionizing photons per unit volume per unit time, $\epsilon$) without involving the escape fractions is through its relation with the photoionization rate ($\Gamma_{\rm bg}$) and the mean free path ($\lambda_{\rm MFP}$) of ionizing photons, $\Gamma_{\rm bg} \propto \lambda_{\rm MFP}\epsilon$. Both of these parameters have been measured in an independent as well as in a dependent way using quasar spectra (Faucher-Giguère et al., 2008; Calverley et al., 2011; Becker et al., 2021; Gaikwad et al., 2023). Measurements of these parameters at high redshifts have only recently become possible and are subject to uncertainties due to parameters associated with the quasar (e.g., Worseck et al., 2014; D'Aloisio et al., 2018, see also discussion in Section 2.1 and Chapter 6). On the other hand, reionization simulations assign sources by assuming an emissivity that is able to match the observed Ly$\alpha$ flux opacity at $z \sim 5$. Simulations find such an emissivity to rise with reionization and fall post reionization (e.g., Keating et al., 2020a).

The other variable of reionization is the sinks of ionizing photons. Equation 1.2.1 accounts for absorption in sinks in an approximate way using the clumping factor to account for the decrease in flux due to recombinations. The sinks are expected to be located within the same overdense regions as the ionizing sources. Many of these sinks are believed to correspond to what we observe at low redshifts as Lyman Limit Systems (LLS). These are high column density absorption systems which are difficult to be resolved in semi-numerical and some of the radiative transfer reionization simulations as they require resolution of $\mathcal{O}$(kpc) or less (e.g., Davies & Furlanetto, 2016; Kulkarni et al., 2019). These are also challenging to detect at high redshifts because they are primarily identified in the Ly$\alpha$ forest, which is saturated in this regime, and are less likely to be identified through their metal lines. Consequently, the abundance of Lyman Limit Systems (LLSs) is extrapolated from that observed at low redshifts (Bahcall et al., 1993;

Songaila & Cowie, 2010; Prochaska et al., 2010; O'Meara et al., 2013) to higher redshifts. Additionally, modeling sinks is crucial, as they can significantly influence the mean free path of ionizing photons during and even after the overlap phase of reionization.

### *When did it happen?*

The redshifts at which the reionization process began and ended is another unknown, as it depends on the evolution of the sources and sinks. Conversely, one could constrain source properties if we can independently determine the duration and start or end of reionization. One of the earliest constraints on the end and duration of the reionization epoch is from the measurements of the temperature and polarisation anisotropies in the CMB (McQuinn, 2016). In particular, the free electrons from reionization have two contributions to the CMB signal. First, they are responsible for washing out or suppressing the existing temperature anisotropies from the era of recombination. The corresponding suppression in the power spectrum is proportional to the square of $\exp(-\tau_{re})$, where $\tau_{re}$ is the Thomson optical depth due to reionization. The Thomson scattering of electrons with CMB photons will occur on scales smaller than the horizon size at reionization, corresponding to large $l$ modes on the CMB temperature. Second, reionization also creates anisotropies in the polarisation, as scattered electrons cause E-mode polarisation at large scales ($l \lesssim 30$). Measurements of the reionization optical depth from the CMB polarisation power spectrum yield a value of $\tau_{re} = 0.054 \pm 0.007$ (Planck Collaboration, 2020), suggesting a mid-point of reionization of $z_{re} = 7.7 \pm 0.7$ (corresponding to an uncertainty $\lesssim 1\%$ of current Hubble time), and consistent with a late and fast reionization. Additionally, Compton scattering of photons with free electrons from reionization can also lead to thermal Sunyaev-Zeldovich (tSZ) and kinetic Sunyaev-Zeldovich (kSZ) effects in the leading order. The tSZ effect creates a y-type distortion in the CMB black body spectrum and is dominated by hot electrons from low redshift galaxy clusters, but studies show cross-correlation can be used to extract high redshift information from this signal in the future (Baxter et al., 2021). The kSZ effect is better suited for reionization (e.g., Iliev et al., 2007; Acharya & Majumdar, 2023), however, it is a weak signal and is much difficult to detect, making it a probable target for future observations (e.g., Abazajian et al., 2019; Ade et al., 2019).

A second probe of the end redshift of reionization comes from quasar spectroscopy (see also Section 2.1). The Gunn-Peterson effect (Gunn & Peterson, 1965) suggests that neutral regions

in the IGM should produce troughs in the Ly$\alpha$ forest observed in the spectrum of distant quasars. The disappearance of such GP troughs at redshifts around $z \sim 5.5$ suggest that reionization is complete around then (e.g., Fan et al., 2006a). Moreover, the scatter in the Ly$\alpha$ optical depth at redshifts as low as $z \lesssim 5.5$ was inconsistent with a homogeneous UV background, indicating that reionization was ongoing until $z \sim 5.3$ (Bosman et al., 2022). Such a late end is also consistent with dark gaps in the Ly$\beta$ forest (e.g., Zhu et al., 2022). Reionization also photoheats the IGM, and given the cooling timescale (dominated by adiabatic cooling due to expansion, $\propto 1/H$), the longer and earlier the reionization is, the cooler we expect the low-redshift IGM to be. The mid-point of reionization was constrained to be around $z \sim 8.5$ using the low redshift Ly$\alpha$ flux power spectrum, which is sensitive to the temperature of the IGM (Boera et al., 2019). The start of reionization is more difficult to be probed and is an open question which will be addressed in the future by 21 cm experiments (see end of this section).

### *How fast or slow was the process?*

The rapidity of the reionization process gives us a hint of the sources involved in the process. A direct probe of the progress of the reionization process is the evolution of the volume averaged neutral hydrogen fraction. Figure 1.4 shows the evolution of $\langle x_{\mathrm{HI}} \rangle$ as we currently understand it from several observations and simulations. Most observations exploit the sensitivity of Ly$\alpha$ photons to the volume-averaged neutral fraction. The luminosity functions (LFs) of Lyman-alpha emitters (LAEs) provide information about the number density of LAEs as a function of redshift. LAEs are detectable only when there is no significant absorption by neutral gas in their vicinity or in the IGM. Additionally, the clustering of LAEs is influenced by the neutral hydrogen fraction ($\langle x_{\mathrm{HI}} \rangle$); we expect to detect LAEs primarily in ionized regions, which biases clustering measurements high when the IGM is more ionized (For a review, see Stark, 2016). The spectra of high redshift galaxies and quasars exhibit a damping wing due to absorption by the neutral gas in the IGM, which has also been used to estimate the neutral gas fraction (e.g., Greig et al., 2017b, 2022; Umeda et al., 2023). Direct model-independent measurements involve measuring the dark gaps or dark pixels in the Ly$\alpha$ forest that have saturated absorption due to neutral IGM (e.g., McGreer et al., 2015; Zhu et al., 2022; Jin et al., 2023). The observations suggest the mid point of the reionization ($Q_{\mathrm{H\,I}}$ or $\langle x_{\mathrm{HI}} \rangle = 0.5$) to be around $z \sim 7$–8.

Lastly, if the reionization process was inside-out, we would expect the mean free path

(MFP) of hydrogen-ionizing photons to increase rapidly with the ionized bubble size. As the ionized bubbles merge, a discontinuity in the mean free path should occur, after which the MFP is determined by the separation between residual absorbers or Lyman limit systems (LLSs). Therefore, the evolution of the MFP during reionization is another way to gauge the rapidity of the reionization process. Direct measurements of the MFP of ionizing photons was only possible in 2021, thanks to the availability of a reasonable number of high quality quasar spectra and a new technique to account for the bias due to the quasar developed by Becker et al. (2021). This value and the follow up measurements (Zhu et al., 2023b) suggest a rapid increase in MFP from $z \sim 6$ and below and are consistent with reionization ending at $z \sim 5.3$ (see also Chapter 6).

Finally, a crucial probe to study both the Dark Ages and the reionization era is the emission and absorption associated with the 21 cm hyperfine transition in hydrogen. The cosmological 21 cm signal is expected to be weak and is therefore challenging to detect owing to contaminants or foregrounds in the radio sky, as well as due to instrumentation limitations. However, if detected, it has the potential to directly probe the dark ages and cosmological structure formation at the earliest. There are currently huge ongoing efforts to measure this signal in atleast two ways, the global 21 cm differential brightness temperature (global referring to the temperature being averaged over the whole sky at any given frequency) and the 21 cm differential brightness temperature power spectrum. The former can probe the start and end of reionization while the latter is a complementary probe of the topology of reionization (For a review, see Morales & Wyithe, 2010; Pritchard & Loeb, 2012). An alternative probe of the IGM small scale structure and ionization state is to measure the 21 cm absorption along the lines of sight to bright radio sources such as quasars, to study the 21 cm forest (analogous to the Ly$\alpha$ forest) (e.g., Šoltinský et al., 2021; Furlanetto, 2006). Post reionization, 21 cm intensity mapping experiments are an important alternative probe of the distribution of galaxies and hence the matter power spectrum.

## 1.3   Supermassive Black Hole Growth

Almost all massive galaxies around us host a Supermassive Black hole (SMBH) that is about a million to billion times more massive than the Sun. In the early 2000s, it was discovered that billion-solar-mass SMBHs existed within just a billion years after the Big Bang (Fan et al., 2000). Since then, several hundreds of these billion-solar-mass objects have been detected (Fan et al.,

2006a; Bañados et al., 2016; Matsuoka et al., 2016, 2018b; Mortlock et al., 2011; Matsuoka et al., 2018a, 2019; Yang et al., 2023), with some reaching masses of up to $\sim 10^{10}$ M$_\odot$ at $z \sim 6$ (e.g., Wu et al., 2015). The formation and growth of these early SMBHs have remained open questions in cosmology and astrophysics (For a review, see Inayoshi et al., 2020). Furthermore, local SMBHs show strong correlation with their host galaxy properties such as its stellar mass and stellar velocity dispersion, indicating co-evolution of SMBH and host galaxies (Kormendy & Richstone, 1995; Ferrarese & Merritt, 2000; Wu et al., 2002; Ferrarese et al., 2006; Novak et al., 2006; Beifiori et al., 2012; Shankar et al., 2016). Thanks to JWST, the evolution of these relationships with redshift has also been a topic of significant interest lately, with evidence suggesting that some high-redshift SMBHs may be over-massive relative to their host galaxies (e.g., Yue et al., 2024; Stone et al., 2024).

Astrophysical black holes that form from stellar collapse are likely to be only as massive as their progenitors. Given that the most massive stars hypothesized to exist are of primordial origin and can range up to a few hundred solar masses, the formation of black holes heavier than this mass requires additional processes. In broad terms, there are two pathways through which black holes could have become supermassive. Firstly, SMBHs can grow by the accretion of matter, predominantly in the form of gas, onto a seed black hole. Dark matter, on the other hand, is generally less efficiently accreted than gas because of its collisionless nature, which doesn't provide a means for it to lose enough energy to reach the event horizon of the BH. Simulations have shown that the upper limit on the amount of dark matter that can be accreted by SMBHs is about 10% of its total mass (Peirani & de Freitas Pacheco, 2008). In the case of gas, galaxy mergers typically enable the transfer of gas closer to the black hole, which is then followed by the loss of angular momentum facilitated by electromagnetic interactions. This process ultimately allows the gas to reach the event horizon for accretion. In more massive black holes, it is possible for stars to be accreted directly without being tidally stripped, leading to gravitational radiation (Peterson, 1997). Gas collapsing under the gravity of SMBH will lose potential energy in the form of radiation, which we observe as quasars (Salpeter, 1964). Quasars are a sub class of Active Galactic Nuclei or AGN, which constitutes the system of SMBH and its accretion disk. If radiative accretion is the dominant mechanism for SMBH growth, then the Soltan argument states that the observed mass density of SMBHs can be related to the radiative energy density upto an efficiency factor (radiative efficiency). Observations of the

quasar luminosity function (QLF) and the local mass density of SMBHs validate the Soltan argument for a radiative efficiency of approximately 0.1 (e.g., Yu & Tremaine, 2002; Marconi et al., 2004; Shankar et al., 2009). Alternative methods based on fitting spectra of individual objects using the thin-disk accretion model (Shakura & Sunyaev, 1973) also suggest similar values for the radiative efficiency (e.g., Davis & Laor, 2011). This supports the notion that radiatively efficient accretion predominantly fuels the growth of SMBHs.

Secondly, SMBHs can also grow via mergers, either of several lower mass stellar-collapse black holes belonging to dense star clusters, or of possibly lower mass SMBHs within galaxies. Following mergers, the resultant black hole may either directly become an SMBH or increase its mass over time through accretion. Mergers of BHs have been observed across the electromagnetic spectrum (For a review, see De Rosa et al., 2019), as well as through gravitational wave observatories. Recently, JWST has detected potential AGN pairs at the very high redshifts (e.g., Maiolino et al., 2023). To date, the Laser Interferometric Gravitational Wave Observatory (LIGO) has detected several binary black hole mergers in the low mass range (approximately $100 M_\odot$) (Bartos et al., 2017; Abbott et al., 2020). Mergers of heavier black holes that can lead to the formation of a $10^9 M_\odot$ SMBH will produce gravitational waves in the frequency range $\sim 10^{-4}$–1 Hz. The upcoming gravitational wave observatory, LISA (Laser Interferometer Space Antenna), is capable of observing such mergers (Thorne, 1995).

We follow the discussion in Mo et al. (2010) to derive the accretion rate of the SMBH under the assumption of spherical accretion. Consider a mass distribution of gas $\rho(r)$ around a black hole of mass $M_{\rm BH}$ with a bolometric luminosity $L$. This mass distribution, if ionized, is accreted as long as the gravitational force exerted by the black hole is larger than the force exerted due to radiation pressure from the central source. The inequality between force density follows

$$F_{\rm rad} < F_{\rm grav}, \tag{1.3.1}$$

where

$$F_{\rm rad} = \frac{L}{4\pi c R^2} \sigma_{\rm T} n_{\rm e}(r), \tag{1.3.2}$$

and

$$F_{\rm grav} = \frac{G M_{\rm BH} \rho(r)}{R^2}, \tag{1.3.3}$$

where $\sigma_T$ is the Thomson scattering cross section and $n_e$ is the electron density of the gas (the force on protons is ignored as the cross-section for protons is much lower, given $\sigma_T \propto 1/m_p^2$). In a nearly ionized medium, the electron mass density can be equated to gas mass density $\rho(r)$, so that the maximum luminosity, also known as the Eddington luminosity, can be derived from 1.3.3 as

$$L < L_{\mathrm{Edd}} = \frac{4\pi G c m_p}{\sigma_T} M_{\mathrm{BH}} \approx 1.26 \times 10^{38} \left( M_{\mathrm{BH}}/M_\odot \right) \mathrm{erg\,s}^{-1}. \tag{1.3.4}$$

The radiative efficiency $\epsilon$ denotes the fraction of potential energy lost by the infalling gas that gets converted into radiation,

$$L = \epsilon \dot{M}_{\mathrm{acc}} c^2. \tag{1.3.5}$$

The maximum accretion rate, called the Eddington-limit, is therefore achieved when the luminosity is maximum and equal to the Eddington luminosity. The Eddington-limit is

$$\dot{M}_{\mathrm{Edd}} = \frac{L_{\mathrm{Edd}}}{\epsilon c^2} = 2.1 \left( \frac{0.1}{\epsilon} \right) \left( \frac{M_{\mathrm{BH}}}{10^8 M_\odot} \right) M_\odot \, \mathrm{yr}^{-1}. \tag{1.3.6}$$

The rest of the rest-mass energy of the accreting gas that is not radiated is then utilized in increasing the black hole mass as

$$\dot{M}_{\mathrm{BH}} c^2 = (1 - \epsilon) \dot{M}_{\mathrm{acc}} c^2. \tag{1.3.7}$$

Combining Equation 1.3.5 and 1.3.7, we have

$$L = \frac{\epsilon}{1 - \epsilon} \dot{M}_{\mathrm{BH}} c^2. \tag{1.3.8}$$

More generally, one can express the luminosity of the gas in units of the Eddington luminosity, so that the black hole mass growth rate is proportional to its mass, as

$$\dot{M}_{\mathrm{BH}} \propto M_{\mathrm{BH}}. \tag{1.3.9}$$

The black hole mass at a given time/redshift is therefore given by

$$M_{\mathrm{BH}}(t) = M_{\mathrm{BH}}(t_0) \exp\left( \frac{f_{\mathrm{Edd}}(1-\epsilon)}{\epsilon t_{\mathrm{Edd}}} t_{\mathrm{q}} \right), \tag{1.3.10}$$

where $f_{\mathrm{Edd}} = L/L_{\mathrm{Edd}}$ is the Eddington ratio and $t_{\mathrm{Edd}} \approx 450$ Myr. $M_{\mathrm{BH}}(t_0)$ is the seed mass of the black hole and $t_{\mathrm{q}}$ is the quasar lifetime or the duration for which the black hole is radiatively accreting. The $e$-folding timescale, also called the Salpeter timescale, can be defined as

$$t_{\mathrm{S}} = \frac{\epsilon c \sigma_{\mathrm{T}}}{(1-\epsilon) f_{\mathrm{Edd}} 4\pi G m_{\mathrm{p}}}. \tag{1.3.11}$$

An estimate for the radiative efficiency can be derived as follows. The gravitation potential energy of the accreting material is converted into radiative energy

$$L \approx \frac{G M_{\mathrm{BH}} \dot{M}_{\mathrm{acc}}}{R} = \epsilon \dot{M}_{\mathrm{acc}} c^2, \tag{1.3.12}$$

such that the radiative efficiency is proportional to the compactness of the black hole ($M_{\mathrm{BH}}/R$)

$$\epsilon \approx \frac{G M_{\mathrm{BH}}}{R c^2}. \tag{1.3.13}$$

The Schwarschild radius of the black hole $r_{\mathrm{S}} = 2 G M_{\mathrm{BH}}/c^2$. The spectrum of the AGN is such that much of the accreting material is within $\sim 5 r_{\mathrm{S}}$ (Mo et al., 2010), so that the radiative efficiency is around $\epsilon \sim 0.1$. While the previous discussion doesn't consider the effects of magnetic fields or black hole spin, higher (lower) radiative efficiency is expected with higher (lower) black hole spins (Novikov & Thorne, 1973). Simulations of SMBH growth are necessary to study feedback and the spin-dependent evolution of radiative efficiency (e.g., Bustamante & Springel, 2019).

With the above formalism, we can now look at the pathways to form billion solar mass black holes by redshift $z \sim 6$. Figure 1.5 shows the evolution of black hole mass as a function of cosmic time or redshift for a range of seed masses and quasar lifetimes, as well as the observed distribution of all SMBHs above $z \sim 6$, as of 2021. It is important to note however, that the quasars discovered post 2021 also lie in the same region as shown in the current figure.

## Light seeds
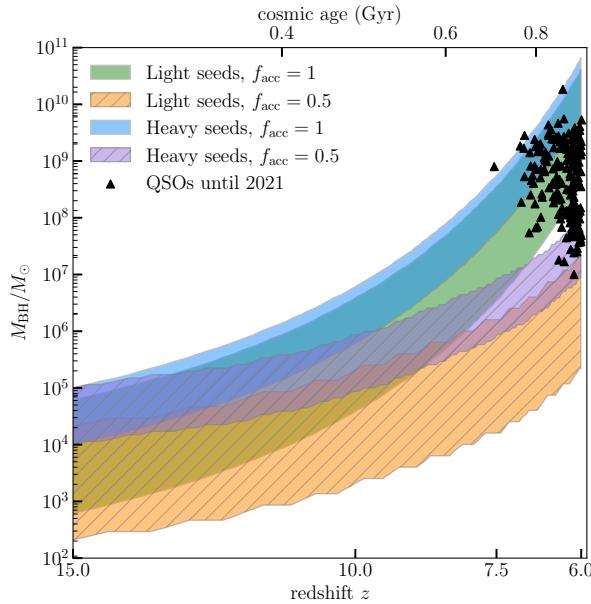
Black holes that form from the collapse of primordial PopIII stars are expected to have masses around a few $\sim 100$ $M_\odot$(e.g., Madau & Rees, 2001). Additionally, mergers of dense stellar clusters can lead to the formation of Intermediate Mass Black Holes (IMBHs) in the mass range of $10^3 - 10^4$ $M_\odot$(e.g., Omukai et al., 2008). It remains unknown when the first stars and their black holes formed. Direct detection of these black holes is challenging, requiring extreme sensitivity due to their presence at high redshifts (e.g., Inayoshi et al., 2020). Given that JWST has detected galaxies around $z \sim 14$, when the Universe was approximately 300 million years old (e.g., Helton et al., 2024), the first black hole seeds are expected to have formed around this time. The green shaded region in Figure 1.5 shows the exponential growth of black hole mass with seed masses in the range of $10^2 - 10^4 \, M_\odot$, starting at a redshift of $z \sim 20$ or cosmic age of $\sim 100$ Myr. The radiative efficiency is assumed to be 0.1. It can be seen that under the assumption that the BH is accreting continuously at the Eddington limit, i.e; $f_{\rm Edd} = 1$, it is possible to achieve SMBH masses of $\sim 10^9 \, M_\odot$ by $z \sim 6$. However, measurements of $f_{\rm Edd}$ are often much less than 1 (Mazzucchelli et al., 2023), and simulations struggle to sustain black hole accretion over extended periods due to feedback (e.g., Weinberger et al., 2018b), prompting the search for an alternative in heavier black hole seeds that require less time to grow. In the case of mergers, gravitational wave emission leads to recoil of the merged BHs into the IGM, where they cannot accrete at sufficiently high rates (e.g., Haiman, 2004).

## Heavy seeds

Direct Collapse Black holes (DCBHs) are black holes that are expected to form through the direct collapse of a molecular cloud via a supermassive star into a black hole. DCBHs are expected to have masses in the range $10^5 - 10^6 \, M_\odot$. The Jeans length sets the length scale at which thermal pressure can counteract gravitational pressure in a molecular cloud. Below the Jeans length scale, the gas cloud disintegrates and cools down, to eventually form stars. For DCBHs to form in Atomic Cooling Halos ($T_{\rm vir} > 10^4$ K), their dominant cooling channel, which is linked to the vibrational transitions of the $H_2$ molecule, must be suppressed. Several theories have been proposed to explain ways to achieve such conditions for DCBH formation, including photodissociation by Lyman-Werner photons (energies between 11.8 and 13.6 eV)

enabled by pairs of synchronized halos or dynamical heating, among other mechanisms (Yoshida et al., 2003; Visbal et al., 2014; Fernandez et al., 2014; Wise et al., 2019; Regan et al., 2020). However, these models are not without challenges, suggesting that such fine-tuned conditions might imply DCBHs are fewer in number, around $10^{-2}$ per cubic Mpc at $z \sim 20$ in the most optimistic scenario (e.g., Inayoshi & Tanaka, 2015; Habouzit et al., 2016). Recent JWST observations of $10^6$ black holes at redshifts around $z \sim 10$, as well as overmassive black holes compared to their host galaxies, have lent more credibility to heavy seed channels (e.g., Larson et al., 2023; Jeon et al., 2024; Maiolino et al., 2024).

In Figure 1.5, we consider a DCBH seed at a later redshift of $z \sim 15$, shown in blue shaded region. Such DCBHs can grow in billion solar mass black holes by redshift 6, given they are accreting at the Eddington-limit. Alternate, more exotic models for BH seeds include Primordial Black Holes (PBH), which are also invoked as a contribution to the dark matter content of the Universe (e.g., Choquette et al., 2019).



**Figure 1.5:** SMBH masses as a function of redshift. Data points represent the quasars discovered before 2021 (Data from Fan et al. 2023). Colored curves show different pathways for the growth of light and heavy black hole seeds, assuming continuous and periodic accretion.

**Episodic accretion**

From Equation 1.3.10, the parameter that decides the growth trajectory apart from the BH seed mass is the quasar lifetime. In the above scenarios, I assumed the lifetime of the quasar to be the time elapsed between the redshift of the seed black hole and observed redshift of the SMBH. In

more realistic models of black hole growth as well in observations, BH growth has found to be rather episodic (e.g., Novak et al., 2011).

The exponential growth model discussed above is the best-case scenario for maximum accretion from all directions (spherical accretion). This does not take into account the effects due to angular momentum. In reality, black holes can have spin and the gas approaching the event horizon can also have angular momentum. As such, the accretion is often expected to proceed through formation of a viscous gaseous disk around the SMBH. For a detailed study of SMBH growth, one therefore needs to employ cosmological simulations that includes dynamics and accretion of the black hole. Such simulations are difficult as they require to resolve physics ranging from cosmological length scales of few 100s of Mpc to accretion length scales of $\mathcal{O}(\mathrm{pc})$. Progress in the last decade has been made using a combination of semi-analytical modeling with zoom-in or constrained Gaussian simulations to first model the cosmological regions where accretion can occur (For a review, see Di Matteo et al., 2023). Due to the dynamic range required to resolve timescales and length scales, these simulations resort to sub-grid modeling for accretion physics. Standard approaches include Eddington-limited Bondi accretion (Bondi, 1952) ($\dot{M}_{\mathrm{BH}} \propto M_{\mathrm{BH}}^2$; $\dot{M}_{\mathrm{acc}} < \dot{M}_{\mathrm{max}} = \dot{M}_{\mathrm{edd}}$) (e.g., Weinberger et al., 2018b; Huang et al., 2020) or gravitational torque driven accretion which has a non-trivial power law dependence on the black hole mass (e.g., Anglés-Alcázar et al., 2017a; Wellons et al., 2023). These subgrid models account for different modes of AGN feedback that can potentially regulate BH growth and include mergers of BHs that allow for a statistical prediction of massive black hole binary (MBHB) merger rates.

All these simulations show evidence for episodic accretion of black holes. Figure 1.5 displays the growth curves for light and heavy black hole seed masses, assuming periodic accretion for simplicity. The accretion duty cycle, $f_{\mathrm{acc}}$, represents the ratio of the accretion period to the total time period, so that a $f_{\mathrm{acc}} = 0.5$ indicates that the black hole spends as much time accreting as it does not. This episodic accretion exacerbates the challenge of growing billion-solar-mass black holes by $z \sim 6$, as evidenced by the inability of the purple and orange curves in Figure 1.5 to reach $\sim 10^9$ M$_\odot$ by $z \sim 6$. Observational constraints on the accretion duty cycle at high redshifts are limited. The only constraints on quasar activity timescales at redshifts of 6 and above are derived from observations of high-redshift quasars, which will be discussed in the next chapter. A major aim of this thesis is to evaluate the robustness of inferences regarding quasar lifetimes

at high redshifts obtained using quasar proximity zone sizes and to link this parameter with the growth of SMBHs.

## 1.4   Open Questions and Focus of this Thesis

Our understanding of the history of the Universe is incomplete without the study of the epoch of cosmic dawn and reionization, which saw the formation and growth of the first luminous sources and the phase transition of the intergalactic gas. Ongoing efforts to study the reionization epoch rely on cosmological simulations and observations of high-redshift quasars and galaxies. The cosmological simulations are largely successful in predicting many of the observations, however, some bottlenecks in modeling remain, particularly associated with the connection between physics at large and small scales. Observations on the other hand have been pushing limits to the era of the first light, to previously unknown territories, and therefore often require several assumptions to be made for the purpose of interpretation of the data. The current paradigm is that reionization is patchy and late, with galaxies being the main drivers of the process. However, few questions such as how gradual or rapid the process is, and whether AGN contribution is negligible, remain to be understood. One of the ongoing debates related to the former question is the direct measurement of the mean free path (MFP) of hydrogen ionizing photons at $z \sim 6$. As we will discuss in Section 2.3, and in Chapter 6, these measurements hinge on the assumptions made about the so-called quasar proximity zones. Revising models of quasar proximity zones in realistic scenarios is therefore necessary for accurately measuring the MFP and understanding its implications for reionization.

A second open question linked to the early Universe is that of the formation and growth of SMBHs within the first billion years. The most accepted mechanism for SMBH growth is through accretion onto a seed black hole, while mergers also playing an important role. The pathways to the growth of SMBHs are often degenerate and require high-redshift observations to be discerned. Major breakthroughs have come recently in the form of detection of lower mass black holes at high redshift by JWST, that could potentially act as SMBH seeds. A recent puzzle concerning the SMBH growth at high-redshifts is that the inferred quasar activity timescales appear to be too short. As we will discuss in Section 2.3, almost all measurements of this parameter at $z \gtrsim 6$ are again based on quasar proximity zones. This further motivates

24
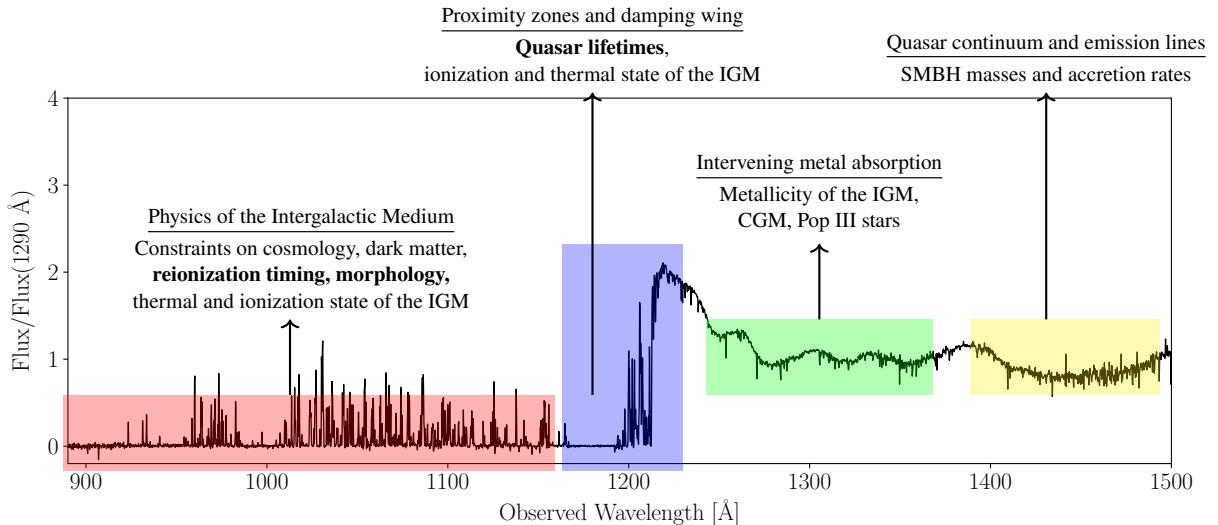
the requirement of realistic models of proximity zones to break the aforementioned degeneracy in the pathways to the formation of SMBHs.

# Chapter 2

# Quasars at High Redshifts

Ever since their discovery in the 1960s (Schmidt, 1963; Matthews & Sandage, 1963; Greenstein, 1963), quasars have fascinated us with their peculiar features: compact like a star but with a broad spectrum, containing a variety of broad and narrow emission lines, sometimes radio-loud, exhibiting huge apparent brightness, and often having larger redshifts than many galaxies discovered in those times. Sixty years later, as we discover what could be the farthest and earliest quasars, they continue to fascinate us with their peculiar features. For example, JWST has recently identified a large number of potential AGNs, referred to as 'little red dots' due to their compactness and extremely red rest-frame optical colors (e.g., Matthee et al., 2024). These AGNs exhibit redder colors compared to high-redshift [O III] emitting galaxies or low-redshift dusty galaxies. Besides their intriguing nature, quasars have enabled the study of numerous cosmological structures along the line of sight from the quasar to us (Fan et al., 2023), thanks to their brightness. From measuring cosmological parameters to determining the temperature-density relation of the IGM, quasar absorption spectroscopy has been an essential probe of cosmology across various length scales and redshifts.

Figure 2.1 summarizes the physics that can be inferred from the rest-frame UV spectrum of a high-redshift quasar as observed today. The observed spectrum is particularly interesting because it encodes information about both the quasar and the intergalactic gas through which the UV radiation has traveled. The intrinsic UV spectrum of the quasar is non-thermal and expected to follow a power-law due to processes such as synchrotron emission in the accretion disk (Peterson, 1997). The gas surrounding the SMBH is also responsible for several emission lines, including the Lyman series (Balmer series in the optical wavelength range) transitions of

**Figure 2.1:** Physics enabled by $z \sim 6$ QSO spectrum. The rest-frame UV spectrum of the QSO is shown in black. The regions shown in red, blue, green and yellow are the Lyman-series forests, the proximity zone blueward of the Ly$\alpha$ emission line, absorption forest blueward of metal emission lines and the intrinsic quasar continuum emission, respectively. The topics of interest in this thesis are highlighted in bold. Spectrum belongs to the quasar PSOJ108+08 from the XQR-30 sample (D'Odorico et al., 2023b).

hydrogen, along with high- and low-ionization emission lines from elements such as carbon, silicon, and magnesium (Osterbrock & Ferland, 2006) The absorption lines are mainly due to intervening gas, which is dominated by hydrogen (mass fraction$\sim 75\%$), with other elements such as helium, carbon and magnesium. As shown in Figure 2.1, different mechanisms are responsible for the characteristics of each observed spectral line and profile. By analyzing the spectrum through equivalent widths, profile fitting, or forward modeling, one can constrain the physics responsible for the generation and propagation of quasar radiation. For example, the continuum in the far-UV, which is relatively free from emission and absorption lines, can be used to estimate SMBH masses and accretion rates (see Section 2.2 for more details).

The strongest emission line, Ly$\alpha$ (the $2p \rightarrow 1s$ transition in hydrogen), occurs in the near-UV. Blueward of this emission line lies the quasar proximity zone, where the quasar continuum, redshifted to the resonant Ly$\alpha$ wavelength, is absorbed by the residual neutral hydrogen in the IGM surrounding the quasar. Beyond the quasar proximity zone, the UV flux is absorbed by the residual neutral hydrogen within the IGM outside the quasar's influence. The flux blueward of the Ly$\alpha$ emission (and similarly higher-order Lyman series lines up to the Lyman limit) thus constrains the amount of neutral hydrogen between the quasar and us.

This chapter provides a brief overview of the various techniques used in quasar spectroscopy to study the epoch of reionization and the growth of supermassive black holes, followed by an

introduction to quasar proximity zones, which is the main focus of this thesis.

## 2.1 Quasars as a Probe of Reionization

The Ly$\alpha$ forest is a series of absorption lines seen blueward of the rest-frame Ly$\alpha$ emission line from the quasar. It results from the absorption of photons from the quasar that are redshifted to the Ly$\alpha$ wavelength when they encounter intervening neutral hydrogen gas clouds. It was realized that the majority of the absorption in the Ly$\alpha$ forest comes from the IGM rather than individual galaxies, as the size of the gas clouds, estimated by observing sightlines of lensed quasars or quasar pairs intersecting the same cloud, are approximately ten times larger than galaxies (Mo et al., 2010). The Ly$\alpha$ optical depth is therefore a result of absorption from the cosmic neutral gas density field established through hierarchical structure formation, including some absorption from collapsed structures leading to absorption systems with high column densities.

A simple model for the Ly$\alpha$ optical depth has been derived by Gunn & Peterson (1965), which is as follows. The Ly$\alpha$ optical depth is an integral of the optical depth along the line of sight to the quasar is

$$\tau_{\text{Ly}\alpha} = \int n_{\text{HI}} \sigma_{\text{Ly}\alpha} \mathrm{d}l = \int_{z=z_q}^{z=0} n_{\text{HI}}(z) \sigma_{\text{Ly}\alpha}(\nu(1+z)) \frac{\mathrm{d}l}{\mathrm{d}z} \mathrm{d}z, \tag{2.1.1}$$

where $n_{\text{HI}}$ is the neutral hydrogen density in proper units. The proper line element $\mathrm{d}l$ can be computed using Equations 1.1.4 and 1.1.5 as $\mathrm{d}l = c/H(1+z)$, where the Hubble parameter in the matter-dominated epoch can be derived using Equation 1.1.2. The Ly$\alpha$ cross-section is evaluated only for those quasar frequencies that, after cosmological redshift, become equal to the Ly$\alpha$ frequency. The Ly$\alpha$ cross section at any frequency $\nu$ is

$$\sigma_{\text{Ly}\alpha}(\nu) = \frac{\pi e^2}{m_e c} f_{\text{Ly}\alpha} \, \phi_\nu, \tag{2.1.2}$$

where $\phi_\nu$ is the intrinsic line profile satisfying $\int \phi(\nu) \mathrm{d}\nu = 1$ and $f_{\text{Ly}\alpha}$ is the oscillator strength for the Ly$\alpha$ transition. Ignoring effects such as thermal and Doppler broadening, the line profile

can be assumed to be sharply peaked at the resonant Ly$\alpha$ frequency $\nu_\alpha$, so that

$$\tau_{\mathrm{GP}} = \frac{\pi e^2 f_{\mathrm{Ly}\alpha}}{m_e H_0 \Omega_m^{1/2}} \int_{z=z_q}^{z=0} n_{\mathrm{HI}}(z)\delta(\nu(1+z)-\nu_\alpha)\frac{\mathrm{d}z}{(1+z)^{5/2}}. \tag{2.1.3}$$

Using the properties of $\delta$-function,

$$\tau_{\mathrm{GP}} = \frac{\pi e^2 f_{\mathrm{Ly}\alpha}}{m_e H_0 \Omega_m^{1/2}\nu_\alpha}\frac{n_{\mathrm{HI}}}{(1+z)^{3/2}}. \tag{2.1.4}$$

Further, one can express the mean neutral hydrogen gas number density in terms of the mean cosmological hydrogen number density, as $n_{\mathrm{HI}} = \langle x_{\mathrm{HI}}\rangle \bar{n}_{\mathrm{H}}$, where the $\bar{n}_{\mathrm{H}} = \bar{\rho}_{\mathrm{gas}} X m_{\mathrm{H}}$. The mean gas density can be rewritten using the critical density and density parameter defined in Section 1.1. The Gunn-Peterson (GP) optical depth is therefore ([Becker et al., 2015a](#))

$$\tau_{\mathrm{GP}} \simeq 2.3\times 10^5 \langle x_{\mathrm{HI}}\rangle \left(\frac{\Omega_b h^2}{0.022}\right)\left(\frac{\Omega_m h^2}{0.142}\right)\left(\frac{X}{0.76}\right)\left(\frac{1+z}{5}\right)^{3/2}, \tag{2.1.5}$$
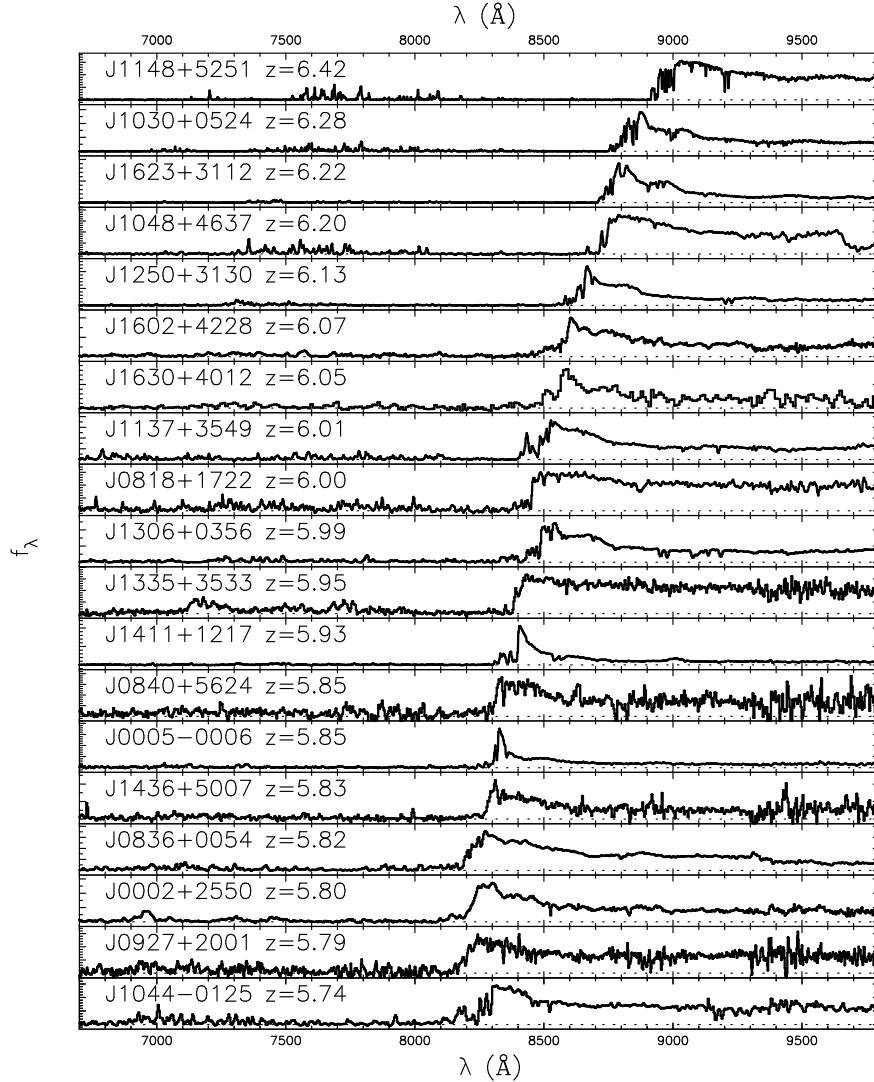
and the Ly$\alpha$ transmission is $e^{-\tau_{\mathrm{GP}}}$. [Gunn & Peterson](#) ([1965](#)) measured the optical depth measurement of the redshift $z \sim 2$ quasar 3C 9 and found, using Equation 2.1.4, that the required gas density is approximately five orders of magnitude smaller than that suggested by the $\Lambda$CDM model. This led to the conclusion that the Universe must have been ionized, with a neutral fraction of about $\sim 10^{-6}$ to explain the observed optical depth. For the post-reionization optically thin gas, one can express $\langle x_{\mathrm{HI}}\rangle$ in terms of the uniform photoionizing background by assuming ionization equilibrium, so that Equation 2.1.4 becomes

$$\tau \propto n_{\mathrm{HI}}(1+z)^{-3/2} \propto n_{\mathrm{H}}^2 T^{-0.7}(1+z)^{-3/2}/\Gamma_{\mathrm{HI}}, \tag{2.1.6}$$

where we assume that the recombination rate coefficient follows $\alpha(T) \propto T^{-0.7}$ ([Hui & Gnedin](#), [1997](#)). Further, accounting for the fluctuations in the gas density field ($\delta = \rho/\bar{\rho} - 1$) and assuming the ionized gas temperature to follow a power law relation with gas density ($T = T_0(1+\delta)^\gamma$) (e.g., [Hui & Gnedin](#), [1997](#)), we have

$$\tau \propto (1+\delta)^\beta (1+z)^{4.5}/\Gamma_{\mathrm{HI}}, \tag{2.1.7}$$

where $\beta = 2 - 0.7\gamma$. This is known as the the Fluctuating Gunn-Peterson Approximation (FGPA, e.g., Weinberg et al. 1998). The FGPA provides a powerful method for measuring the matter power spectrum from the Ly$\alpha$ optical depth in the low-redshift Universe, which can constrain cosmological parameters, the temperature-density relation, and dark matter (e.g., Iršič et al., 2017). Similarly, wavelet analysis and curvature statistics (Becker et al., 2011; Gaikwad et al., 2018; Wolfson et al., 2021) have been employed to study the temperature-density relation of the post-reionization IGM using the Ly$\alpha$ forest.



**Figure 2.2:** Spectra of 19 SDSS quasars between $5.74 < z < 6.42$ discovered by Fan et al. (2006b). The transmission in the Ly$\alpha$ forest is decreasing with increase in redshift, indicating an increasingly neutral Universe above $z \sim 6$. Figure credit: Fan et al. (2006b). ©AAS. Reproduced with permission.

From Equation 2.1.5, it is evident that an $x_{\mathrm{HI}} \sim 10^{-4}$ saturates the Ly$\alpha$ forest ($e^{-\tau_{\mathrm{GP}}} \approx 0$). Therefore, we cannot rely on post-reionization methods during the reionization epoch, where the

optical depth is high, limiting the Lyα transmission. Instead, I will discuss different techniques that are applicable in this context. A useful parameter defined in observations is the effective GP optical depth, which is

$$\tau_{\mathrm{GP}}^{\mathrm{eff}} = -\ln(\mathcal{T}); \quad \mathcal{T} = \langle e^{-\tau_{\mathrm{GP}}} \rangle, \tag{2.1.8}$$

where the average is over the redshift or wavelength range of interest along the quasar sightline.

Fan et al. (2006b) measured the GP optical depth of 19 quasars between high-redshifts of $5.7 < z < 6.4$ for the first time using SDSS, and found the effective optical depth to evolve with redshift at a rate faster than what is suggested by Equation 2.1.7. This suggested the background UV photoionization rate and thus the gas neutral fraction and the mean free path of ionizing photons were rapidly evolving at those redshifts. Additionally, they discovered a large scatter in the effective optical depth measured along the 19 sightlines (see Figure 2.2), suggesting large sample variance in the ionization state of the IGM at the end of reionization. However, Lidz et al. (2006b) argued, using simulations with a homogeneous UV background, that this scatter could be explained by fluctuations in the matter density alone. This picture struggled to hold true as the scatter in the effective optical depth measurements grew larger with newer measurements at high-redshifts, requiring the need for a spatially fluctuating UV background as well as a mean free path (e.g. Becker et al., 2015b; Eilers et al., 2018a; Bosman et al., 2022) to explain them (Davies & Furlanetto, 2016). The origin of scatter in the effective optical depth was also explained using temperature fluctuations alone (e.g., D'Aloisio et al., 2015). To break the degeneracy between photoionization-driven and temperature-driven scatter in opacities, Becker et al. (2018) measured galaxy overdensities along high- and low-opacity sightlines, and found that the low opacity sightlines were overdense, preferring the UVB fluctuation models. Kulkarni et al. (2019) and Keating et al. (2020a) showed using reionization simulations that the fluctuations of Lyα opacity and presence of long troughs as seen in Becker et al. (2018) can be explained in a scenario where reionization ends late around $z \sim 5.3$ such that 'neutral islands' in the IGM persist down to that redshift.

Equation 2.1.4 for the Lyα optical depth ignores line-broadening effects. To compare with observations, it is essential to compute the optical depth incorporating both Doppler and Natural

broadening. The cross-section can be written as

$$\sigma(\nu) = \frac{\pi e^2 f_\alpha}{m_e c} \frac{1}{\Delta \nu_D} \phi_\nu, \tag{2.1.9}$$

where the Doppler width $\nu_D = b/c\nu_\alpha$, and the Doppler parameter $b = \sqrt{2k_B T/m_H}$. The line profile is a Voigt function, which is convolution of Gaussian and Lorentzian functions:

$$\phi_\nu = \frac{a}{\pi^{3/2}} \int_{-\infty}^{\infty} \frac{\exp(-y^2)}{(x-y)^2 + a^2} dy, \tag{2.1.10}$$

where $a = \Gamma/4\pi\Delta\nu_D$ is the ratio of the decay constant to the Doppler width. The fractional shift from the line center is given by $x = (\nu - \nu_\alpha)/\Delta\nu_D$ and the parameter $y$ represents the ratio of gas velocity to the Doppler parameter, $y = v/b$. The integral is called the Hjerting function, and usually analytically approximated in simulations (e.g. Tepper-García, 2006). The total optical depth is then the line profile function rescaled by the column density of neutral hydrogen $N_{H\,I}$.

At a wavelength $\Delta\lambda$ redward from the line-center, the optical depth can be approximated as (Miralda-Escudé & Rees, 1998)

$$\tau(\Delta\lambda) \propto \tau_{GP} \left(\frac{\Delta\lambda}{\lambda}\right)^{-1}. \tag{2.1.11}$$

The absorption away from the line-center in the Lorentzian wings of the line profile, which becomes significant for large column densities $N_{H\,I}$, is referred to as the damping wing. The observed flux $\mathcal{F} = \exp(-\tau)$ will show extended wings due to the damping factor, leading to the name 'damping wings' (e.g., Miralda-Escudé, 1998) . The damping wing observed redward of the Ly$\alpha$ emission line, resulting from neutral IGM in individual quasar sightlines, has been utilized to constrain the volume-averaged neutral hydrogen fraction (e.g., Mortlock et al., 2011; Greig et al., 2019) (see also Section 1.2 and Figure 1.4). However, several challenges arise when using this phenomenon as a probe of the IGM, including uncertainties in the intrinsic spectrum of the quasar, the quasar lifetime, and sample variance due to the limited number of available sightlines (e.g., Greig et al., 2024). The redward damping wing has been detected in $z \sim 7$ quasars and, more recently, in $z \sim 10$ galaxy spectra obtained using JWST (e.g., Umeda et al., 2023). Malloy & Lidz (2015) proposed searching for neutral islands by stacking dark gaps

in the Ly$\alpha$ forest and looking for damping wing signatures at the edges of these dark regions. Notably, recent detections of red and blue damping wings attributed to neutral islands in the Ly$\alpha$ forest have been reported by Zhu et al. (2024); Spina et al. (2024), who also provide a lower limit on $\langle x_{HI} \rangle$.

At high redshifts, where the Ly$\alpha$ forest is saturated, the Ly$\beta$ forest serves as a more effective probe due to its lower bound-bound cross-section. The statistics of dark gaps in the Ly$\beta$ forest have been utilized to constrain the volume-averaged neutral fraction. Dark gaps represent absorption troughs or extended regions of absorption in the Ly$\alpha$ forest. By comparing the distribution of dark gap lengths and their frequency (referred to as dark gap statistics) with reionization simulations, one can infer the timing of reionization (e.g., Zhu et al., 2021, 2022). The absorption within these dark gaps may originate from neutral regions of the IGM, which can either be underdense areas with low UV background or overdense regions that have yet to reionize. A more model-independent approach involves considering all absorption in the dark pixels of the Ly$\alpha$ forest to arise from neutral islands of reionization, allowing for the placement of an upper limit on $\langle x_{HI} \rangle$ (e.g., Jin et al., 2023).

In the predominantly saturated Ly$\alpha$ forest, transmission manifests as 'transmission spikes', which arise from low-opacity underdense regions (Bolton & Becker, 2009). These spikes are complementary to dark gaps. The dependence of the shape parameters (height and width) and the distribution of these transmission spikes on the gas density, ionization state, temperature, and photoionization background has been explored through numerical simulations (e.g., Garaldi et al., 2019; Gaikwad et al., 2020). These spikes primarily result from overionized underdense regions, which may arise from the transverse proximity effect of galaxies or AGN (e.g., Kakiichi et al., 2018; Meyer et al., 2020). Yang et al. (2020b) utilized transmission spikes to derive a lower limit on the flux or an upper limit on the effective optical depth, assuming that all optical depth is attributable to these transmission spikes.

Finally, the quasar proximity zone is the only useful probe blueward of the Lyman series emission lines at $z \gtrsim 6$ as the Lyman-series forest is completely saturated. The ionizing radiation of the quasar is responsible for ionized overdense regions around the quasar, which translate to excess transmission immediately blueward of the rest-frame Lyman-series emission line. The proximity effect of the quasars (see Section 2.3 for detailed discussion) has been used effectively to constrain the background UV photoionization rate (e.g., Calverley et al., 2011), the volume

averaged neutral fraction $x_{HI}$ (e.g., Fan et al., 2006b), the IGM mean temperature $T_0$ (e.g., Bolton et al., 2012) and the mean free path of hydrogen ionizing photons $\lambda_{MFP}$ (e.g., Prochaska et al., 2009). The primary challenge in utilizing proximity zones to study reionization is that the quasar can outshine the ultraviolet background (UVB), along with the unique overdense environments in which they are found, given that overdense regions tend to experience reionization earlier than underdense regions (see Section 1.2).

Other constraints on reionization using quasar spectra include the measurements of the metallicity of the IGM (e.g., Oh, 2002; Simcoe, 2006). The ionizing-radiation-producing stars and galaxies also give rise to supernovae that can produce metals (elements other than H, He, most commonly studied ones include C, Si, O, Mg), which can be carried into the CGM and IGM through galactic winds. The metals are studied in terms of their column density distribution function or equivalent widths. Disappearance of high-ionization lines of C IV but presence of low-ionization counterparts such as C II suggests that the ionization state of the CGM is also transforming during the H I reionization epoch (e.g., Cooper et al., 2019; Davies et al., 2023b). The metal ionization lines also produce forests during reionization in the quasar spectra similar to the Ly$\alpha$ but will not saturate owing to their lower abundances (e.g., Hennawi et al., 2021).

The above observational constraints can be categorized into model-independent (dark pixels) and model-dependent (dark gaps, transmission spikes, proximity zones). Numerical simulations are necessary to obtain model-dependent constraints (see also discussion in Section 1.2). Such simulations should be sufficiently converged on length scales that are comparable to the data. While some reionization simulations are able to achieve high resolution (e.g., Gnedin & Kaurov, 2014), they compromise on the volume and vice versa (e.g., Kulkarni et al., 2019). Simulations of damping wings and proximity zones are additionally run with a 1D radiative transfer simulation on top of the 3D hydrodynamical cosmological radiative transfer reionization simulation (Davies et al., 2018c), or sometimes run using semi-numerical simulations to model the damping wing (Greig et al., 2022), to include the effect of quasars which are absent in the 3D simulations. While post-processing for quasars is computationally cheap and effective, it might not be suitable for timescales larger than 100 Myr (see also discussion in Section 3).

While quasar optical spectra are valuable for characterizing reionization, they are not without challenges. Issues such as the decreasing number density of quasars with redshift, selection effects, and the saturation of the Ly$\alpha$ forest make it a difficult probe at redshifts $z > 8$. Com-

bining multi-wavelength data from radio observations of the 21 cm forest and CMB secondary anisotropies could enhance our inferences in these contexts. For further discussion on the topics introduced in this section, readers may refer to the recent review by Fan et al. (2023).

## 2.2 Quasar lifetimes and SMBH masses

Single-epoch virial mass estimates of SMBHs assume that the ionization emission lines originating from the BLR (Broad Line Region) clouds are Doppler broadened due to being in virial equilibrium with the SMBH. In such a case,

$$M_{\mathrm{BH}} \propto v^2 R_{\mathrm{BLR}}, \tag{2.2.1}$$

where $v$ is the velocity dispersion estimated from the FWHM (Full Width at Half Maximum) of the emission line, and $R_{\mathrm{BLR}}$ is the size of the BLR region. The BLR size has been measured using the technique of reverberation mapping for several quasars. This technique determines the light travel time from the accretion disk to the BLR by measuring the time lags between the correlated variations of the continuum emission and broad line emission. Direct measurements of the size of the BLR are only available for a few ($\sim 50$) quasars. For others, the size is often derived based on the established scaling relation of $R_{\mathrm{BLR}}$ with the quasar bolometric luminosity $L$. However, the parameter space in which these scaling relations are established is rather narrow and might lead to extrapolation errors when applied to high-redshift quasars which are generally much brighter. The black hole mass can be derived as:

$$\frac{M_{\mathrm{BH}}}{\mathrm{M}_\odot} = 10^a \left( \frac{\mathrm{FWHM}_{\lambda_{\mathrm{rest}}}}{10^3 \ \mathrm{km\,s^{-1}}} \right)^2 \left( \frac{\lambda_{\mathrm{rest}} L_{\lambda_{\mathrm{rest}}}}{10^{44} \ \mathrm{erg\,s^{-1}}} \right)^b, \tag{2.2.2}$$

where the parameters $a, b$ are scaling factors derived using reverberation mapping. Typically, these values are taken to be $a = 6.86$, $b = 0.5$ (e.g., Vestergaard & Peterson, 2006). The most commonly used emission lines are $\mathrm{H}\alpha, \mathrm{H}\beta$, Mg II and C IV. Although Ly$\alpha$ is the strongest emission line, it is not preferred for this purpose due to IGM contamination at high-redshifts blueward of the line-center. The C IV line might display huge blueshifts compared to the systemic redshift of the quasar due to galactic winds, which can lead to an overestimate of the

SMBH masses by about ~ 5 times (e.g., Coatman et al., 2016). The typical error on the virial mass estimates is 0.5 dex, dominated by systematic errors, while the error from line fitting is much smaller comparatively (e.g., Farina et al., 2022). $L_{\lambda_{\text{rest}}}$ is the specific luminosity at the corresponding rest-frame wavelength of the emission line in focus.
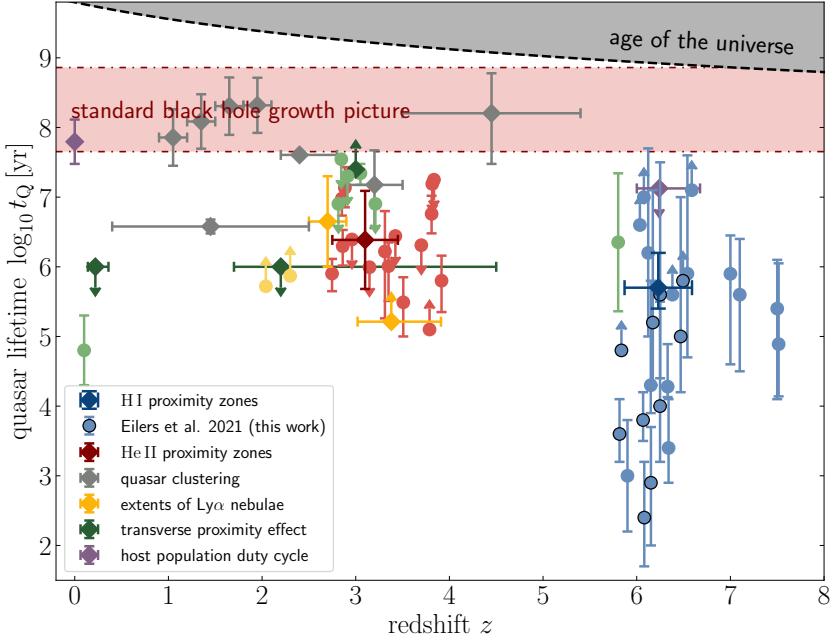
The bolometric luminosity can be derived by applying the bolometric correction to the specific luminosity (e.g., Shen et al., 2011)

$$L_{\text{bol}} = 5.15\lambda L_{\lambda,3000\,\text{Å}} \ \text{erg}\,\text{s}^{-1}. \tag{2.2.3}$$

Using the bolometric luminosity and black hole mass, the Eddington ratio is computed as $f_{\text{Edd}} = L_{\text{bol}}/L_{\text{Edd}}$, where $L_{\text{Edd}}$ is the Eddington luminosity as defined in Equation 1.3.4. Black hole masses of redshift $z \sim 6$ objects is shown in Figure 1.5. At redshift $z \sim 6$, the average Eddington ratio is about ~ 0.8, about ~ 0.3 higher than the values at low redshift (Mazzucchelli et al., 2023). Using a local Salpeter argument, radiative efficiency of SMBHs has been measured from their ionizing output (measured by spectral profile fitting) and mass, which has been found to be << 1 (e.g. Davies et al., 2019).

The other important parameter connecting SMBH masses to SMBH growth is the quasar lifetime (Equation 1.3.10). Quasar lifetimes represent the duration for which the supermassive black hole is visible to us as a quasar. Figure 2.3 shows measurements of quasar lifetimes across redshifts. These lifetimes can be inferred through both direct and indirect methods.

The indirect method relies on the measurements of quasar clustering (Martini & Weinberg, 2001). Since more massive halos are expected to be more clustered, the clustering length determines the minimum host halo mass of the quasar. Assuming all dark matter halos above this mass host a SMBH, a fraction $t_{\text{q}}/t_{\text{H}}$ of them will have quasars, where $t_{\text{H}}$ is the halo lifetime. Integrating over all possible host halo masses will lead to the quasar number density which can be compared to that measured independently from the QLF, to constrain the quasar lifetime. Typical estimates of the population average of quasar lifetimes measured from clustering range between $10^7$–$10^8$ yr (e.g., Croom et al., 2004; Shen et al., 2007; Laurent et al., 2017). DiPompeo et al. (2014) follow the same method to measure angular clustering of obscured and unobscured quasars, and report that the obscured phase of BH growth is ~ 2 times longer than the unobscured phase. As we go to higher redshifts $z \sim 6$, however, measurements of quasar clustering become

**Figure 2.3:** Compilation of measurements of quasar lifetimes and duty cycles from observations of H I proximity zones in blue (Eilers et al., 2018b, 2021; Davies et al., 2019, 2020; Andika et al., 2020; Morey et al., 2021), He II proximity zones in red (Khrykin et al., 2021; Worseck et al., 2021; Khrykin et al., 2019), quasar clustering in grey (Shen et al., 2007; Shankar et al., 2010; White et al., 2012; Laurent et al., 2017), Lyα nebulae in yellow (Cantalupo et al., 2014; Hennawi et al., 2015; Trainor & Steidel, 2013; Borisova et al., 2016), transverse proximity effect in green (Keel et al., 2012; Schmidt et al., 2018; Bosman et al., 2020; Kirkman & Tytler, 2008; Schmidt et al., 2017; Oppenheimer et al., 2018) and host population duty cycle in lavender (Yu & Tremaine, 2002; Chen & Gnedin, 2018). For each colour, lifetimes estimated as population averages are indicated as a darker shade, while individual measurements are shown as a lighter shade. The red band shows the range of lifetimes expected (1–16 Salpeter times) while assuming exponential accretion with $\epsilon = 0.1$ and $f_{\rm Edd} = 1$ (see Equation 1.3.10). The dotted black curve is the age of the Universe. Figure credit: Eilers et al. (2021). ©AAS. Reproduced with permission.

increasingly difficult as the number density of quasars decreases. Kauffmann & Haehnelt (2002) use simulations to argue that at high-redshifts, the galaxy-quasar correlation function is a better probe of the quasar lifetime, given the larger abundance of galaxies. Recent efforts to measure both types of clustering at high redshifts using JWST and Subaru (e.g., Arita et al., 2023; Eilers et al., 2024) indicate a duty cycle that is much less than 1. Given the Hubble time at $z \sim 6$, this translates to integrated lifetimes of around $10^6$ yr, shorter than even the Salpeter timescale, which is around $\sim 45$ Myr for a radiative efficiency of 0.1 and Eddington ratio of 1.

In the direct methods, the quasar activity timescales were inferred based on the feedback of the quasar activity on its surroundings (Hogan et al., 1997). Therefore, they might not always be sensitive to the integrated lifetime of the quasar but rather to the episodic lifetime, as the integrated lifetime of the quasar is irrelevant beyond the ionization equilibration timescale, after which the ionization front will not travel any farther from the quasar. These lifetimes are also often based on individual objects, unlike the population averages estimated using indirect methods. The proximity effect around quasar is the enhanced ionization caused by the quasar

on its surroundings extending out to the IGM (Cen & Haiman, 2000) (see next section for a detailed discussion). Several constraints on the quasar lifetimes have been derived from their proximity zones as observed along parallel and perpendicular directions to the quasar (e.g., Adelberger, 2004). At $z \sim 4$ and below, transverse proximity zones of the quasar in the Ly$\alpha$ forest spectrum of a background galaxy or quasar have been used to estimate quasar lifetimes to be around $\gtrsim 10^6$ yr (e.g., Worseck & Wisotzki, 2007; Gonçalves et al., 2008). The transverse proximity effect has also been probed using the He II Ly$\alpha$ forest (e.g., Schmidt et al., 2017, 2018). The longer ionization timescale of He II makes it sensitive to longer lifetimes than the H I Lyman series forests. The measured lifetimes from the line-of-sight proximity effect in the He II spectrum of quasar were found to be around $10^6$–$10^7$ yr (Khrykin et al., 2016; Worseck et al., 2021; Khrykin et al., 2019). At higher $z \sim 6$, the only constraints on quasar lifetimes were all derived using the line-of-sight H I Ly$\alpha$ proximity zone sizes of the quasars (e.g., Eilers et al., 2017; Davies et al., 2020) (see also next section for more details). These measurements find average lifetimes of around $\sim 10^6$ yr (Morey et al., 2021). A complementary probe is to look for the Ly$\alpha$ emission around quasars as a result of re-emission of their ionizing radiation. The illumination from quasars makes it possible to detect Ly$\alpha$ emission from galaxies within their ionized bubble. Bosman et al. (2020) detected 3 LAEs in the field of a $z \sim 5.7$ quasar, and argued that the quasar active phase had to be $< 10^7$ yr. Similarly, detection of scattered ionizing radiation from the quasar seen in the form of Ly$\alpha$ emission (fluorescent emission) from neutral hydrogen clouds in the IGM (Hada et al., 2024) has been used to constrain the quasar lifetimes to be $\lesssim 20$ Myr (e.g., Trainor & Steidel, 2013). Extended emission from Ly$\alpha$ nebulae or 'blobs' that were as large as few 100s of kpc was thought to originate from quasar-photoionized gas and hence used to constrain their lifetimes (e.g., Cantalupo et al., 2014). In some cases where an associated quasar was not found around the Ly$\alpha$ blobs, obscuration has been invoked to explain their detection (e.g., Steidel et al., 2000).

To summarize, while the integrated lifetime measurements are around $10^6$–$10^8$ yr, the direct measurements indicate a lifetime around $10^6$ yr, with some reporting timescales shorter than $\lesssim 10^4$ yr.

## 2.3 Proximity Zones

The Lyman series forests and the emission lines redward of the Ly$\alpha$ have been useful probes of the IGM and SMBH growth, respectively, as discussed in the previous two subsections. A unique probe that has been studied in the context of both the IGM and SMBH growth simultaneously is the quasar proximity zone. Quasar proximity zones, or near-zones, correspond to the regions in the IGM that were ionized by the quasar. For a quasar emitting $\dot{N}$ photons per unit time in an IGM that consists of only hydrogen, which is initially nearly neutral ($x_{\mathrm{HI}} \sim 1$) and uniformly distributed with density $n_{\mathrm{H}}$, the total number of photons ($N$) per unit time within a shell can be accounted for by ionizations and radiative recombinations as

$$\frac{\mathrm{d}N}{\mathrm{d}t} = 4\pi R^2 n_{\mathrm{HI}} \frac{\mathrm{d}R}{\mathrm{d}t} + \frac{4\pi}{3} R^3 n_{\mathrm{HII}}^2 \alpha. \tag{2.3.1}$$

Conversely, the rate of change in the ionized volume can be written as

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{\dot{N} - \frac{4\pi}{3} R^3 n_{\mathrm{HII}}^2 \alpha}{4\pi R^2 n_{\mathrm{HI}}}. \tag{2.3.2}$$

Equation (2.3.2) can be solved analytically assuming uniform $n_{\mathrm{HII}} = n_{\mathrm{H}}$ as

$$R_{\mathrm{ion}}(t) = R_{\mathrm{s}} \left[ 1 - \exp\left( \frac{-t}{x_{\mathrm{HI}} t_{\mathrm{rec}}} \right) \right]^{1/3}, \tag{2.3.3}$$

where

$$R_{\mathrm{s}} \equiv \left( \frac{3\dot{N}}{4\pi \alpha n_{\mathrm{H}}^2} \right)^{1/3}, \tag{2.3.4}$$

and

$$t_{\mathrm{rec}} = \frac{1}{\alpha n_{\mathrm{H}}}, \tag{2.3.5}$$
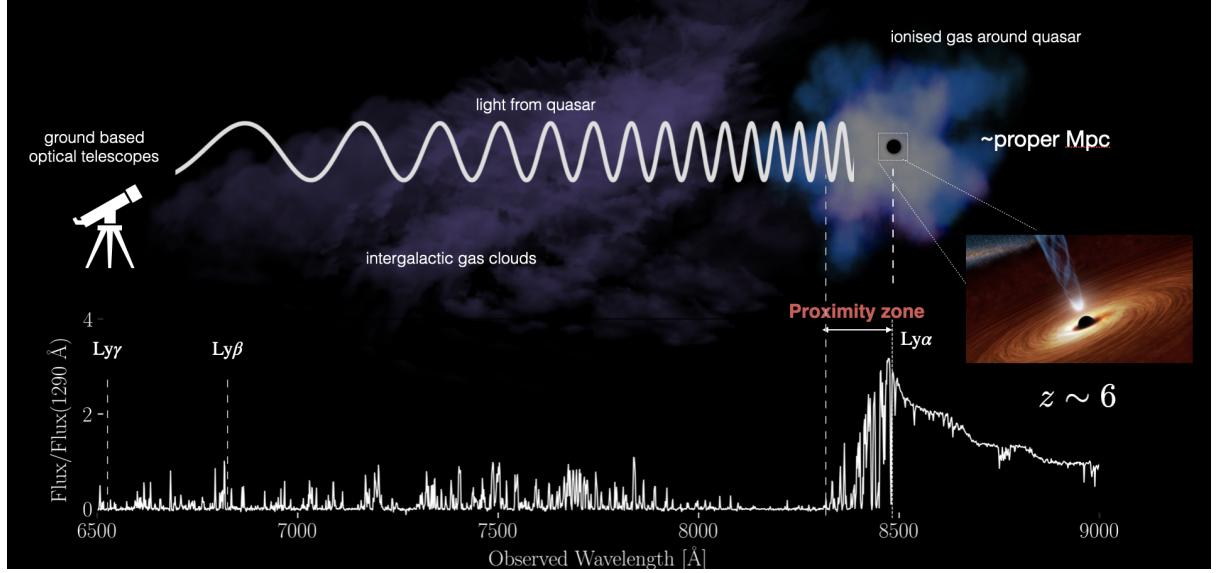
where $\alpha$ is the temperature-dependent recombination coefficient. The ionization front increases in size until the number of recombinations balance out the photoionizations, at which point, the ionization front radius saturates to the Stromgren value $R_{\mathrm{s}}$. Under the condition that $t \ll t_{\mathrm{rec}}$,

$$R_{\mathrm{ion}} \propto \left( \frac{3\dot{N} t_{\mathrm{q}}}{n_{\mathrm{HI}}} \right)^{1/3} \propto \dot{N}^{1/3} t_{\mathrm{q}}^{1/3} x_{\mathrm{HI}}^{-1/3} (1+z)^{-1/3}. \tag{2.3.6}$$

Thus, the size of the ionized region depends on the total ionizing output of the quasar, which depends on the quasar magnitude and lifetime, as well as the neutral hydrogen density in the quasar environment. In an inhomogeneous medium, the recombination timescale in Equation 2.3.4 has to be modified in terms of the clumping factor $C_{\mathrm{H\,II}}$ as defined in Equation 1.2.4 of Section 1.2. This size of the ionized region was derived assuming the boundary between ionized and unionized regions is sharp. In the context of reionization, quasars are likely located in already ionized regions (e.g., Yu & Lu, 2005), so that the transition is between a highly ionized region due to the quasar and an ionized region created by the background radiation dominated by stars and galaxies. This makes it difficult to define the size of the ionized region. Rather, what one defines is the proximity zone size, $R_{\mathrm{p}}$, which can be consistently defined between simulations and observations. In general, several definitions were used to define the proximity zone size (e.g., Cen & Haiman, 2000; Lidz et al., 2006b). Fan et al. (2006b) were the first to give a definition for the proximity zone size that is used in this thesis and several other recent works (see also Chapter 5). Fan et al. (2006b) defined it as the distance from the Ly$\alpha$ emission line at the quasar systemic redshift to the location at which the Ly$\alpha$ transmitted flux blueward, smoothed by a 20Å boxcar, first drops below 10%. The motivation to use such a definition was that the average transmission in a UVB-reionized region in the Ly$\alpha$ forest was 0.04 at $z \sim 6$, and hence any transmission above this value at redshifts closer to that of the quasar would correspond to the region that has been ionized by the quasar. The proximity zone size defined thus is always lesser than or equal to $R_{\mathrm{ion}}$, since the assumption is that $x_{\mathrm{HI}} = 1$ outside $r > R_{\mathrm{ion}}$, while $R_{\mathrm{p}}$ is defined by the boundary at which $\tau_\alpha = 2.3$, which can correspond to an $x_{\mathrm{HI}}$ that is not necessarily equal to 1. An analytic expression for the maximum proximity zone size has been derived by Bolton & Haehnelt (2007a) by equating the photoionization rate due to quasar with the background photoionization rate $\Gamma_{\mathrm{HI}}$ at 10% transmission, computed using the GP optical depth (Equation 2.1.4) (see also discussion in Section 3.3.1).

Over the years, measured proximity zone sizes have been used to constrain both $x_{\mathrm{HI}}$ and $t_{\mathrm{q}}$. Fan et al. (2006b) measured proximity zone sizes for a sample of 19 quasars with redshifts between $5.7 < z < 6.4$, as shown in Figure 2.2. Using their constraints on the $x_{\mathrm{HI}}$ value at $z \sim 5.7$ and assuming a quasar lifetime of $10^7$ yr, they obtained an $x_{\mathrm{HI}}$ value of $\sim 1.3 \times 10^{-3}$ by considering the scaling of the proximity zone size with redshift from Equation 2.3.6. The FGPA optical depth in Equation 2.1.6 is proportional to $1/\Gamma_{\mathrm{H\,I}}$. This property has been utilized to

**Figure 2.4:** Schematic diagram showing features of the quasar spectrum and their physical origin. The proximity zone is highlighted in the spectrum as well as the ionized region around the quasar.



compare the transmission within and outside the proximity zone, allowing measurement of the photoionization rate $\Gamma_{HI}$ given $\Gamma_{qso}$ (e.g., Calverley et al., 2011). Bolton et al. (2012) measured the temperature at mean density ($T_0$) by fitting a Voigt profile to the transmission lines within the proximity zones of $z \sim 6$ quasars, while carefully accounting for various mechanisms that can cause line broadening using simulations. As discussed later in this thesis (Chapters 3 and 5), the techniques to infer $x_{HI}$, $T_0$, and $\Gamma_{HI}$ from proximity zone sizes are associated with uncertainties related to the quasar lifetime, as well as sample variance and significant sightline-to-sightline scatter in the density and ionization state of the IGM.

Finally, the attenuation of the Lyman-continuum flux has been used to measure the mean free path of hydrogen ionizing photons (Prochaska et al., 2009). At high redshifts, blueward of the Lyman-continuum flux, the only non-zero transmission occurs within the proximity zone of the quasar. Hence, it was necessary to disentangle the contribution of the quasar to the transmission before inferring that of the reionized IGM. Becker et al. (2021) measured the $\lambda_{MFP}$ at the highest redshift in the EoR at $z \sim 6$ by accounting for the contribution of the quasar analytically. The indirect method for measuring the $\lambda_{MFP}$ involved comparison of the evolution of the mean and scatter of the Ly$\alpha$ effective optical depth between simulations and data, to jointly infer $x_{HI}$, $\Gamma_{H\,I}$ and $\lambda_{MFP}$ (Gaikwad et al., 2023; Davies et al., 2024). Interestingly, as I will discuss in detail in Chapter 6, the direct measurement of $\lambda_{MFP}$ was found to be higher than $\sim 2\sigma$ in tension with several reionization simulations, and lower than the $\lambda_{MFP}$ inferred at the

same redshift using indirect measurements.

Measurements of proximity zone sizes have been crucial for inferring quasar lifetimes (e.g., Khrykin et al., 2016). The size of the ionized region, and consequently the proximity zone size, depends on the total number of ionizing photons, which is influenced by the quasar lifetime (Equation 2.3.6). As a quasar begins emitting into the IGM, two timescales are relevant, the photoionization timescale ($\propto 1/\Gamma_{\text{H I,qso}}$) and the recombination timescale ($\propto 1/n_{\text{HII}}\alpha$, Equation 2.3.5). As long as the photoionization timescale is much longer than the recombination timescale, the size of the ionized region will grow with cosmic time. Therefore, the longer a quasar has been active prior to observation, the larger its ionized region is expected to be. With an increase in the number of photoionizations, the number of recombinations also rises until they eventually balance each other. At this point, the size of the ionized region will no longer increase with time.

Numerical simulations are required to predict the evolution of quasar lifetimes with proximity zone sizes in an inhomogeneous medium that may have also been inhomogeneously reionized. Eilers et al. (2017) measured proximity zone sizes for approximately 30 quasars at $z \sim 6$ and found that around 10% of them were smaller than $\sim 2$ pMpc. By using simulations and accounting for metal absorbers that could obstruct the growth of proximity zone sizes, they concluded that the small proximity zone sizes likely indicate quasar lifetimes of $\lesssim 10^4$ yr, with average lifetimes from all proximity zone sizes estimated to be around $\sim 10^6$ yr. Follow-up measurements and simulations that fit the Ly$\alpha$ transmission profile within quasar proximity zones of stacked spectra also suggested an effective lifetime of $\sim 10^6$ yr (e.g., Morey et al., 2021). Such quasar lifetimes fall below the Salpeter timescale discussed in Section1.3, posing a challenge for SMBH growth.

Davies et al. (2020) examined analytic models of proximity zone sizes for quasars with episodic lifetimes (as discussed in Section 1.3) and concluded that the observed distribution of proximity zone sizes disfavours variations in quasar luminosity on timescales longer than $\sim 10^4$ yr. Chen & Gnedin (2021) analyzed the distribution of proximity zone sizes in the CROC reionization simulation and found that the presence of LLS can result in $1 - 2\%$ of 30 Myr old quasars having small proximity zone sizes $< 1$ pMpc, similar to those observed by Eilers et al. (2017, 2020). In all these analyses, the modeling of proximity zone sizes from quasar spectra was performed using cosmological hydrodynamical simulations with a uniform $\Gamma_{\text{H I}}$ or UVB

during reionization, or using cosmological hydrodynamical radiative transfer simulations with small box sizes that are insufficient to capture the large-scale reionization topology or the sizes of ionized bubbles, while also following an early reionization history.

## 2.4   This Thesis

The main goals of the thesis were to address the following questions using simulations and observations of quasar proximity zones at high-redshifts of $z \sim 6$.

1. **Modeling Quasar Proximity Zones:** Previous modeling of quasar proximity zones has been limited by assumptions about the ionization state of the IGM and the small volumes of reionization simulations. In this thesis, we examine proximity zones in our reionization simulation, which reproduces several other observables from the EoR (Satyavolu et al., 2023a). Proximity zone sizes in such a late reionization model have not been investigated, and the extent to which proximity zones are obstructed by neutral islands at the later stages of reionization remains uncertain.

2. **Quasar Proximity Zone Sizes and SMBH Masses:** The observed distribution of quasar proximity zone sizes has been challenging to reconcile with their SMBH masses when assuming a 'lightbulb' lightcurve for the quasar. We incorporate quasar variability in our simulations of proximity zones and compare them with observations. Additionally, we propose possible scenarios for SMBH growth to reconcile the proximity zone model with SMBH masses (Satyavolu et al., 2023a).

3. **New Measurements of Proximity Zones Using High-Quality Spectra:** Despite the presence of over 200 quasars in the EoR, fewer than half have had high-quality spectra. The XQR-30 survey, an ESO large program targeting quasars in the EoR, provides the highest quality spectra for such objects (D'Odorico et al., 2023b). We measure proximity zone sizes for the E-XQR-30 sample, increasing the total number of proximity zone sizes in the EoR to 87 (Satyavolu et al., 2023b).

4. **Impact of Quasar Proximity Zones on the Direct Measurements of the MFP:** Direct measurements of the mean free path of hydrogen ionizing photons in the EoR were not

only smaller than indirect measurements but also inconsistent with several reionization simulations. These direct measurements relied on an analytic model of quasar proximity zones to estimate the contribution of quasar radiation to the background and the $\lambda_{\mathrm{MFP}}$. However, this model does not account for the inhomogeneously ionized and overdense IGM surrounding the quasar, which could lead to an over-prediction of proximity zone size. We analyze the robustness of this analytic model by comparing it with our late reionization simulations. Furthermore, we propose a new direct method to measure the $\lambda_{\mathrm{MFP}}$ using simulated model stacks, aiming to reconcile the measurements of Becker et al. (2021) with those from indirect measurements and simulations (Satyavolu et al., 2023c).

In Chapter 3, we will discuss our model for quasar proximity zones in a partially ionized IGM during reionization. In Chapter 4, we will discuss the need for obscured growth of SMBHs in the early Universe based on comparisons between proximity zone models that are consistent with data and SMBH mass measurements. In Chapter 5, we discuss our measurements of quasar proximity zone sizes of the XQR-30 sample. In Chapter 6, we discuss the robustness of the $\lambda_{\mathrm{MFP}}$ measurements in the EoR. We conclude with a summary and future outlook on quasars in the EoR. Our measurements as well as theoretical models assume $\Lambda$CDM cosmology with $\Omega_{\mathrm{b}} = 0.0482$, $\Omega_{\mathrm{m}} = 0.308$, $\Omega_{\Lambda} = 0.692$, $h = 0.678$, $n_{\mathrm{s}} = 0.961$, $\sigma_8 = 0.829$, and $Y_{\mathrm{He}} = 0.24$ (Planck Collaboration XVI, 2014).

# Chapter 3

# Quasar Proximity Zones in the EoR

Measurements of proximity zone sizes in rest-frame UV spectra of high-redshift quasars are valuable ingredients for constraints on the growth history of SMBH. Proximity zone sizes have now been measured for a handful of quasars with $z > 6$ (Fan et al., 2006b; Carilli et al., 2010; Mortlock et al., 2011; Venemans et al., 2015; Mazzucchelli et al., 2017; Eilers et al., 2017; Bañados et al., 2018; Ishimoto et al., 2020; Reed et al., 2017; Bañados et al., 2021). The resultant values range from 10 pMpc to 0.14 pMpc across redshifts 5–7. The highest redshift at which a proximity zone size has been measured is at $z = 7.54$ (Bañados et al., 2018) for a quasar of magnitude $M_{1450} = -26.7$ for which the proximity zone size is 1.3 pMpc, a factor of three to four smaller than typical proximity zones measured at redshift $z \sim 6$. These measured proximity zones have been used to estimate lifetimes of redshift-6 quasars to be around $10^6$ yr on average (Eilers et al., 2017; Davies et al., 2020; Eilers et al., 2020; Morey et al., 2021). Interestingly, Eilers et al. (2017) and Eilers et al. (2020) reported the discovery of seven quasars with extremely small proximity zone sizes that appear to imply very short quasar lifetimes of about $10^4$ yr.

Lifetimes as small as $10^4$ yr (Eilers et al., 2017, 2020; Andika et al., 2020) are challenging for SMBH formation models. The Salpeter time (Salpeter, 1964), or the e-folding time, for a black hole growing exponentially at the Eddington limit with a radiative efficiency of 0.1 is $4.5 \times 10^7$ yr. Therefore, if the quasar lifetime is only $10^4$ yr, which is $\sim 0.005 \, t_{\mathrm{Salpeter}}$, then the black hole hardly grows, as

$$M_{\mathrm{BH}} = M_{\mathrm{seed}} \exp\left(\frac{t_{\mathrm{q}}}{t_{\mathrm{Salpeter}}}\right). \tag{3.0.1}$$

This requires the black hole seed to be heavier than even the most massive direct collapse black hole seeds suggested ($\sim 10^6 \, \mathrm{M_\odot}$; Inayoshi et al. 2020).

It is therefore important to critically examine the inference of quasar lifetimes from observed proximity zone sizes. Three important uncertainties that affect this inference are the large-scale ionization environment of the quasars, their large-scale cosmological density environment, and quasar variability.

First, the redshift range inhabited by these quasars is also witness to a rapid, large-scale change in the ionization state of the Universe due to reionization. While the details of how reionization occurs and what causes it remain uncertain, it has been argued recently that the spatial fluctuations in the $z \sim 5$–6 Ly$\alpha$ forest require reionization to end as late as $z \sim 5.3$ (Kulkarni et al., 2019). It has also been a common, conservative, assumption in the literature that reionization is caused by the hydrogen-ionizing radiation produced by young massive stars in star-forming galaxies. As a result, the ionization and thermal state of the medium in which a proximity zone is produced is already affected by a complex interplay of stellar radiation and the intergalactic hydrogen. Therefore, it is necessary to include a proper model of reionization while simulating proximity zones. Previous models of quasar proximity zones often made simplifying assumptions about the ionization and thermal environment of high-redshift quasars. They either assumed the initial ionization state around the quasar to be set by a homogeneous UV background or to be uniformly ionized or neutral (Bolton & Haehnelt, 2007a; Maselli et al., 2007; Bolton et al., 2011; Keating et al., 2015; Eilers et al., 2017). Lidz et al. (2007) were the first to point out that this assumption would not be representative of the inhomogeneous IGM at $z \sim 6$. They performed three-dimensional radiative transfer simulations to obtain the patchy UV background at this redshift, but found that the patchy ionization structure of the IGM around the quasar has little effect on quasar proximity zone sizes as quasars tend to reside in regions that are already ionized. Davies et al. (2018c) have modelled proximity zones and damping wings in two $z > 7$ quasars with the help of semi-numerical reionization simulations. Recently, Chen & Gnedin (2021) also implemented patchy ionization in their quasar proximity zone models by means of the CROC radiative transfer simulations, although the models considered by them reionize too early to be consistent with Ly$\alpha$ forest measurements.

Second, similar to the ionization structure of the quasar environment, the uncertain cosmological density structure around high-redshift quasars can also potentially play a role in setting

the proximity zone size. The distribution of host halo masses of high-redshift quasars is not well understood. It is often assumed that the most luminous quasars reside in the most massive halos (e.g., Springel et al. 2005), but observationally the evidence is uncertain (Coil et al., 2007; Kim et al., 2009). Shen et al. (2007) measured clustering around $z \sim 3$ quasars, which suggested that they lived in massive halos with a minimum mass of $\sim 10^{12}$ M$_\odot$. García-Vergara et al. (2022) have also reported strong clustering of galaxies around redshift $z \sim 4$ quasars. However, based on the spatial correlation of quasars with protoclusters, Uchiyama et al. (2018) inferred that luminous quasars around $z \sim 4$ do not reside in the most overdense regions. Mignoli et al. (2020) found an overdensity of galaxies around a quasar at redshift as high as $z \sim 6.31$. There is no consensus on the overdensity around quasars in simulations either. Costa et al. (2014) report that quasars must reside in the most massive halos in highly over dense regions to be able to grow as massive as $10^9 M_\odot$ by redshift 6 without requiring super-Eddington accretion. The BlueTides simulations (Di Matteo et al., 2017; Tenneti et al., 2018) find that massive black holes are formed not in massive halos, but in halos with low tidal fields. They suggest that the most massive black holes should also have formed in environments similar to low mass black holes. Habouzit et al. (2019) use the Horizon-AGN simulation to study the environment of high-redshift quasars and conclude that statistically most massive black holes reside in regions with high galaxy counts. Fanidakis et al. (2013) used semi-analytic models to study the dark matter environment of quasars and conclude that they live in average mass halos. Ren et al. (2021) used semi-analytical modelling of the relationship between the quasar luminosity and the host halo mass to predict clustering around high-redshift quasars. Keating et al. (2015) were the first to study the role of halo environment on quasar proximity zones. They argued that proximity zone properties do not depend strongly on the host halo mass of the quasars.

Third, accretion onto SMBHs can be episodic. This variability is conventionally quantified using duty cycles and episodic times, where the duty cycle is the fraction of the quasar lifetime for which a quasar is shining, while the episodic time is the duration of each luminous episode. Worseck et al. (2021) inferred short episodic times $\lesssim 1$ Myr for four of the thirteen quasars they studied at redshifts $z \sim 3$ from He II proximity zones. These small episodic times were independent of the quasar magnitude, black hole mass, and Eddington ratio, which suggested that their observations must have sampled quasars with short episodic times and large duty cycles. They also remark that if high-redshift quasars follow a similar trend, then most of the black hole

growth must have happened during the obscured phase. Schawinski et al. (2015) estimate that each accreting phase of SMBHs should last around $10^5$ yr based on the time lag between AGN switching on, becoming visible in X-rays, and becoming visible through photoionized narrow lines of the host galaxy. SMBH simulations also suggest that the accretion occurs in episodes shorter than 1 Myr (Anglés-Alcázar et al., 2017b; Beckmann et al., 2019; Massonneau et al., 2023). Shen (2021) constrained the episodic phases to last for $10^3$–$10^5$ yr based on statistics of 'turned-off quasars' and massive galaxies with orphan broad Mg II emission. So far, the modelling of proximity zones using cosmological simulations post-processed with radiative transfer codes have mostly assumed simple light curves for the quasar, namely the 'lightbulb' model where the quasar is shining at a constant luminosity throughout its lifetime. Davies et al. (2020) describe an analytical model to predict proximity zone sizes of quasars with blinking light curves as well as for more general light curves and found their model to be in good agreement with their simulations. They conclude that the distribution of proximity zone sizes in such scenarios should allow one to put constraints on the episodic lifetime and duty cycle of the quasar, with their model disfavoring large variations in quasar luminosity below $< 10^4$ yr.

In this chapter and the next chapter, we investigate the effect of the cosmological density environment of quasars, their large-scale ionization and thermal environment, and episodic accretion activity on quasar proximity zones. We develop and use a one-dimensional radiative transfer scheme together with a high-dynamic-range cosmological radiation transfer simulation of reionization that is calibrated to the Ly$\alpha$ forest at $z > 4$. The dynamic range of the simulation allows us to span a wide range of host halo masses. The calibration to the Ly$\alpha$ forest measurements brings a level of realism to our reionization model.
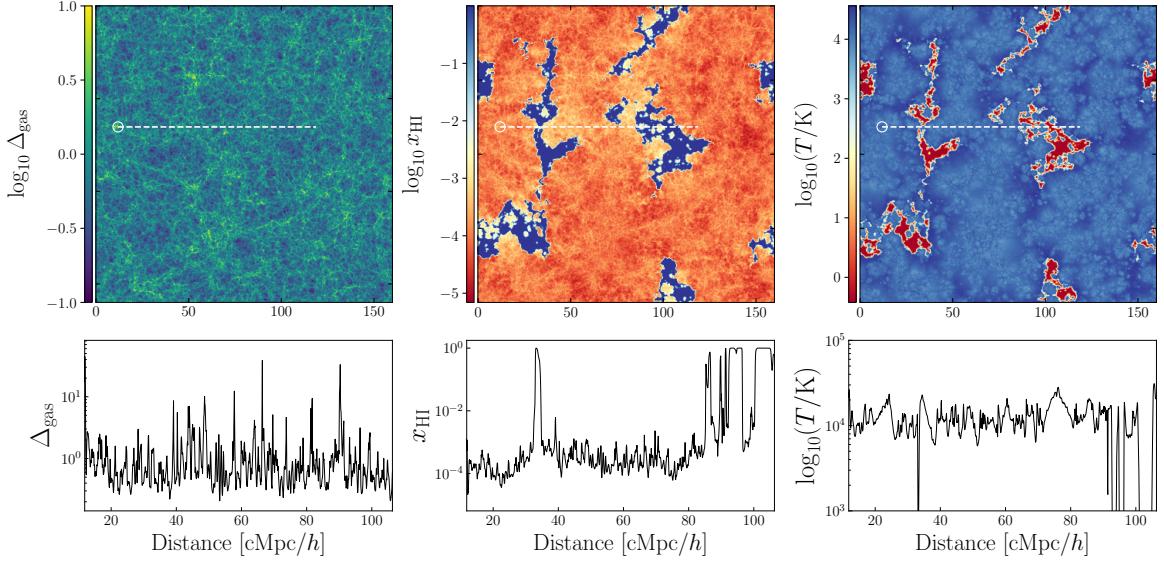
## 3.1   Reionization Simulation

We use the cosmological simulation previously presented by Kulkarni et al. (2019) to set up initial conditions around quasars for redshifts $5 < z < 9$. This model consists of a cosmological hydrodynamical simulation developed using P-GADGET-3 (which is a modified version of GADGET-2, described by Springel 2005), post-processed for three-dimensional radiative transfer using the ATON code (Aubert & Teyssier, 2008, 2010). The box size is 160 cMpc/$h$ with $2048^3$ gas and dark matter particles. The output of the radiative transfer computation is

obtained on a $2048^3$ uniform Cartesian grid with the same box size. The simulation is run from $z = 99$ to 4 with initial conditions chosen to be identical to the 160–2048 simulation from the Sherwood Simulation Suite (Bolton et al., 2017). Snapshots are saved in intervals of 40 Myr. Sources of ionizing radiation are placed at the centers of masses of friends-of-friends groups with mass $> 10^9 \, \mathrm{M_\odot}/h$, with the luminosity of each source proportional to the host halo mass with a mass-independent constant of proportionality (Hassan et al., 2022; Chardin et al., 2015). As discussed by Kulkarni et al. (2019), this model agrees with the measurement of the distribution of Ly$\alpha$ opacities at $z > 5$, and also agrees with several other observations such as the CMB optical depth (Planck Collaboration, 2020), the large-scale radial distribution of galaxies around opaque Ly$\alpha$ troughs (Keating et al., 2020a; Becker et al., 2015b), quasar damping wings (Greig et al., 2017b, 2019; Davies et al., 2018c; Wang et al., 2020), measurements of the IGM temperature (Keating et al., 2020b), and the luminosity function and clustering of Ly$\alpha$-emitters (Weinberger et al., 2018a, 2019). Hydrogen reionization ends at $z = 5.3$ in this model, with the process half-complete at $z = 7$. This picture continues to be consistent with newer Ly$\alpha$ opacity measurements (Bosman et al., 2022). The locality of the moment-based M1 radiative transfer scheme used in this set-up allows the use of GPUs, which speeds up the radiative transfer computation by a factor of more than 100 relative to CPUs (Aubert & Teyssier, 2010). This enables us to enhance the dynamic range of the simulation to include high-mass halos without unduly sacrificing small-scale resolution. At $z = 5.95$, the smallest halo mass resolved in the simulation is $2.32 \times 10^8 \, \mathrm{M_\odot}$, while the largest halo mass is $4.59 \times 10^{12} \, \mathrm{M_\odot}$. We refer the reader to Kulkarni et al. (2019) and Keating et al. (2020a) for further details.

## 3.2 1D Radiative Transfer Simulation

For investigating quasar proximity zones, we use simulation snapshots at $z = 5.95$, 6.60, 7.14, and 8.15 in this thesis. The volume averaged neutral hydrogen fraction at these redshifts is 0.13, 0.37, 0.53 and 0.75 respectively. As an example, Figure 3.1 shows distributions of the gas overdensity $\Delta_{\mathrm{gas}} = \rho_{\mathrm{gas}}/\bar{\rho}_{\mathrm{gas}}$, neutral hydrogen fraction $x_{\mathrm{HI}}$, and gas temperature $T$ at $z = 5.95$. The small white circle in the top three panels of Figure 3.1 marks the location of a halo of mass $6.97 \times 10^{11} \, \mathrm{M_\odot}$. The bottom panels of the figure show the same quantities as the top panel along a one-dimensional skewer drawn from the three-dimensional snapshot, starting

**Figure 3.1:** Distribution of the gas density $\Delta_{\text{gas}}$, neutral hydrogen fraction $x_{\text{HI}}$, and gas temperature $T$ at $z = 5.95$ from our simulation. The small white circle in the top panels marks the location of a halo of mass $6.97 \times 10^{11}$ M$_\odot$ . The bottom panels show the same quantities as the top panel along a one-dimensional skewer drawn from the three-dimensional snapshot, starting from the halo location highlighted in the top panels. The skewer is also shown in the top panels with a dashed white line.

from the halo location highlighted in the top panels. The skewer is also shown in the top panels with a dashed white line, to illustrate the large-scale cosmological environment of this sightline. Large neutral hydrogen patches of up to 100 cMpc/$h$ in size can be clearly seen at this redshift. The gas temperature in these regions is less than 10 K. Although the neutral hydrogen patches in Figure 3.1 are in the deepest voids, making these regions opaque to Ly$\alpha$ photons, the relationship between Ly$\alpha$ opacity and large-scale overdensity is non-linear. This is because regions that are ionized in the most recent past are also in voids, but these regions have higher-than-average temperature, which makes them more Ly$\alpha$ transparent than overdense regions (Keating et al., 2020a). Our simulation set-up allows us to study how this affects quasar proximity zones. In order to obtain model quasar spectra, we place quasars inside massive halos and perform one-dimensional radiative transfer along skewers starting from the halo, similar to the skewer shown in Figure 3.1. The advantage of post-processing using one-dimensional radiative transfer as opposed to three-dimensional radiative transfer is that the computational expense is smaller by several orders of magnitude. Comparisons between three-dimensional and one-dimensional radiative transfer for studying large-scale reionization show little difference in the neutral hydrogen fraction between the two methods (Ghara et al., 2018). The details of our one-dimensional radiative transfer method are discussed in the next section.

We post-process sightlines obtained from the simulation described above with a one-

dimensional radiative transfer computation. The basic equations for the hydrogen and helium ionization chemistry in the presence of photoionization, collisional ionization, and radiative recombination are (Bolton & Haehnelt, 2007a; Rosdahl et al., 2013)

$$\frac{dn_{HII}}{dt} = n_{HI}\left(\Gamma_{HI} + \Gamma_{bg}^{HI} + n_e\beta_{HI}(T)\right) - n_e n_{HII}\alpha_{HII}(T), \tag{3.2.1}$$

$$\frac{dn_{HeII}}{dt} = n_{HeI}\left(\Gamma_{HeI} + \Gamma_{bg}^{HeI} + n_e\beta_{HeI}(T)\right) - n_e n_{HeIII}\alpha_{HeIII}(T)$$
$$- n_{HeII}\left(\Gamma_{HeII} + \Gamma_{bg}^{HeII} + n_e\beta_{HeII}(T)\right) - n_e n_{HeII}\alpha_{HeII}(T), \tag{3.2.2}$$

$$\frac{dn_{HeIII}}{dt} = n_{HeII}\left(\Gamma_{HeII} + \Gamma_{bg}^{HeII} + n_e\beta_{HeII}(T)\right) - n_e n_{HeIII}\alpha_{HeIII}(T). \tag{3.2.3}$$

Here, $\Gamma$ and $\Gamma_{bg}$ denote the photoionization rates of various species induced by quasars and by background sources, respectively. The temperature-dependent collisional ionization rates are denoted by $\beta$. Each $\alpha$ refers to the temperature-dependent recombination rates of respective species, and each $n$ denotes their physical number densities from which the electron number density can be computed as $n_e = n_{HII} + n_{HeII} + 2n_{HeIII}$. Hydrogen and helium abundances were assumed to be of primordial ratio,

$$n_{He} = \frac{Y}{4(1-Y)}n_H, \tag{3.2.4}$$

where the helium mass fraction is $Y = 0.24$.

The photoionization rates $\Gamma_i$ ($i =$ H I, He I, He II) in a shell of volume $dV$ at distance $r$ from the central source are calculated as (Bolton & Haehnelt, 2007a)

$$\Gamma_i(r) = \frac{1}{n_i(r)\,dV(r)}\int_{\nu_i}^{\infty}\frac{L_\nu}{h_P\nu}\exp\left(-\tau_\nu(r)\right)P_i(r)\,d\nu, \tag{3.2.5}$$

where the $\nu_i$'s denote the frequencies corresponding to respective ionization thresholds. The total optical depth is given by $\tau_\nu = \tau_\nu^{HI} + \tau_\nu^{HeI} + \tau_\nu^{HeII}$, where $\tau_\nu^{HI}, \tau_\nu^{HeI}$ and $\tau_\nu^{HeII}$ are the cumulative sums of the respective optical depths of the three species in all previous shells within radius $r$. These are calculated by summing over the opacities of all shells as

$$\tau_\nu^i = \sum_{<r}\Delta\tau_\nu^i = \sum_{<r}n_i\sigma_\nu^i\Delta r, \tag{3.2.6}$$

51

where $i = \mathrm{HI, HeI, HeII}$, with $\sigma_\nu^i$ the respective ionization cross-sections, and the sum is over all shells with equal width $\Delta r$. The quantity $P_i$ in Equation (3.2.5) represents the conditional probability of a photon being absorbed by species $i$ in the shell at $r$ under the condition that the photon is not absorbed by the other two species in that shell. For the three species, this probability is given by

$$P_{\mathrm{HI}} = p_{\mathrm{HI}} \, q_{\mathrm{HeI}} \, q_{\mathrm{HeII}} \left( 1 - e^{-\Delta \tau_\nu^{\mathrm{tot}}} \right) / D, \tag{3.2.7}$$

$$P_{\mathrm{HeI}} = q_{\mathrm{HI}} \, p_{\mathrm{HeI}} \, q_{\mathrm{HeII}} \left( 1 - e^{-\Delta \tau_\nu^{\mathrm{tot}}} \right) / D, \quad \text{and} \tag{3.2.8}$$

$$P_{\mathrm{HeII}} = q_{\mathrm{HI}} \, q_{\mathrm{HeI}} \, p_{\mathrm{HeII}} \left( 1 - e^{-\Delta \tau_\nu^{\mathrm{tot}}} \right) / D, \tag{3.2.9}$$

where $p_i = 1 - e^{-\Delta \tau_\nu^i}$ is the probability that a photon is absorbed by species $i$ in this shell, and $q_i = e^{-\Delta \tau_\nu^i}$ is the probability that the photon is not absorbed by species $i$ in this shell. The quantity $1 - e^{-\Delta \tau_\nu^{\mathrm{tot}}}$, with $\Delta \tau_\nu^{\mathrm{tot}} = \sum_i \Delta \tau_\nu^i$ denotes the total probability that a photon is absorbed in the current cell due to all species. The factor $D = p_{\mathrm{HI}} q_{\mathrm{HeI}} q_{\mathrm{HeII}} + q_{\mathrm{HI}} p_{\mathrm{HeI}} q_{\mathrm{HeII}} + q_{\mathrm{HI}} q_{\mathrm{HeI}} p_{\mathrm{HeII}}$ normalizes the probabilities such that the total number of photons absorbed per unit time in a given cell due to all species is (Mellema et al., 2006)

$$\sum_i n_i(r) \mathrm{d}V(r) \Gamma_i = \int_{\nu_i}^\infty \mathrm{d}\nu \left[ \frac{L_\nu}{h_{\mathrm{P}} \nu} \times \exp(-\tau_\nu(r)) \times \left( 1 - e^{-\Delta \tau_\nu^{\mathrm{tot}}(r)} \right) \right]. \tag{3.2.10}$$

In Equation (3.2.5), $L_\nu$ is the specific luminosity of the quasar. This is related to the total number of photons emitted per unit time as

$$\dot{N} = \int_{\nu_{\mathrm{HI}}}^\infty \frac{L_\nu}{h_{\mathrm{P}} \nu} \mathrm{d}\nu. \tag{3.2.11}$$

For a quasar source, we assume the specific luminosity to be a broken power-law in frequency

$$L_\nu = L_{\mathrm{HI}} \left( \frac{\nu}{\nu_{\mathrm{HI}}} \right)^{-\alpha_s}; \quad \nu > \nu_{\mathrm{HI}}. \tag{3.2.12}$$

The spectral index $\alpha_s$ is chosen as 1.7 based on the profile of quasars observed around $z \sim 3$ (Lusso et al., 2015). By assuming a power law with slope $-0.61$, the specific luminosity at hydrogen ionizing edge $L_{\mathrm{HI}}$ can be computed from the specific UV luminosity $L_{1450}$ at 1450 Å (Lusso et al., 2015), which in turn can be derived from the observed UV magnitude

$M_{1450}$ as

$$L_{1450} = 10^{(51.60 - M_{1450})/2.5} \mathrm{erg\,s^{-1}Hz^{-1}}. \tag{3.2.13}$$

We set the photoionization rate due to background sources, $\Gamma^i_{\mathrm{bg}}$, by using the gas density values of our simulation and assuming equilibrium with the IGM before the quasar is turned on (Chen & Gnedin, 2021). This background ionization rate is of the order of $\sim 10^{-12}\,\mathrm{s^{-1}}$. The size of the proximity zone turns out not to have a strong dependence on the background photoionization for the quasar luminosities that we consider (cf. Eilers et al., 2017; Davies et al., 2020).

The electron collisional ionization rate coefficient values are taken from Hui & Gnedin (1997). We use Case A recombination coefficients (Hui & Gnedin, 1997), which take into account the radiative recombination to all energy levels including the ground state (Bolton et al., 2009; Davies et al., 2016). Secondary ionizations can decrease the temperature within the front and somewhat increase the ionization front size in an initial mostly neutral medium (Davies et al., 2016). But since our medium is initially mostly ionized, the timescale on which photoelectrons lose their energy through collisional ionizations is $t_{\mathrm{loss}} \propto x_{\mathrm{HI}}^{-1} \sim$ a few hundred Myr, secondary electrons do not play a significant role. We ignore them in our computation.

The gas temperature is given by

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{2}{3} \frac{\mu m_{\mathrm{H}}}{\rho k_{\mathrm{B}}} \left( \mathcal{H} - \Lambda \right) - 2HT - \frac{T}{n} \frac{\mathrm{d}n}{\mathrm{d}t} \tag{3.2.14}$$
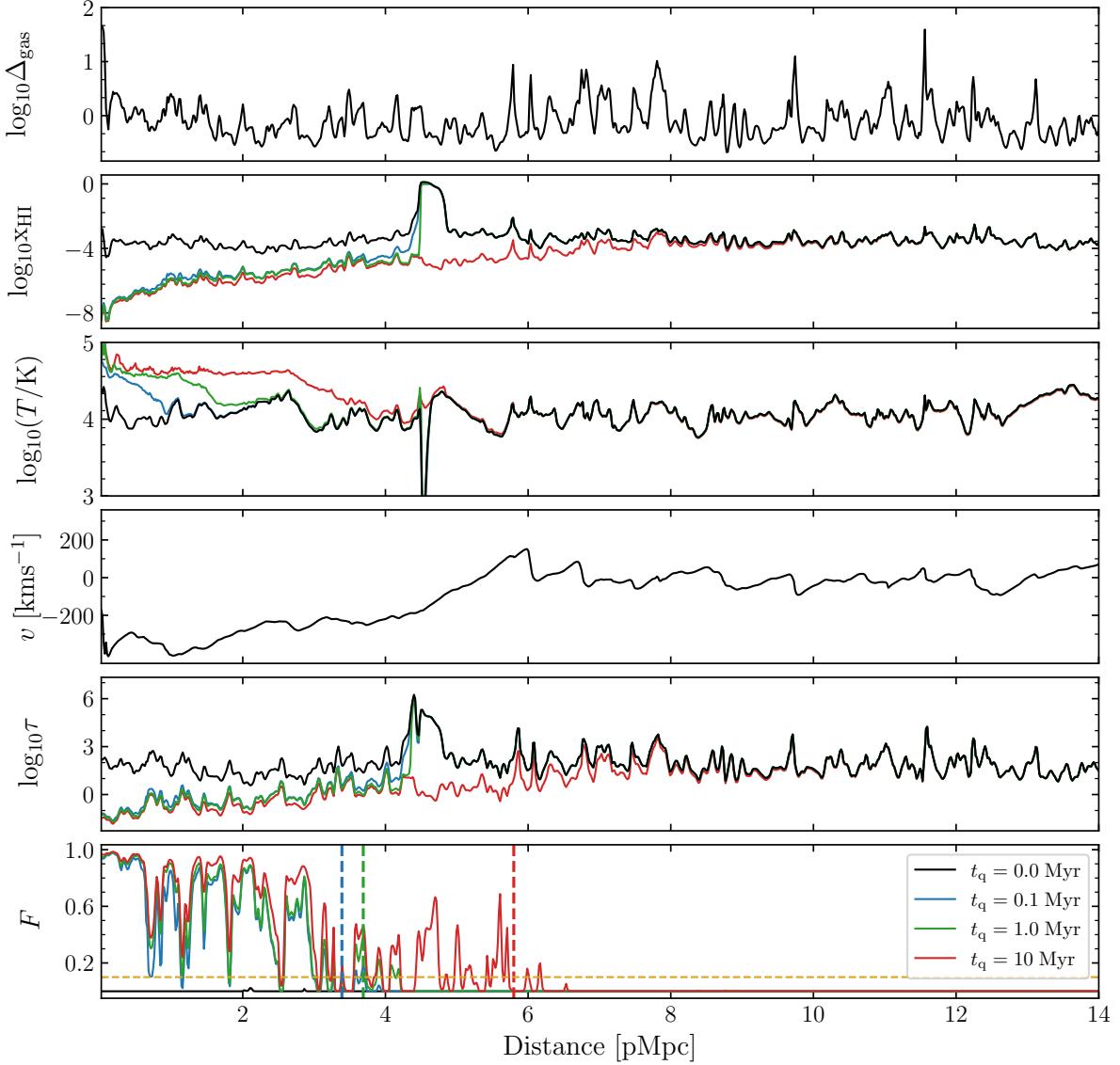
where the heating $\mathcal{H}$ is

$$\begin{aligned}
\mathcal{H} &= \mathcal{H}_{\mathrm{bg}} + \sum_i n_i \epsilon_i \\
&= \mathcal{H}_{\mathrm{bg}} + \sum_i \frac{1}{\mathrm{d}V} \int_{\nu_i}^{\infty} \mathrm{d}\nu \left( h\nu - h\nu_i \right) \frac{L_\nu}{h\nu} e^{-\tau_\nu} P_i.
\end{aligned} \tag{3.2.15}$$

The heating from background sources was set by assuming thermal equilibrium before quasar turn-on (Haardt & Madau, 2012). Similar to Bolton & Haehnelt (2007a), the cooling term $\Lambda$ includes radiative cooling by recombination (Hui & Gnedin, 1997), free-free emission (Cen, 1992), inverse Compton scattering (Peebles, 1971), collisional excitation (Hui & Gnedin, 1997), collisional ionization (Cen, 1992), cooling due to adiabatic expansion of the universe, as well as due to redistribution of heat between different species.

Equations (3.2.1)–(3.2.3) and Equation (3.2.14) are coupled and have to be solved numerically. In order to do this, we discretize each line of sight into uniform cells. The cell size is fixed and is set by our simulation. Following Davies et al. (2016), the photoionization and heating rate integrals are evaluated by sampling the frequencies into 80 logarithmic bins between $\nu_{\mathrm{HI}}$ and $40\nu_{\mathrm{HI}}$. The equations are updated using a modified version of the implicit Euler method (Rosdahl et al., 2013) using a fixed global time step that uniformly applies to all cells in a sightline. After solving the thermochemistry equations, we compute the Ly$\alpha$ optical depth $\tau$ along the line of sight assuming a Voigt absorption profile. We use peculiar velocities from the underlying hydrodynamical simulation in this process. The transmitted flux is calculated as $F = \exp(-\tau)$. We smooth the obtained flux with a boxcar filter of 20 Å and use the definition given by Fan et al. (2006b) to calculate the proximity zone size as the distance at which the smoothed flux drops below 0.1. Appendix A gives details of our algorithm. We present results of several code tests in Appendix B.
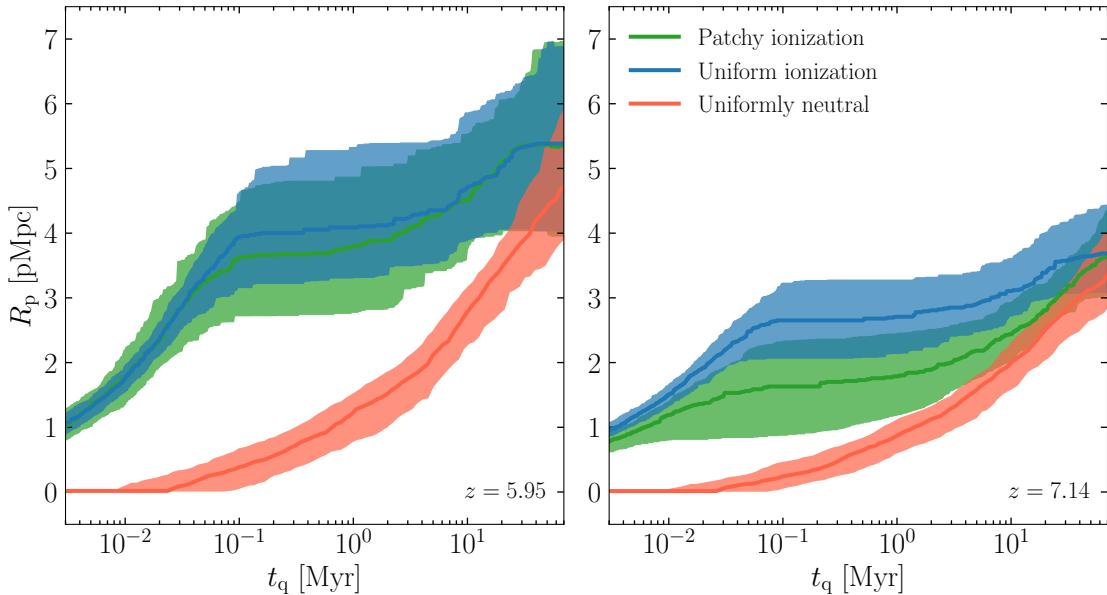
Figure 3.2 shows the result of post-processing the sightline at $z = 5.95$ shown in Figure 3.1 with the above one-dimensional radiative transfer scheme. The figure shows the neutral gas density, neutral hydrogen fraction, gas temperature, peculiar velocities, Ly$\alpha$ optical depth, transmitted flux along the sightline, before the quasar turns on, and 0.1, 1, and 10 Myr after the quasar has turned on. The quasar is of magnitude $M_{1450} = -26.4$ and has a power law spectral index of $\alpha = -1.7$. We assume that the quasar is on throughout its lifetime $t_{\mathrm{q}}$ with constant luminosity (the 'lightbulb model'). The quasar is placed in a halo of mass $6.97 \times 10^{11}$ M$_\odot$. At this redshift, the sightline is initially almost fully ionized with a few patches of neutral hydrogen that can be seen at $\sim 5$ pMpc. This results in a temperature of nearly $\sim 10^4$ K everywhere except in the neutral regions. Given the high cross-section for absorption of the Ly$\alpha$ line, a neutral fraction of $10^{-4}$ is sufficient to entirely absorb radiation along the sightline before the quasar turns on. The flux $F$ is therefore uniformly zero at $t_{\mathrm{q}} = 0$. After the quasar has turned on, the quasar radiation further ionizes the gas as recombinations are negligible ($\alpha \propto T^{-0.7}$), resulting in a decrease of neutral hydrogen fraction, while photoionization concomitantly heats up the gas. As the number of ionizations increase, the recombination rate increases and eventually an equilibrium is reached where the quasar and background radiation balance the recombinations. This results in a temperature of nearly $\sim 10^4$ K almost everywhere. The proximity zone size $R_{\mathrm{p}}$ shown at each of the quasar ages $t_{\mathrm{q}} = 0.1$, 1, and 10 Myr thus increases with lifetime, making it

**Figure 3.2:** The result of applying our one-dimensional radiative transfer scheme to the sightline at $z = 5.95$ shown in Figure 3.1. From top to bottom, the panels show the gas density $\Delta_{\mathrm{gas}}$, neutral hydrogen fraction $x_{\mathrm{HI}}$, gas temperature $T$, the peculiar velocity $v$, the Ly $\alpha$ optical depth $\tau$, and the transmitted flux $F$ along the sightline. Black curves show the quantities before the quasar turns on; colored curves correspond to quasar lifetimes of $t_{\mathrm{q}} = 0.1$, 1, and 10 Myr. In the bottom panel, the dashed horizontal line shows the 10% transmission cut-off used to define the proximity zone size, following Fan et al. (2006b). The corresponding proximity zone sizes are demarcated by the vertical dashed lines in this panel.

sensitive to lifetime measurements until an equilibrium is reached.

We test the numerical convergence of our $R_{\mathrm{p}}$ computation by projecting the density field on grids with successively finer spatial resolution. We find that the proximity zone size computed at our base resolution of 160–2048 is converged to a relative error of 15%, with the proximity zone sizes being smaller at higher resolution.

**Figure 3.3:** Proximity zone size $R_p$ as a function of quasar lifetime at redshifts $z = 5.95$ (left panel) and $z = 7.14$ (right panel). The green curve shows the median evolution of proximity zone size $R_p$ with quasar lifetime $t_q$ from a sample of 100 sightlines with the initial ionization conditions taken from our reionization model. Shaded region shows the $1\sigma$ (68.26% equal-tailed credible interval) scatter across sightlines. The blue curve and shaded region show the same quantities from a case in which the IGM around the quasar is uniformly ionized. Similarly, the red curve and shaded regions show results from an initial fully neutral IGM.

## 3.3 Proximity Zone Sizes in Late Reionization Models
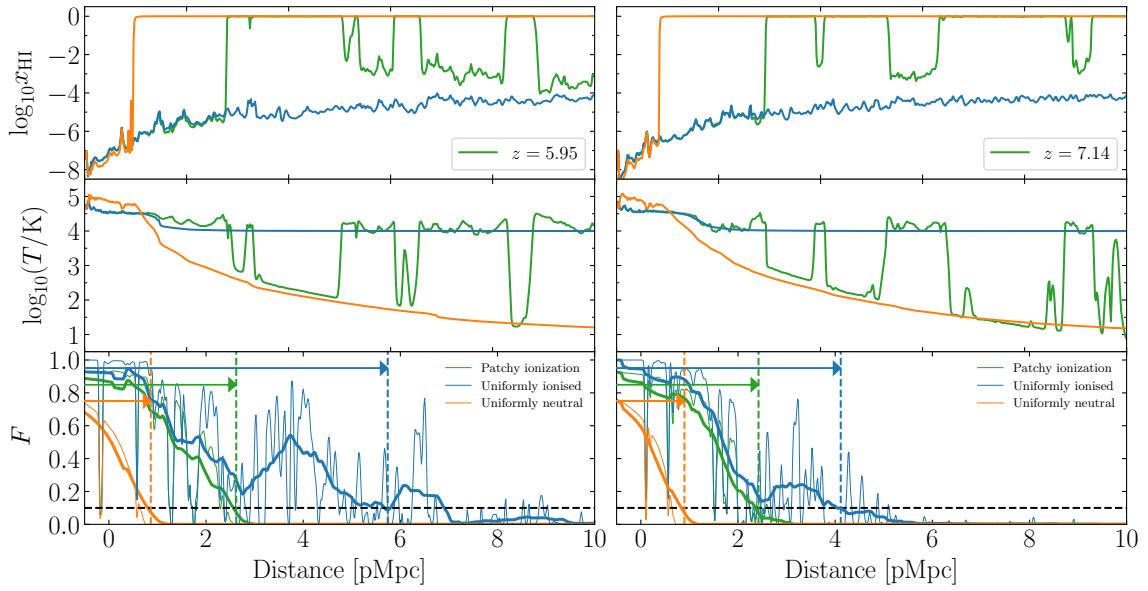
The computational set-up described above now allows us to discuss the effect of the ionization and cosmological environment of quasars as well as quasar variability on the quasar proximity zones which will be discussed in the below Sections 3.3.1 and 3.3.2 and section 4.1 in the following Chapter.

### 3.3.1 Effect of reionization topology on proximity zone size

Previous studies (Eilers et al., 2017; Davies et al., 2020; Keating et al., 2015) have assumed that quasar proximity zones grow in an environment that is either completely neutral or already ionized by a background of hydrogen ionizing radiation. In the latter case, these models set the initial ionization fraction of IGM by invoking a homogeneous UV background. The resultant ionization distribution therefore follows the cosmological gas density distribution and has values around $x_{HI} \sim 10^{-4}$. The CROC simulations (Chen & Gnedin, 2021) also study proximity zones in the presence of an inhomogeneous ionization background at redshift $\sim 6$. However, they consider the reionization to be mostly complete by redshift $z \sim 6$. Lidz et al. (2007) consider effects of patchy reionization at $t_q = 1$ Myr and conclude that it can lead to longer proximity zone

size than in a uniformly ionized medium since quasars are likely to be born inside massive halos that reionize earlier than a typical region. They also point out that the huge sightline-to-sightline scatter in such models might lead to misinterpretations about the ionization state of IGM from $z \sim 6$ quasar observations. Our goal here is to examine if inhomogeneous reionization resulting from late-end reionization models such as ours can result in small proximity zone sizes even for longer quasar lifetimes or lead to longer proximity zone sizes for smaller lifetimes. To study the effects of inhomogeneous reionization on proximity zones, we consider quasars situated in the most massive halos in our simulation with masses between $10^{11}$ $M_\odot$ and $10^{12}$ $M_\odot$. All quasars are assumed to have $M_{1450} = -26.4$ corresponding to $\dot{N} = 1 \times 10^{57}$ photons/s.

Figure 3.3 shows the median evolution of the proximity zone size $R_{\rm p}$ for 100 sightlines, for different initial ionization conditions around the quasar, at $z = 5.95$ and $z = 7.14$. The figure also shows the $1\sigma$ spread in the proximity zone sizes (68.26% equal-tailed credible interval). Our fiducial computation uses the initial ionization and temperature values for the IGM from our underlying radiative transfer simulation (see Section 3.1). We also compare this with a scenario in which the IGM is initially uniformly ionized. In this case, the ionized hydrogen fraction is assumed to be $x_{\rm HI} = 10^{-4}$ throughout the box, with a temperature of $T = 10^4$ K. Finally, we consider a case in which the IGM is initially completely neutral, so that initially $x_{\rm HI} = 1$ and $T = 10$ K throughout the box. The overall evolution shows well-known behavior: $R_{\rm p}$ initially increases due to the small photoionization timescale and then becomes constant as ionization equilibrium is reached as recombinations increase and become equal to ionizations. This is followed by a slight increase in $R_{\rm p}$ at later times, which is because of the larger photoionization timescale of He II and He III that delays their ionization. The associated increase in temperature leads to a decrease in recombination rate and neutral hydrogen fraction, which results in increased transmitted flux. We see that the initial ambient ionization conditions have a large effect on the evolution of $R_{\rm p}$. When the initial ionization state of the IGM is fully neutral, proximity zones are smaller and grow slowly owing to the damping wing as seen in Figure 3.4. However, it is also interesting that patchy ionization conditions also have an effect on the growth of $R_{\rm p}$. The difference between the fully ionized and the patchy ionized cases is relatively stronger at $z = 7.14$, when reionization is half complete in our model. In the patchy ionization case, the upper bound of the $R_{\rm p}$ distribution reaches equilibrium sooner, while the lower bound takes longer than a few Myr to equilibrate. There is a steeper increase
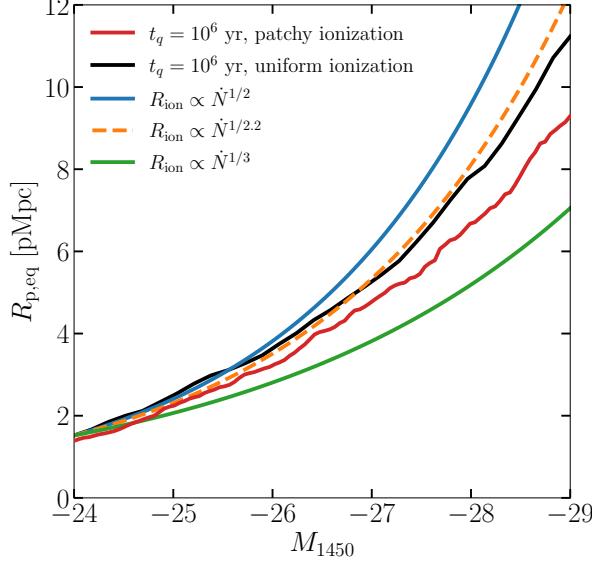
**Figure 3.4:** Proximity zone size $R_\mathrm{p}$ for different initial conditions: The two panels show neutral hydrogen fraction, temperature and transmitted Ly $\alpha$ flux at a quasar age $t_\mathrm{q} = 1$ Myr at two redshifts $z = 5.95$ and $7.14$ for the lightbulb model along example sightlines. Green, blue, and orange curves show various quantities from the cases with a patchy initial ionization, uniform initial ionization, and no initial ionization. In the bottom panels, the thick curves show the transmitted flux after being smoothed by a 20 Å boxcar filter and the dashed horizontal line shows the 10% transmission cut-off used to define the proximity zone size. The corresponding proximity zone sizes are demarcated by the vertical dashed lines in these panels. Neutral hydrogen patches in the IGM impede the growth of the proximity zone relative to the uniformly ionized case.

in $R_p$ post-equilibrium. The difference between the patchy and uniform cases is also clearly seen in Figure 3.4, which shows two example sightlines at redshifts $z = 5.95$ (left panel) and $z = 7.14$ (right panel). Green, blue, and orange curves show various quantities from the cases with a patchy initial ionization, uniform initial ionization, and no initial ionization. It can be clearly seen that patches of intergalactic hydrogen along the sightline impede the growth of the proximity zone.

It is instructive to investigate the dependence of the equilibrium proximity zone sizes $R_\mathrm{p,eq}$ on quasar magnitude $M_{1450}$ and redshift $z$. For the case shown in Figure 3.3, we can safely assume equilibrium at a quasar age of $t_\mathrm{q} = 1$ Myr. Once equilibrium is reached, the ionization front radius in a homogenous medium will be equal to the Stromgren radius

$$
R_\mathrm{ion} = \left( \frac{3\dot{N}}{4\pi n_\mathrm{H}^2 \alpha} \right)^{1/3}.
\tag{3.3.1}
$$

The proximity zone size, $R_\mathrm{p}$, as discussed in the previous section, is defined not by the position of the ionization front but by the point at which the Lyman-$\alpha$ flux crosses an assumed threshold. But one might still expect $R_\mathrm{p,eq} \propto \dot{N}^{1/3}$ following the Stromgren argument. Bolton & Haehnelt

**Figure 3.5:** The scaling of the proximity zone size at equilibrium, $R_{\mathrm{p,eq}}$, with the quasar magnitude at redshift $z = 5.95$. Black curve shows the mean $R_{\mathrm{p,eq}}$ in the case of an initially uniformly ionized IGM; the red curve shows the mean $R_{\mathrm{p,eq}}$ in the case of patchy ionization. The other curves show the scaling motivated by previous models in the literature. Patchy ionization has an influence on this scaling, such that the proximity zone sizes for bright quasars are relatively smaller compared to the uniformly ionized case.

(2007a) derive an analytic estimate for $R_{\mathrm{p,eq}}$ by considering the flux to be set by the Gunn-Peterson optical depth,

$$R_{\mathrm{p,eq}} \quad = \quad \frac{3.14}{\Delta_{\mathrm{lim}}} \left( \frac{\dot{N}}{2 \times 10^{57}\,\mathrm{s}^{-1}} \right)^{1/2} \quad \times \quad \left( \frac{T}{2 \times 10^4\,\mathrm{K}} \right)^{0.35} \left( \frac{1+z}{7} \right)^{-9/4} \mathrm{pMpc}, \quad (3.3.2)$$
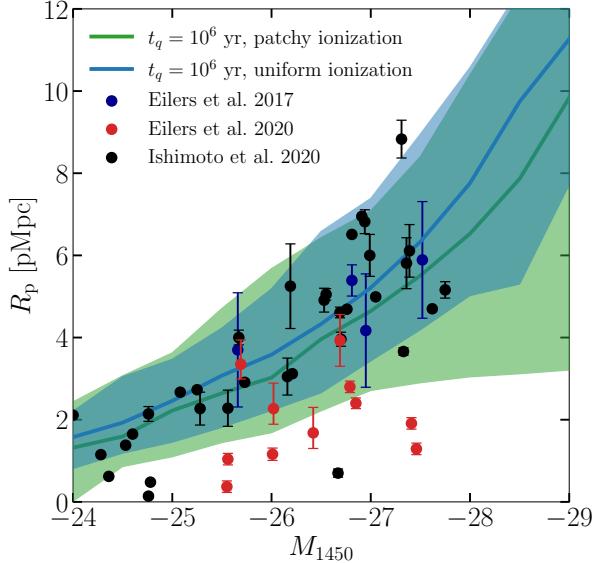
where $\Delta_{\mathrm{lim}}$ is the gas density corresponding to the flux threshold used to define the proximity zone. Davies et al. (2020) use the scaling of effective optical depth derived from their simulations instead of Gunn-Peterson optical depth. They derive an analytical value for equilibrium $R_{\mathrm{p}}$ as

$$R_{\mathrm{p,eq}} = r_{\mathrm{b}} \left[ \left( \frac{\tau_{\mathrm{bg}}}{\tau_{\mathrm{lim}}} \right)^{1/\alpha} - 1 \right]^{-1/2}, \quad (3.3.3)$$

where $\tau_{\mathrm{bg}}$ is the effective optical depth in the absence of quasar and $r_{\mathrm{b}}$ is the distance at which quasar radiation equals the background radiation, derived for their simulations as

$$r_{\mathrm{b}} = 11.3 \left( \frac{\Gamma_{\mathrm{bg}}}{2.5 \times 10^{-13}\,\mathrm{s}^{-1}} \right)^{-1/2} \left( \frac{\dot{N}}{1.73 \times 10^{57}\,\mathrm{s}^{-1}} \right)^{1/2} \mathrm{pMpc}. \quad (3.3.4)$$
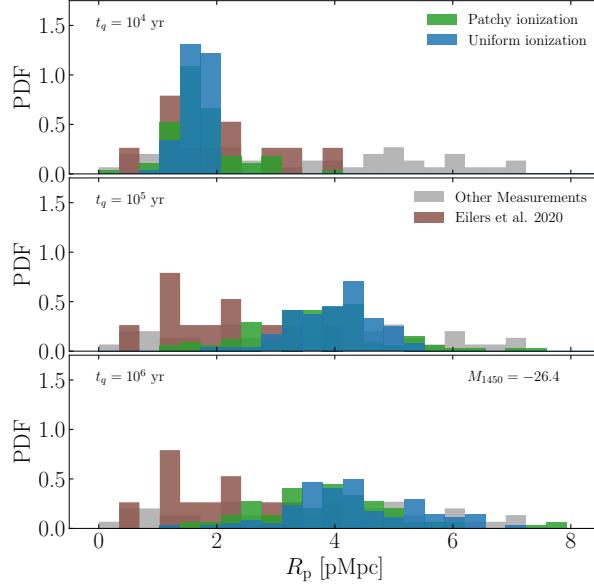
Nonetheless, both of the above estimates suggest a steeper $R_{\mathrm{p,eq}} \propto \dot{N}^{1/2}$. The radiative transfer simulations of Davies et al. (2020) suggest a slightly modified dependence of $R_{\mathrm{p,eq}} \propto \dot{N}^{1/2.2}$.

59

**Figure 3.6:** Effect of patchy reionization on proximity zone sizes at $z \sim 6$ in comparison to data. The curves and the shaded regions show the median with 90% scatter from a sample of 100 sightlines, with the uniform ionization case shown in blue and patchy ionization case shown in green.

Figure 3.5 compares these estimates to our simulated results by showing the mean evolution of $R_{\rm p}$ as a function of magnitude for different initial conditions. We observe that the $R_{\rm p,eq}$ is proportional to $\dot{N}^{1/2.2}$ in good agreement with Davies et al. (2020) in the case of an initially uniformly ionized case. With patchy ionization conditions, the proximity zone sizes are reduced, preferentially for bright quasars, and the dependence of $R_{\rm p,eq}$ on $\dot{N}$ is much shallower, as illustrated by the red curve in Figure 3.5. The topology of reionization can thus have a considerable effect on proximity zone sizes. Note that the effect of patchiness seen in Figure 3.5 is strongly redshift-dependent, with the scaling moving closer to the $\dot{N}^{1/3}$ curve at higher redshifts. This suggests that the scaling of proximity zone sizes invoked in the literature (e.g., Eilers et al., 2017) has limited validity.

In order to examine if this reduction in proximity zone sizes helps reconcile models with data, Figure 3.6 compares the proximity zone size distributions with measurements. The data points in this figure span redshifts between $z = 5.7$ and 6.5. Eilers et al. (2017) have performed analysis on a homogeneous sample of 34 quasars, both new and archival, to obtain proximity zones. Eilers et al. (2020) measured proximity zones by targeting pre-selected quasars that could be potentially young, and performed a multi-wavelength analysis in Eilers et al. (2021) to rule out reduction in proximity zone sizes due to proximate DLAs. Ishimoto et al. (2020) have included only quasars with [CII] and Mg II redshifts in their sample, leading to most precise

**Figure 3.7:** Effect of patchy reionization on the number of small proximity zone sizes at $z \sim 6$ in comparison to data. The brown histogram shows results from the targeted survey by Eilers et al. (2020). The grey histogram shows the other data (Eilers et al., 2017; Ishimoto et al., 2020), which form a homogeneous untargeted sample.

estimates of redshifts and thus proximity zones. They have also updated most measurements from Eilers et al. (2017) with the latest redshifts. All of them use the same definition for proximity zone size $R_{\rm p}$ (Fan et al., 2006b). We see that the data are for relatively faint quasars with $M_{1450} > -28$ that are affected by reionization topology to a lesser degree as compared to brighter quasars. The median proximity zone size shows only a moderate change between the patchy and ionized cases. However, the enhanced spread in the proximity zone sizes in the patchy model can potentially ease the tension between the models and the data. The change in proximity zone sizes because of assuming uniform initial conditions is around 0.29 pMpc at $t = 1$ Myr for a quasar of magnitude $-26.4$, with a maximum change of 0.40 pMpc considering all quasar ages. Meanwhile, the uncertainties in measured proximity zone sizes can range from 0.14–1.43 pMpc (Eilers et al., 2017) for redshift 6 quasars. The uncertainty on $R_{\rm p}$ due to instrumental noise for fainter quasars is of the same order as redshift errors, while for brighter quasars this error is unknown but expected to be small because of the better signal-to-noise ratio (Ishimoto et al., 2020). Therefore, both patchy and uniform ionization models are potentially consistent to within the experimental uncertainty at this redshift.
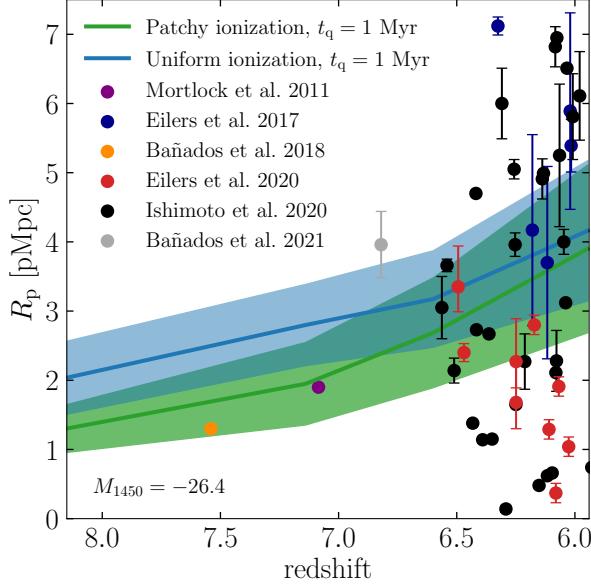
Figure 3.7 shows the reduction in proximity zone sizes in the patchy reionization model for different quasar lifetimes in comparison with data. The incidence of small proximity zones is greater at longer lifetimes in the patchy ionization model relative to the uniform ionization one.

However, in either case, as found previously, the lightbulb model is not sufficient to explain the population of small proximity zones observed.

The reduction in model proximity zone sizes is more significant at high redshifts due to the relatively smaller size of ionized regions around the quasars. Figure 3.8 illustrates this by showing the evolution of the proximity zone size $R_{\rm p}$ with redshift from our model in comparison with data from Mortlock et al. (2011), Eilers et al. (2017, 2020), Bañados et al. (2018, 2021), and Ishimoto et al. (2020). Analytically, the dependence of $R_{\rm p}$ on redshift in a uniform density field can be read from Equation (3.3.2) as $R_{\rm p} \propto (1+z)^{-2.25}$. In the Davies et al. (2020) model, the redshift dependence comes through $\tau_{\rm bg}$ as $(1+z)^{-3.2}$. In both the uniform and patchy cases, there is reduction in the proximity zone size towards high redshift. This is partly driven by the density evolution (cf. Equation 3.3.2). However, the evolution is significantly more rapid in the patchy ionization case due to an additional contribution due to the patchiness. There is also an associated increase in the scatter in the proximity zone sizes. Early measurements of the proximity zone sizes argued for a rapid evolution between redshifts 5.7 and 6.5 (Carilli et al., 2010; Venemans et al., 2015) while more recent observations (Mazzucchelli et al., 2017; Eilers et al., 2017; Ishimoto et al., 2020) have suggested a shallower trend at the same redshifts. More data seems to be necessary to measure the average proximity zone size evolution. With LSST, Euclid, and Roman expected to detect up to several thousand quasars in this redshift range, future follow-up observations will be crucial.

### 3.3.2 Effect of quasar host halo mass on proximity zone size
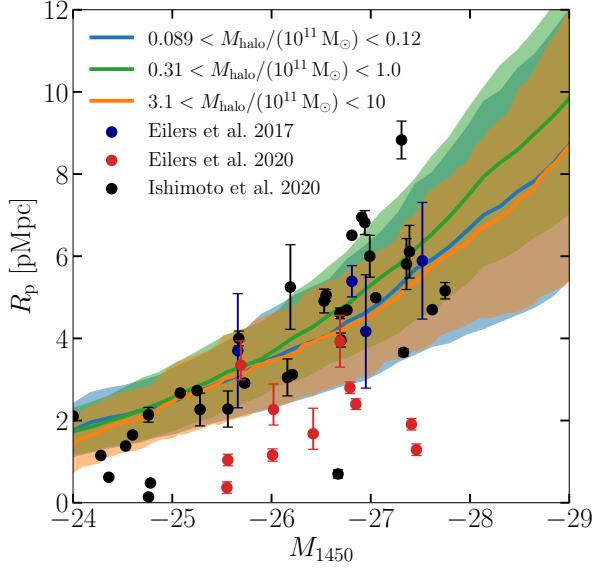
The placement of quasars in the cosmological large-scale structure environment is another point on which models have had to make untested assumptions. This is partly because the host halo masses of $z \sim 6$ quasars are not known, and partly because of the limited dynamic range of simulations. We now discuss the effect of quasar host halo mass on the size of its proximity zone. We consider three mass ranges for the host masses $8.9 \times 10^9 < M_{\rm halo}/{\rm M}_\odot < 1.2 \times 10^{10}$, $3.1 \times 10^{10} < M_{\rm halo}/{\rm M}_\odot < 1.0 \times 10^{11}$, $3.1 \times 10^{11} < M_{\rm halo}/{\rm M}_\odot < 1.0 \times 10^{12}$ and examine proximity zone sizes while assuming equal luminosity for all quasars. Overdensities will typically form galaxies sooner, and reionize earlier, so we would expect these regions to be mostly ionized and have larger proximity zones. However, we do not see a significant scaling of the proximity
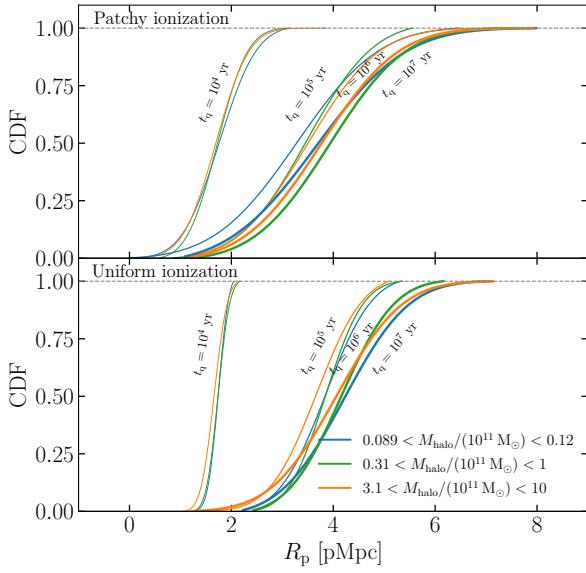
**Figure 3.8:** Evolution of the proximity zone size $R_{\mathrm{p}}$ as a function of redshift for quasar magnitude $M_{1450} = -26.4$ and lifetime of 1 Myr for the lightbulb model. The curve shows the mean from a sample of 100 sightlines. Shaded region shows the $1\sigma$ (68.26% equal-tailed credible interval) scatter across sightlines.

zone size $R_{\mathrm{p}}$ with the host halo mass, as shown in Figure 3.9. This figure shows the median proximity zone size with the associated $1\sigma$ scatter in a sample of 100 sightlines in each mass bin at $z = 5.95$ for a range of quasar luminosities. The quasar age is fixed at 1 Myr in all cases. The lack of a dependence of the proximity zone sizes on the halo mass seen in Figure 3.9 is consistent with previous results (Lidz et al., 2007; Keating et al., 2015).

The cumulative distribution functions (CDFs) of proximity zone sizes helps understand this. Figure 3.10 shows the CDFs of proximity zone sizes in our three halo mass bins for different quasar lifetimes at $z = 5.95$. CDFs for the patchy ionization case as well as the uniform ionization case are shown. The quasar magnitude is fixed at $M_{1450} = -26.4$, and 100 random sightlines are used in each case. We see that the proximity zone sizes in the patchy ionization case are smaller than those in the uniform ionization case, although, as we discussed in the previous section above, this reduction is rather moderate at this redshift. We also see that in the patchy ionization case, quasars in halos with smaller masses have a greater incidence of small proximity zone sizes. This is because regions around small-mass halos reionize later (Keating et al., 2020a), as evidenced by the bottom panel of Figure 3.10. Any dependence of the proximity zone sizes on the halo masses is thus indirect and is caused due to the different ionization conditions around halos of different mass.

**Figure 3.9:** The median proximity zone size $R_{\mathrm{p}}$ and its associated $1\sigma$ scatter at $z = 5.95$ when quasars are placed in one of our three chosen halo mass bins. The quasar age is fixed at 1 Myr here, and 100 sightlines are used in each case. The proximity zone size is not strongly affected by the host halo mass, although there is a weak preference for smaller proximity zones around smaller halos due to patchy reionization, as seen in Figure 3.10.



**Figure 3.10:** The cumulative distribution functions of proximity zone sizes at $z = 5.95$ for three quasar age values and three choices of the quasar host halo mass. The quasar magnitude is fixed at $M_{1450} = -26.4$, and 100 sightlines are used in each CDF. The top panel shows the distributions in the patchy ionization case, while the bottom panel shows the distributions when the IGM is uniformly ionized. There is an enhancement in the incidence of small proximity zone sizes for small host halo masses. But this enhancement is significant only in the patchy ionization case, which suggests that it is indirectly caused by small-mass halo environments getting reionized relatively later.
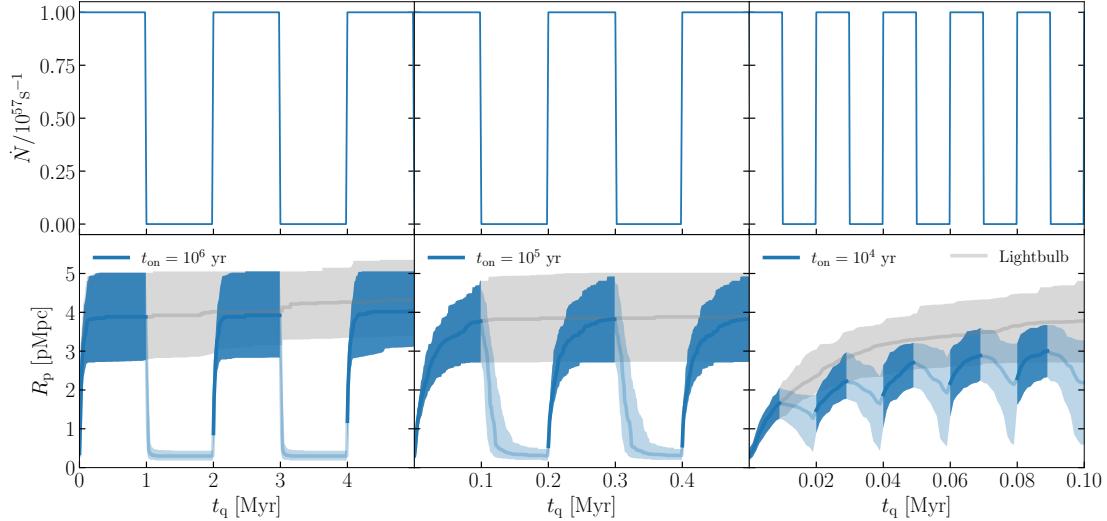
# Chapter 4

# Need for Obscured Growth of SMBHs

In this Chapter, we confront our proximity zone models from Chapter 3 with measurements of proximity zone sizes as well as black hole masses to infer requirements on black hole growth.

## 4.1 Effect of episodic quasar activity on proximity zone size

We now investigate the effect of quasar variability on proximity zone sizes. Observations and simulations suggest that quasars are not constant 'lightbulbs'; they flicker on time scales of $\sim 10^5$ yr or less (Novak et al., 2011; Gabor & Bournaud, 2013; Schawinski et al., 2015; King & Nixon, 2015; Oppenheimer et al., 2018; Shen, 2021). So far, we have been assuming lightbulb quasars in this thesis. While realistic light curves will be much more complex and cannot be described by a constant duty cycle, we consider the simpler 'blinking lightbulb' scenario, where the quasar periodically turns on for a duration of $t_{\rm on}$ and off for $t_{\rm off}$. Once the quasar turns off, the neutral hydrogen fraction will relax to its equilibrium value due to background ionization, on a timescale of $t_{\rm eq} \sim 1/\Gamma_{\rm bg}$. A quasar can be off either because it is obscured by one of several possible mechanisms, or because the black hole is not accreting. The subsequent evolution of the proximity zone size is very different from the lightbulb case if this equilibration timescale is shorter than the time for which the quasar is in the off state. Davies et al. (2020) analytically solved for the behavior of proximity zone sizes in the presence of a flickering quasar with a uniform background radiation. They concluded that the proximity zone sizes are sensitive to episodic lifetime and duty cycle of quasars. Here, we investigate how the distribution of proximity zone sizes changes for varying episodic lifetimes in our patchy

**Figure 4.1:** Proximity zone size evolution in periodically varying quasars. The three columns show quasars that flicker between zero luminosity and $\dot{N} = 10^{57}$ s$^{-1}$ with duty cycle $f_{\text{duty}} = 0.5$ and episodic on time $t_{\text{on}} = 10^6$ yr (left column), $t_{\text{on}} = 10^5$ yr (middle column), and $t_{\text{on}} = 10^4$ yr (right column). Top panels show the quasar light curves; bottom panels show the corresponding evolution of the proximity zone size. The blue curves and blue shaded regions in the bottom panels show the median evolution in a sample of 100 sightlines, and the $1\sigma$ scatter. Regions shaded in lighter blue in the bottom panels show the proximity zone size evolution during the quasar's off period. For comparison, the bottom panels also show the proximity zone size evolution for a corresponding lightbulb quasar.

reionization simulations.

We first consider a simple model in which the quasar flickers periodically between zero and a fixed luminosity corresponding to $\dot{N} = 1 \times 10^{57}$ s$^{-1}$ ($M_{1450} = -26.4$). We consider quasars with $t_{\text{on}}$ between $10^4$ and $10^6$ yr, where the episodic lifetime $t_{\text{on}}$ is the duration of one bright episode in the quasar light curve. The quasar light curve is assumed to be periodic, so that each cycle is characterized by a bright phase with duration $t_{\text{on}}$ and an obscured phase with duration $t_{\text{off}}$. The duty cycle $f_{\text{duty}}$, defined as the fraction of quasar lifetime that the quasar is on, is then $t_{\text{on}}/(t_{\text{on}} + t_{\text{off}})$.[1] The top panels of Figure 4.1 show three example light curves describing this scenario. The quasars are hosted by halos with masses between $10^{11}$ and $10^{12} \text{M}_\odot$. The initial ionization and thermal state are set by our patchy reionization model.

The lower panels of Figure 4.1 show the evolution of the proximity zone size $R_{\text{p}}$ for such periodic quasars. The shaded region shows the $1\sigma$ scatter in $R_{\text{p}}$ among 100 sightlines. The evolution of the proximity zone size during the black hole's bright phase is shown in a bolder colour to distinguish it from the evolution during the obscured phase. The proximity zone size $R_{\text{p}}$ clearly follows the quasar light curve for all three episodic times. This can be understood

---

[1] Alternative, but related, definitions of the duty cycle are also used in the literature. For instance, the duty cycle has also been defined as the fraction of the Hubble time for which the quasar is shining (Haiman et al., 2004), or the ratio of the number of active and quiescent SMBHs (Shankar et al., 2012; Bolgar et al., 2018).

as follows. When the quasar is on, the equilibration time for the ionization front, under the assumption of a constant photoionization rate, is (Khrykin et al., 2016)

$$t_{eq}^{on} \gg \frac{1}{\Gamma_{qso} + \Gamma_{bg} + n_e \alpha}. \tag{4.1.1}$$

This is the time taken for the gas within the proximity zone to settle in the equilibrium state in presence of the quasar. The behaviour of the proximity zone when the quasar is off can be understood as follows. When the quasar turns off, the number of ionizations is reduced to that only due to background photoionization. If the quasar turns off after an equilibrium is reached, then the number of recombinations are higher than the photoionizations immediately after the quasar turns off. This leads to decrease in the ionization fraction until a new equilibrium is reached between the background photoionizations and recombinations. The timescale to reach this new equilibrium therefore depends on the recombination rate as $t_{eq}^{off} \propto 1/n_H \alpha$. On the other hand, if the quasar turns off before an equilibrium is reached between ionizations and recombinations, such that the number of recombinations is still smaller than the background photoionization before quasar turn-off, then post quasar turn-off, the ionized fraction continues to increase, although at a slower rate of $t_{eq}^{off} \propto 1/\Gamma_{bg}$, till it reaches the new equilibrium value. Therefore, the timescale to reach the new equilibrium post quasar turn-off depends on when the quasar turns off once it is turned on.

The timescale for proximity zone to disappear on the other hand depends on the time for the neutral fraction to increase to $\sim 10^{-4}$ once the quasar is turned off, so that the Ly$\alpha$ absorption is saturated. We have

$$\frac{dx_{HI}}{dt} = -(\Gamma_{bg} + \Gamma_{qso})x_{HI} + (1 - x_{HI})^2 \alpha(T)n_H \tag{4.1.2}$$

$$\approx -(\Gamma_{bg} + \Gamma_{qso})x_{HI} + \alpha n_H. \tag{4.1.3}$$

When the quasar turns off, the equation becomes

$$\frac{dx_{HI}}{dt} \approx -\Gamma_{bg}x_{HI} + \alpha n_H. \tag{4.1.4}$$

Integrating until a time $t$ after the quasar turns off, the equation becomes

$$\int_{t_{\mathrm{on}}}^{t_{\mathrm{on}}+t} \frac{\mathrm{d}x_{\mathrm{HI}}}{-\Gamma_{\mathrm{bg}}x_{\mathrm{HI}}+\alpha n_{\mathrm{H}}} \approx \int_{t_{\mathrm{on}}}^{t_{\mathrm{on}}+t} \mathrm{d}t. \tag{4.1.5}$$

Assuming $\alpha, n_{\mathrm{H}}$ and $\Gamma_{\mathrm{bg}}$ to be constants, the solution can be written as

$$x_{\mathrm{HI}}(t) \approx \frac{\alpha n_{\mathrm{H}}}{\Gamma_{\mathrm{bg}}}\left(1-e^{-\Gamma_{\mathrm{bg}}t}\right)+x_{\mathrm{HI}}(t_{\mathrm{on}})e^{-\Gamma_{\mathrm{bg}}t} \tag{4.1.6}$$

The time $t_{\mathrm{vanish}}$ at which the proximity zone disappears is such that

$$x_{\mathrm{HI}}(t = t_{\mathrm{vanish}}) \sim 10^{-4}. \tag{4.1.7}$$

$t_{\mathrm{vanish}}$ can then be computed as
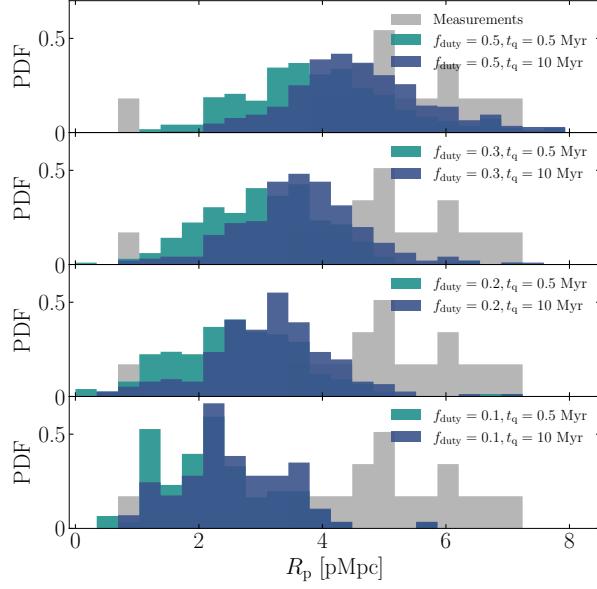
$$t_{\mathrm{vanish}} \approx -\frac{1}{\Gamma_{\mathrm{bg}}}\ln\left(\frac{\Gamma_{\mathrm{bg}}x_{\mathrm{HI}}(t=t_{\mathrm{vanish}})-\alpha n_{\mathrm{H}}}{\Gamma_{\mathrm{bg}}x_{\mathrm{HI}}(t=t_{\mathrm{on}})-\alpha n_{\mathrm{H}}}\right) \tag{4.1.8}$$

Assuming $x_{\mathrm{HI}}(t = t_{\mathrm{on}})$ to be $\sim 10^{-8}$ and substituting $\Gamma_{\mathrm{bg}} \sim 2.5 \times 10^{-13}\,\mathrm{s}^{-1}$, $\alpha$ at $T = 10^{4}$ K as $4.5 \times 10^{-13}\,\mathrm{cm}^{3}\mathrm{s}^{-1}$, and an average $n_{\mathrm{H}} \sim 10^{-4}\,\mathrm{cm}^{-3}$ at redshift $z \sim 6$, we find

$$t_{\mathrm{vanish}} \approx 0.1\,\mathrm{Myr} \tag{4.1.9}$$

For a small $t_{\mathrm{on}} \sim 10^{4}$ yr, $x_{\mathrm{HI}}(t_{\mathrm{on}})$ can be comparable to $10^{-4}$. Assuming $x_{\mathrm{HI}}(t=t_{\mathrm{on}}) \sim 9 \times 10^{-5}$, we obtain $t_{\mathrm{vanish}} \approx 0.01$ Myr. Therefore, $t_{\mathrm{vanish}}$ can have values between $0.01-0.1$ Myr depending on $x_{\mathrm{HI}}(t_{\mathrm{on}})$. This explains why the proximity zone is destroyed more quickly between cycles than it builds up in the two leftmost panels of Figure 4.1 but not in the rightmost panel. This allows some proximity zone growth to accumulate over multiple cycles, but as we will see below this growth is not significant for an on time of $10^{4}$ yr.

If the on-time $t_{\mathrm{on}}$ is greater than the $t_{\mathrm{eq}}^{\mathrm{on}}$ defined in Equation (4.1.1), $R_{\mathrm{p}}$ follows the lightbulb distribution, as seen in the $t_{\mathrm{on}} = 10^{6}$ yr panel in Figure 4.1. For $t_{\mathrm{on}} = 10^{4}$ yr, as in the rightmost panel of Figure 4.1, the on-time is too short for the proximity zone size to equilibrate to its lightbulb value. Consequently, if in this scenario the duty cycle is small enough, so that $t_{\mathrm{off}}$ is greater than $t_{\mathrm{vanish}}$, then the proximity zone size remains much smaller than its lightbulb value

**Figure 4.2:** Distribution of the proximity zone size $R_p$ for a periodically flickering quasar with peak magnitude $M_{1450} = -27$, for various duty cycles and quasar lifetimes. The episodic on time is held fixed to $t_{on} = 10^4$ yr. The proximity zone size is measured only when the quasar is bright. Grey histograms show the homogeneous sample of measurements by Eilers et al. (2017) and Ishimoto et al. (2020).

at all times. This provides a viable mechanism to explain small proximity zones.
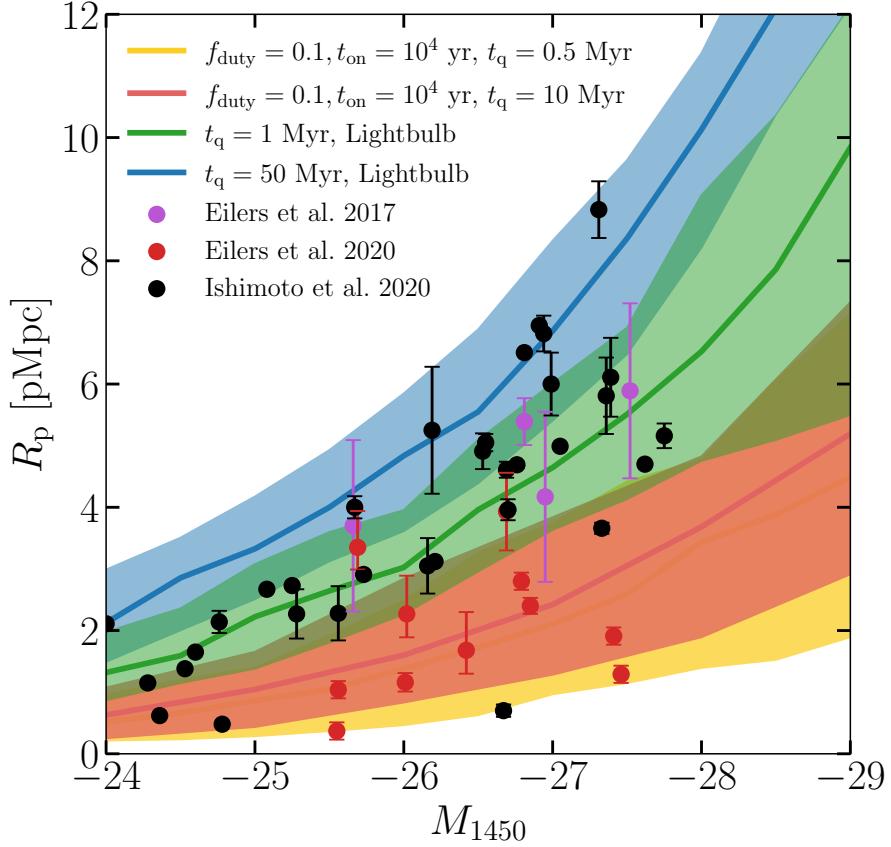
Figure 4.2 demonstrates this by showing the distribution of proximity zone sizes for a quasar with peak magnitude $M_{1450} = -27$, for various duty cycles and quasar lifetimes. The proximity zone size distributions are derived only when the black hole is shining and is accessible to the observer. We show distributions at quasar lifetimes of $t_q = 0.5$ Myr and $t_q = 10$ Myr, to investigate if the small proximity zone sizes vanish at large times. We randomly sample 100 sightlines from the simulation box, add a random relative temporal offset to each quasar light curve, and use the ionization conditions given by our patchy reionization model. The episodic on time is held fixed to $t_{on} = 10^4$ yr, as longer on times will simply take the proximity zone size distribution to the lightbulb value, as we saw in Figure 4.1. We see that smaller values of the duty cycle $f_{duty}$ yield smaller proximity zone sizes. This is as expected from our discussion above. Smaller duty cycles imply longer off times for the quasar, which allows the proximity zone to disappear as the gas in the proximity zone has enough time to equilibrate to the background photoionization rate. We also see that the proximity zone sizes do not increase significantly even for long quasar lifetimes of $t_q = 10$ Myr. There is a small increase in $R_p$ at large $t_q$ for large values of the duty cycle because these large duty cycles correspond to smaller off times, which prevent complete equilibration. But there is virtually no change in the distribution of proximity zone sizes for $f_{duty} = 0.1$ between $t_q = 0.5$ Myr and $t_q = 10$ Myr. Figure 4.2 also compares these

69

proximity zone size distributions with the homogeneous sample of measurements by Eilers et al. (2017) and Ishimoto et al. (2020). We see that models with smaller duty cycles can readily explain even the smallest proximity zones in the data. It has to be noted that the data histograms do not include the redshift errors on the proximity zone measurements which can go up to 40%. Also, the smallest $R_p$ measurements shown in Figure 4.2 are at fainter magnitudes compared to our model which explains the slight discrepancy between models and data for $R_p$ less than $\sim 1$ pMpc.

This picture of variable quasars is put to a more stringent test in Figure 4.3, which aims to model all currently measured proximity zone sizes at $z \sim 6$. This figure shows measurements by Eilers et al. (2017), Ishimoto et al. (2020), and Eilers et al. (2020), in comparison with four models. For each quasar magnitude, we assume a periodic light curve with the given duty cycle and the episodic on time. As before, we randomly sample 100 sightlines from the simulation box, add a random relative temporal offset to each quasar light curve, and use the ionization conditions given by our patchy reionization model. (Figure 4.3 does not show the proximity zone size measured by Ishimoto et al. (2020) for the quasar J1406–0116. This quasar shows no Ly$\alpha$ emission line, making it hard to fit a continuum spectrum. Indeed, Ishimoto et al. (2020) find that the $R_p$ measurement for this quasar changes significantly, increasing by a factor of $\sim 7$, if the continuum fitting method is changed.) We see that the large proximity zones are well fit by the lightbulb model at all magnitudes. These data are therefore also consistent with flickering quasars with long on times. The smaller proximity zones cannot be fit by lightbulb models with $t_q$ as large as $10^7$ yr. But these can be fit by models that have a small episodic on time $t_{on} = 10^4$ yr and small duty cycle $f_{duty} = 0.1$. Furthermore, this model continues to describe the small-$R_p$ data reasonably well even at large quasar lifetimes of $t_q = 10^7$ yr, thus avoiding the need for fine-tuning. The performance of these models can be further improved by reducing the episodic on time $t_{on}$ and the duty cycle $f_{duty}$. We leave a more rigorous development of such models for future work.

## 4.2  Consequences for black hole growth

Although we now see that it is possible to explain the observed small proximity zone sizes via episodic light curves with $t_{on} \sim 10^4$ yr, we should now ask whether such a scenario allows

**Figure 4.3:** Proximity zone size measurements at $z \sim 6$ by Eilers et al. (2017), Ishimoto et al. (2020), and Eilers et al. (2020), in comparison with four models. For each model, we show the median proximity zone size at each quasar magnitude, along with the $1\sigma$ scatter (68.26% equal-tailed credible interval). We randomly sample 100 sightlines from the simulation box, add a random relative temporal offset to each quasar light curve, and use the ionization conditions given by our patchy reionization model. For reasonably large values of $t_q$, large proximity zones can be fit by lightbulb models, while the smaller proximity zones need variable quasars with short episodic on times and small duty cycle. These variable quasar models continue to fit the small proximity zone sizes well even at longer quasar lifetimes $t_q$.

formation of black holes with inferred masses by redshifts $z \sim 6$. This appears to be difficult for the values considered in Figure 4.3. For example, under the assumption that the accretion rate is proportional to the black hole mass, and assuming a radiative efficiency of $\epsilon = 0.1$, a seed of mass $10^3$ $M_\odot$ will require 2 Gyr to grow into a $10^9$ $M_\odot$ black hole with a duty cycle of $f_{duty} = 1/3$, while accreting at the Eddington rate. For a duty cycle of $f_{duty} = 1/10$, equal to what we needed in Figure 4.3 for the smallest observed proximity zones, the required time increases to $\sim 7$ Gyr. Not only are these lifetimes greater than those inferred in Figure 4.3 by more than an order of magnitude, they are also longer than the age of the Universe at $z \sim 6$ by factors of at least two. As discussed extensively in the literature, growing to masses of $10^9$ $M_\odot$ or more during the optically bright phases requires larger seed masses, larger duty cycles, super-Eddington rates, or a combination thereof (Eilers et al., 2021).

One way to alleviate this problem is by having the black hole grow also during obscured

phases. During such phases of obscured growth, the black hole does not shine as a luminous quasar along the observer's sightline. In this scenario, we can discriminate between a 'luminosity duty cycle', $f_{\mathrm{duty,lum}}$, which quantifies the fraction of the black hole's lifetime for which it shines as an optically bright quasar, and an 'accretion duty cycle', $f_{\mathrm{duty,acc}}$, which is the fraction of the black hole's lifetime for which it accretes and grows. If $f_{\mathrm{duty,lum}} < f_{\mathrm{duty,acc}}$, the black hole undergoes obscured growth, whereas if the two duty cycles are equal, the black hole only grows while it is in the luminous quasar phase. Using this terminology, the duty cycle $f_{\mathrm{duty}}$ discussed in the previous section can now be understood as $f_{\mathrm{duty,lum}}$.

Invoking obscured growth now allows us to solve the black hole growth crisis. For example, for the cases discussed above, in which we assumed accretion on to the black hole to be proportional to the black hole's mass, a radiative efficiency of $\epsilon = 0.1$, a luminosity duty cycle of $f_{\mathrm{duty,lum}} = 1/3$ or $1/10$, we now find that an accretion duty cycle of $f_{\mathrm{duty,acc}} = 0.7$ readily allows a seed mass of $10^3$ M$_\odot$ at redshift $z = 15$ to grow into a $10^9$ M$_\odot$ by $z \sim 6$ within the Hubble time ($\sim 670$ Myr) while accreting at a moderately super-Eddington rate of $\sim 1.5$. Pushing the accretion duty cycle closer to unity can even remove the requirement of super-Eddington accretion.

The combination of measurements of quasar proximity zone sizes and the black hole masses thus seem to necessitate obscured growth of SMBHs at high redshifts. Davies et al. (2019) and Worseck et al. (2021) have also argued for obscured black hole growth from their interpretations of proximity zones in hydrogen and helium Ly$\alpha$ forest spectra. For a given observing sightline, obscured black hole growth can occur due to (*a*) orientation effects due to dusty torus close to the AGN, so that the black hole continues to accrete and shine, but not along the given sightline (Antonucci, 1993), or (*b*) small-scale physics near the black hole, such as photon trapping, in which photons are unable to escape because efficient accretion of optically thick material impedes photon diffusion (Begelman, 1979), or (*c*) obscuration by dust, accumulated due to supernovae on the scale of the galaxy (Riechers et al., 2013; Wang et al., 2013; Yajima et al., 2017). Obscuration is also likely to occur at a range of different radii at different times due to different mechanisms over the growth history of SMBHs (Buchner et al., 2015). For instance, by cross-correlating the brightest UV-selected AGN from the GOODS sample at redshifts $z \sim 1$–3 with X-ray measurements from Chandra, Del Moro et al. (2017) inferred an obscured AGN fraction of about 0.67. More generally, the obscured AGN fraction is suggested to vary widely

between 0.1 and 1 with luminosity (Treister et al., 2008; Buchner et al., 2015). The photon trapping picture has been supported by later analytical and numerical work (e.g., Quataert & Gruzinov, 2000; Igumenshchev et al., 2003; Blandford & Begelman, 2004; Takahashi & Ohsuga, 2015) although some models suggest a reduced efficiency of photon trapping with an associated emission of radiation from polar regions of the accreting black hole (Jiang et al., 2014).

The scenario of photon trapping discussed above is usually associated with super-Eddington accretion and therefore low radiative efficiency. Low radiative efficiency is an alternative to the obscuration scenario, but the required radiative efficiency for a $10^3$ $M_\odot$ seed with $f_{\mathrm{Edd}} = 1$ and duty cycle 0.1 to grow into a $10^9$ $M_\odot$ black hole within 1 Gyr would be 0.015. This is consistent with the low radiative efficiencies suggested by Davies et al. (2019), but this radiative efficiency is a factor of more than five smaller than theoretical predictions for standard accretion models (Shapiro & Teukolsky, 1983) and observational measurements (Shankar et al., 2004; Davis & Laor, 2011; Trakhtenbrot et al., 2017).

# Chapter 5
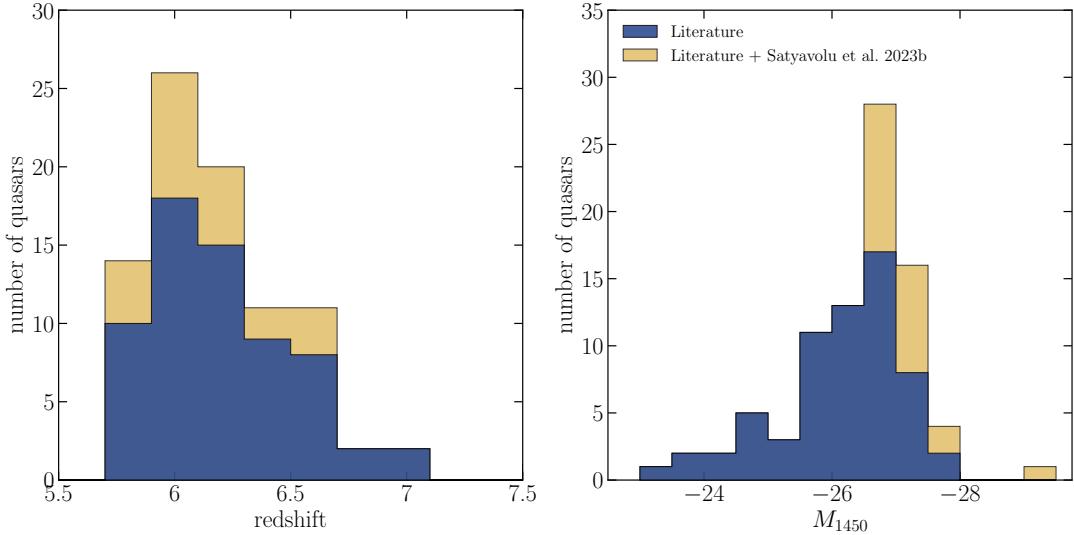
# New Measurements of Proximity Zone Sizes

Proximity zone sizes have so far been measured in 75 quasars between redshifts 5.7 and 7.5. Fan et al. (2006b) were the first to define and measure proximity zones for 16 SDSS $z \sim 6$ quasars. They also defined the luminosity-scaled proximity zones, where the measured proximity zones were corrected to the values that would be measured if all quasars were at a magnitude of $M_{1450} = -27.0$. They found that the luminosity-scaled proximity zone sizes decrease with increasing redshift, and attributed the decline to the evolution of the neutral hydrogen fraction in the IGM at those redshifts. Bolton & Haehnelt (2007b) measured the proximity zone sizes for four SDSS quasars in both Ly$\alpha$ and Ly$\beta$ forests and suggested that for a large enough sample, their ratio could be used to estimate the volume-averaged neutral fraction. Following the definition given by Fan et al. (2006b), proximity zones for quasars with redshifts $z > 5.7$ have since been measured by Willott et al. (2007), Mortlock et al. (2009), Willott et al. (2010), Carilli et al. (2010), Mortlock et al. (2011), Venemans et al. (2015), Reed et al. (2015), Eilers et al. (2017), Reed et al. (2017), Mazzucchelli et al. (2017), Bañados et al. (2018), Eilers et al. (2020), Ishimoto et al. (2020), and Bañados et al. (2021). The highest redshift quasar for which the proximity zone size has been measured is the redshift 7.54 quasar ULAS J1342+0928 (Bañados et al., 2018), with a proximity zone size of 1.3 pMpc. The quasars at $z = 7.085$ and 7.54 have proximity zone sizes that are three times smaller than the typical values at redshift $z \sim 6$. This is because these quasars show damped Ly$\alpha$ absorption by the intergalactic hydrogen. All of these proximity zone size measurements use similar methods, although they often differ in data

quality and some procedural details. For instance, all measurements exclude broad-absorption-line (BAL) quasars, as the outflow-induced broad absorption lines in these objects can bias the proximity zone size measurement (Eilers et al., 2020). The quasar continuum estimation methods are different in each of the measurements, but while this could lead to differences in the reported proximity zone sizes, Eilers et al. (2017) found that in practice the differences are negligible.

Interpretation of these proximity zone size measurements has led to interesting constraints on the properties of quasars and the IGM. Willott et al. (2007) estimated luminosity-scaled proximity zone sizes and found them to be relatively large (6.4 and 10.8 pMpc). Following Bolton & Haehnelt (2007b), they concluded that these quasars must be in an already ionized IGM with a neutral hydrogen fraction less than 0.3 at redshifts 6.1 and 6.43 respectively. Eilers et al. (2017) measured proximity zones of 30 quasars between $5.7 \lesssim z \lesssim 6.5$ and found a much shallower evolution of the luminosity-scaled proximity zone size as a function of redshift, unlike the previous measurements. They found that this evolution is independent of the IGM around the quasar, suggesting that contrary to previous analyses, the proximity zone size is set by the quasar properties and is relatively insensitive to the neutral hydrogen fraction of the IGM. Mazzucchelli et al. (2017) and Ishimoto et al. (2020) also found a shallow evolution of proximity zone sizes with redshift.

Eilers et al. (2017) also discovered 3 quasars with proximity zone sizes < 1 pMpc. After confirming that there is no truncation of the proximity zone size due to proximate absorbers or patchy neutral hydrogen islands, they concluded that these quasars must be young with lifetimes $t_q < 10^5$ yr. Such small proximity zones were also found by Reed et al. (2017), who measured proximity zones for four quasars. Two of their quasars showed small luminosity-corrected proximity zones, which they suggest could imply that the quasar is young with $< 10^7$–$10^8$ yr age, or that they are located in a region where the average hydrogen neutral density is a factor of 10 higher. Eilers et al. (2020) pre-selected and measured proximity zone sizes for 13 quasars, including two quasars from Reed et al. (2017) and one from Eilers et al. (2017), between $5.8 < z < 6.5$, that were likely to be young after ruling out spurious truncation of proximity zones. They conclude that 5 of their quasars are likely very young quasars with lifetimes $< 10^5$ yr. Such short quasar lifetimes have been found to be hard to reconcile with the estimates of the central supermassive black hole masses (Davies et al., 2019; Eilers et al., 2021).

**Figure 5.1:** Distribution of quasar redshifts (left panel) and UV magnitudes (right panel) for the 22 quasars studied in this thesis. The blue histograms show the distributions for the 65 quasars for which proximity zones have been measured previously, after excluding the 10 quasars for which we have updated the proximity zone size measurements in this thesis. The yellow histograms show the distributions for all 87 quasars for which proximity zones sizes are now available, including the 22 that have been measured in this thesis. .

Overall, the picture that emerges is that supermassive black holes spend a long time growing in an obscured phase (Satyavolu et al., 2023a) or undergo radiatively inefficient accretion at super/hyper-Eddington rates (Davies et al., 2019; Eilers et al., 2021). Increasing the sample size of proximity zone studies may therefore enable us to tighten the constraints on black hole growth.

In this thesis, we add 22 measurements to the above set of proximity zone size measurements using the XQR-30 sample. This is one of the largest set of proximity zone measurements based on homogeneous, high quality quasar spectra. We use the traditional definition of the proximity zone given by Fan et al. (2006b), and examine how the resultant proximity zone sizes correlate with the quasar luminosity, redshift, and black hole mass.

## 5.1 The XQR-30 survey

XQR-30 is an European Southern Observatory (ESO) Large Programme (ID: 1103.A-0817, P.I. V. D'Odorico) that targeted 30 quasars with redshifts between 5.8 and 6.6 using VLT/XSHOOTER (Vernet et al., 2011) to obtain high-resolution, high-SNR rest-frame UV spectra. The target quasars are some of the brightest quasars known in the southern hemisphere in this redshift range (D'Odorico et al., 2023b). The spectra were taken with slit widths of 0.9 arcsec and

| | Object | $z_{qso}$ | Line | $M_{1450}$ | Ref. (redshift) | Ref. (magnitude) |
|---|---|---|---|---|---|---|
| 1 | J0408-5632 | $6.033^{+0.0107}_{-0.006}$ | Mg II | $-26.56$ | Bischetti et al. (2022) | Reed et al. (2017) |
| 2 | PSOJ029-29 | $5.976^{++0.0106}_{-0.006}$ | Mg II | $-27.32$ | Bischetti et al. (2022) | Bañados et al. (2016) |
| 3 | ATLASJ029-36 | $6.013^{+0.0106}_{-0.006}$ | Mg II | $-27.00$ | Bischetti et al. (2022) | Bañados et al. (2016) |
| 4 | VDESJ0224-4711 | $6.525^{+0.0114}_{-0.0064}$ | Mg II | $-26.98$ | Bischetti et al. (2022) | Reed et al. (2017) |
| 5 | PSOJ060+24 | $6.17^{+0.0109}_{-0.0061}$ | Mg II | $-26.95$ | Bischetti et al. (2022) | Bañados et al. (2016) |
| 6 | PSOJ108+08 | $5.9647\pm0.0023$ | C II | $-27.59$ | Bosman et al. (in prep.) | Bañados et al. (2016) |
| 7 | SDSSJ0842+1218 | $6.0754\pm 0.0024$ | C II | $-26.91$ | Schindler et al. (2020) | Bañados et al. (2016) |
| 8 | PSOJ158-14 | $6.0687\pm 0.0024$ | C II | $-27.32$ | Bosman et al. (in prep.) | Bañados et al. (2023) |
| 9 | PSOJ183-12 | $5.893^{+0.0105}_{-0.0059}$ | Mg II | $-27.49$ | D'Odorico et al. (2023b) | Bañados et al. (2016) |
| 10 | PSOJ217-16 | $6.1466\pm 0.0024$ | C II | $-26.94$ | Bosman et al. (in prep.) | Bañados et al. (2016) |
| 11 | PSOJ242-12 | $5.8468\pm 0.0023$ | C II | $-26.92$ | Bosman et al. (in prep.) | Bañados et al. (2016) |
| 12 | PSOJ308-27 | $5.799^{+0.0103}_{-0.0058}$ | Mg II | $-26.78$ | D'Odorico et al. (2023b) | Bañados et al. (2016) |
| 13 | PSOJ323+12 | $6.5872\pm0.0025$ | C II | $-27.07$ | Schindler et al. (2020) | Mazzucchelli et al. (2017) |
| 14 | PSOJ359-06 | $6.1719\pm 0.0024$ | C II | $-26.79$ | Schindler et al. (2020) | Bañados et al. (2016) |
| 15 | SDSSJ0927+2001 | $5.7722\pm0.0023$ | CO | $-26.76$ | Wang et al. (2010) | Bañados et al. (2016) |
| 16 | SDSSJ0818+1722 | $5.967^{+0.0105}_{-0.0059}$ | Mg II | $-27.52$ | D'Odorico et al. (2023b) | Bañados et al. (2016) |
| 17 | SDSSJ1306+0356 | $6.033\pm0.0023$ | C II | $-27.15$ | Decarli et al. (2018) | Nanni et al. (2017) |
| 18 | ULASJ1319+0950 | $6.1347\pm0.0024$ | C II | $-27.05$ | Venemans et al. (2020) | Bañados et al. (2016) |
| 19 | SDSSJ1030+0524 | $6.309^{+0.0111}_{-0.0062}$ | Mg II | $-26.99$ | Jiang et al. (2007) | Bañados et al. (2016) |
| 20 | SDSSJ0100+2802 | $6.3269\pm0.0024$ | C II | $-29.14$ | Wang et al. (2016) | Bañados et al. (2016) |
| 21 | ATLASJ025-33 | $6.3373\pm0.0024$ | C II | $-27.50$ | Decarli et al. (2018) | Carnall et al. (2015) |
| 22 | PSOJ036+03 | $6.5405\pm0.0025$ | C II | $-27.33$ | Venemans et al. (2020) | Bañados et al. (2016) |

**Table 5.1:** Properties of the 22 quasars studied in this thesis (Satyavolu et al., 2023b). The columns show the serial number, quasar name, quasar redshift with the total $1\sigma$ uncertainty, the emission line used for determining the quasar redshift, quasar absolute UV magnitude at 1450 Å, and references for the quasar redshift and magnitude.

0.6 arcsec, nominal resolution $R \sim 8900$ and 8100, and median resolution of $R \sim 11400$ and 9800 in the visible (VIS) and near-infrared (NIR) arms of XSHOOTER, with pixel size of 10 km/s in both arms (Resolution, however, is not a deciding factor in our analysis, since we smooth the spectra by a 20 Å boxcar for obtaining the proximity zone size). The observing time on target ranged from 4 h to 11 h. The median SNR per pixel in the rest-frame 1600–1700 Å wavelength range is between 25 and 160 for spectra rebinned to 50 km/s. Data reduction, which includes optimal sky subtraction, telluric absorption correction, optimal extraction and direct combination of exposures, was done using a custom IDL pipeline developed for the XQ-100 survey (Becker et al., 2019) with minor improvements, mainly for the NIR arm. Further details about data reduction are discussed by D'Odorico et al. (2023b). We also include 12 archival VLT/XSHOOTER spectra in our sample, that, together with the 30 XQR-30 quasars, form the enlarged XQR-30 sample. These have similar redshifts, magnitudes, SNR, and comparable spectral resolution as the XQR-30 sample. The data reduction for these additional quasars was done with the same pipeline that was used for the XQR-30 sample. The full sample is described in Bosman et al. (2022) and D'Odorico et al. (2023b).

Of the 42 quasars in the enlarged XQR-30 sample, we use 22 in this study. We exclude 12 quasars that show strong broad absorption lines (BALs; Bischetti et al., 2022) and 7 quasars with proximate damped Ly$\alpha$ systems (pDLAs; Davies et al., 2023; Bañados et al., 2019).

We exclude BAL quasars because their proximity zones may be affected by unseen strong N V associated absorption. Proximate damped Lyman-$\alpha$ systems are absorption systems with neutral hydrogen column density $N_{\text{HI}} > 2 \times 10^{20} \text{cm}^{-2}$ at a velocity separation $\Delta v < 3000 \text{km s}^{-1}$ from the quasar (Prochaska et al., 2008). pDLAs can prematurely truncate the quasar flux, leading to spuriously small proximity zones. We exclude all quasars with pDLAs at a velocity separation $\Delta v < 5000 \text{km s}^{-1}$ from the quasar, that have been identified by the presence of neutral oxygen tracing the neutral hydrogen or by their associated ionised absorbers (Davies et al., 2023, Sodini et al. in preparation). They are also not modelled in our simulations, making them not suitable for comparison. We also exclude the heavily reddened quasar J1535+1943, which is most likely obscured (Yang et al., 2021). The large error on the systemic redshift of this quasar makes a reliable measurement of its proximity zone size difficult.

We obtain the normalised transmitted flux by fitting continuum spectra redward of the quasar's Ly$\alpha$ line using the log-PCA approach of Davies et al. (2018d), as described in Chen et al. (2022) and Bosman et al. (2022). This method improves upon the original PCA-based continuum fitting introduced by Suzuki (2005) and Pâris et al. (2011). We note however that the choice of continuum fitting method has been found to have a negligible impact on the proximity zone size measurement (Eilers et al., 2017).

### 5.1.1 Quasar redshifts and magnitudes

Table 5.1 summarises the redshifts and magnitudes of the 22 quasars in our study. Accurately measuring the redshifts of these quasars is difficult but also necessary for accurate estimates of the proximity zone sizes. For 13 of the 22 quasars, we use redshifts determined from the emission lines due to the transitions of CO or C II from the host galaxy (Decarli et al., 2018; Wang et al., 2010, Bosman et al. in preparation). We assign an uncertainty to this redshift measurement of $\Delta v \sim 100$ km/s, corresponding to blueshift of the emission line from the quasar's systemic rest-frame. The uncertainty associated with the fit to the emission lines is negligible. For the remaining 9 quasars, we use the redshifts measured from the quasar's Mg II emission line (D'Odorico et al., 2023b; Bischetti et al., 2022), with a typical associated uncertainty of $\Delta v \sim 391$ km/s (Schindler et al., 2020).

The absolute magnitude at 1450 Å ($M_{1450}$) is measured from the apparent magnitude $m_{1450}$,
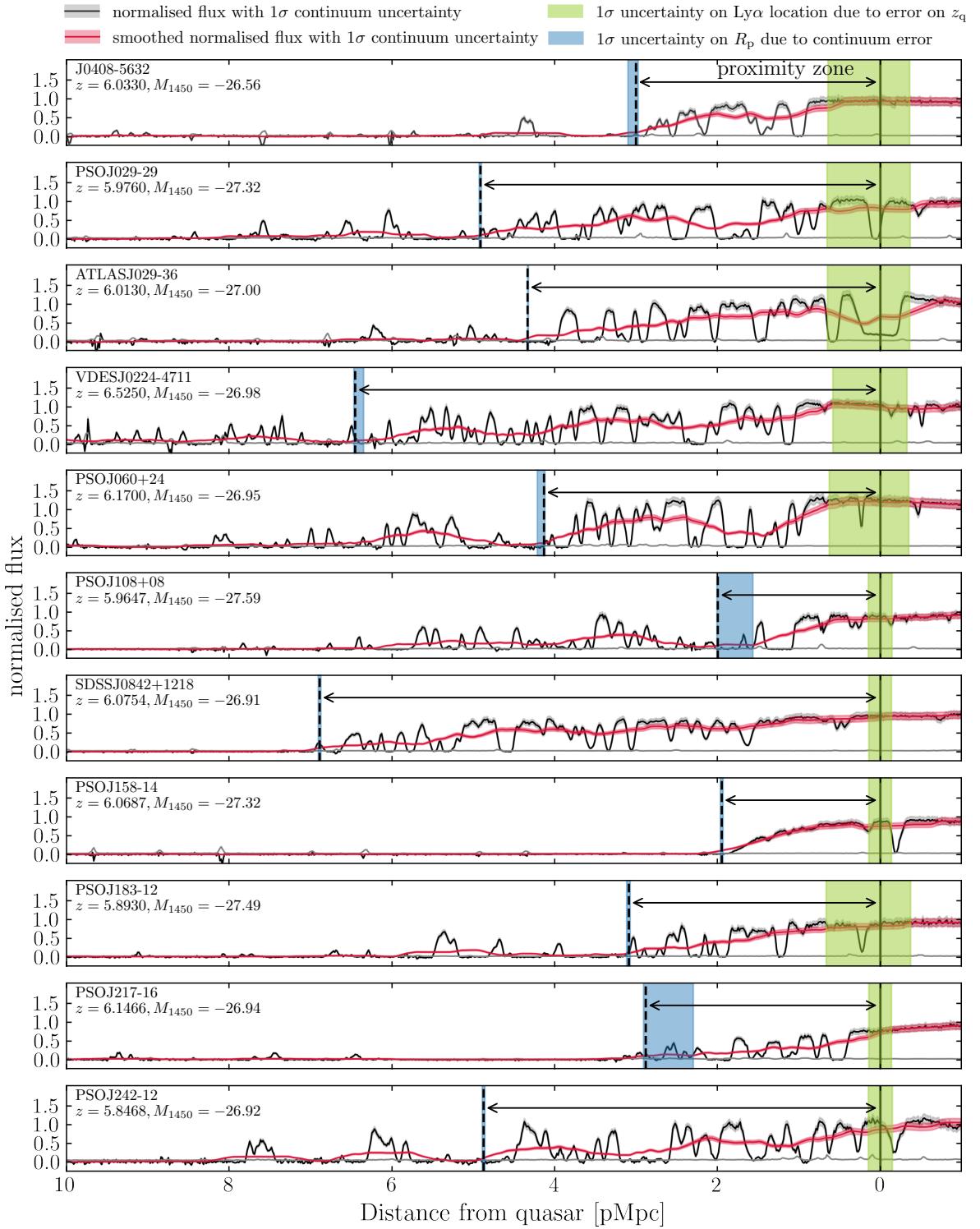
which is obtained by extrapolating the magnitude in the $y_{P1}$ or $J$ bands, depending on where contamination due to emission lines is lower, using a power law shape for continuum with spectral index $\alpha = -0.3$ (Bañados et al., 2016). The references for absolute magnitudes for each of the quasars are listed in Table 5.1.

Figure 5.1 shows the distribution of redshifts and magnitudes of quasars for which proximity zones have been previously measured (see Introduction of this Chapter) and our addition to this distribution. Our sample significantly increases the number of proximity zone sizes measured for quasars with redshifts $5.9 < z < 6.1$ and with magnitudes $-27.5 < M_{1450} < -26.5$.
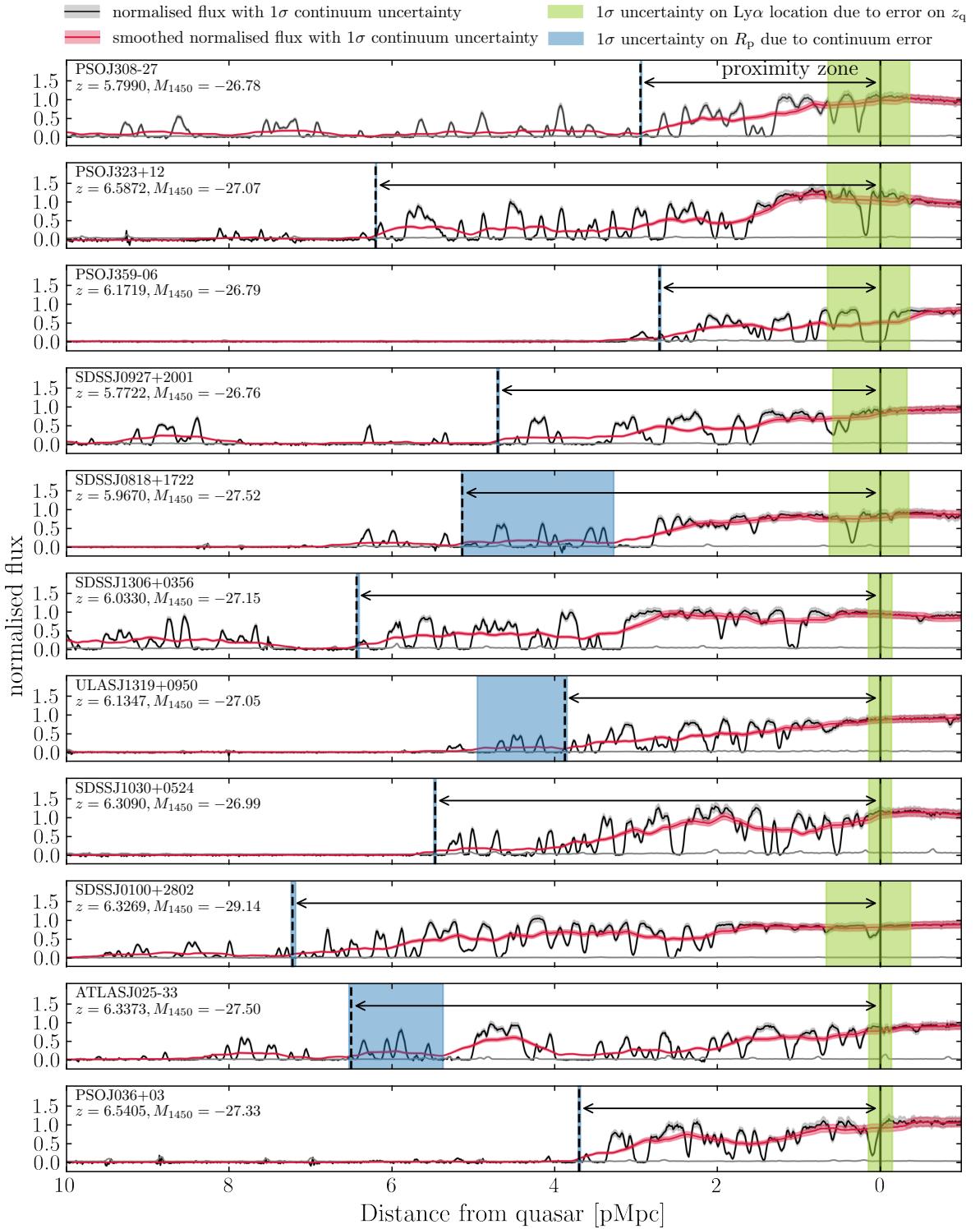
## 5.2   Proximity zone sizes of the E-XQR30 sample

To measure the proximity zone sizes of the quasars in our sample, we follow the convention introduced by Fan et al. (2006b). We smooth the continuum-normalised flux of each quasar by a 20 Å boxcar in the observer's frame, and locate the pixel with redshift $z_{GP}$ at which the smoothed normalised flux first drops below 0.1. The proximity zone size $R_p$ is then calculated by dividing the comoving line of sight distance between $z_{qso}$ and $z_{GP}$ by $(1+z_{qso})$ to obtain the proper distance. Figures 5.2 shows the resulting proximity zones. Table 5.2 lists the proximity zone sizes. Figure 5.3 shows their distribution.

Figure 5.2 shows the spectra and corresponding proximity zones for all the quasars in our sample. The red curves show the smoothed spectrum with shaded regions showing $1\sigma$ spread due to continuum uncertainties. Instrumental noise on the spectrum is negligible and hence we do not propagate this error onto the proximity zone size. Following Eilers et al. (2017), the error on the proximity zone size due to redshift uncertainty is calculated as $\Delta R_p = \Delta v/H(z)$, where $\Delta v$ is the redshift uncertainty in velocity units. The quasars in our sample have $\Delta v = 100$ km/s (for [C II] redshifts) which corresponds to an uncertainty of $\Delta R_p \sim 0.14$ pMpc in the proximity zone size at redshift 6. The uncertainty is larger for quasars with Mg II redshifts, with a median value of $\Delta R_p \sim 0.5$ pMpc. The continuum errors are computed by measuring the proximity zone sizes of the $1\sigma$ upper and lower bounds of the continuum-normalised flux using the same definition. For most of our quasars, the redshift uncertainty errors dominate over the continuum uncertainty errors on the proximity zone sizes, as shown in Table 5.2. All previous analyses are thus justified in neglecting the continuum errors. The largest error on $R_p$ due to continuum

**Figure 5.2:** Proximity zones of the quasars in our sample. The normalised flux obtained by dividing measured flux by continuum, is shown in black. Red curves show the smoothed spectra with shaded region showing the $1\sigma$ uncertainty in the continuum. Black solid and dotted lines show the quasar location and the extent of proximity zones, respectively. The blue shaded regions show the $1\sigma$ uncertainty on proximity zone sizes due to continuum uncertainties. Green shaded regions show redshift errors as the uncertainty on the location of the expected Ly$\alpha$ emission of the quasar.

**Figure 5.2** – *continued*

uncertainties is observed in the archival quasar SDSSJ0818+1722 to be 1.86 proper Mpc ($\sim 36\%$ of the measured value), even though the $1\sigma$ uncertainty on the continuum is not significant. This is because the definition of proximity zone size is such that even though the smoothed flux is quite close to 0.1 due to the uncertainty of the continuum placement, $R_\mathrm{p}$ is not defined until the smoothed flux becomes equal to or less than 0.1. Likewise, even though the proximity zones of some quasars are of similar size (e.g PSOJ158-14 and PSOJ108+08), their flux outside the proximity zone size is quite different. In order to better constrain quasar lifetimes based on proximity zone sizes, we will study the use of multiple definitions for proximity zone sizes based on the flux threshold in future work (Satyavolu et al., in preparation).
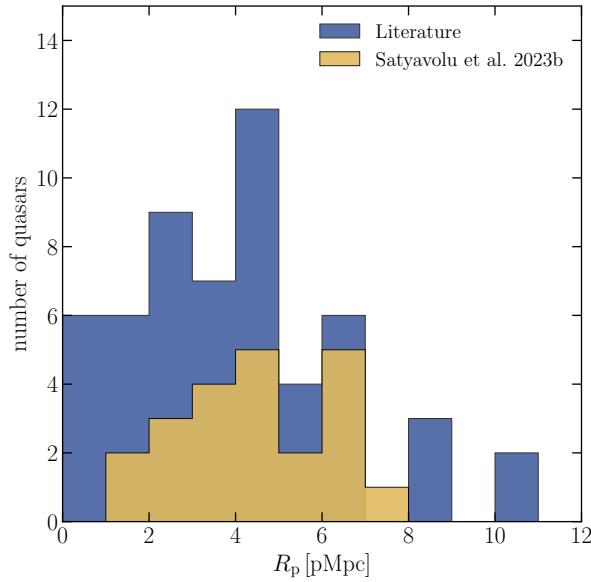
The total error on the proximity zone sizes of the quasars was obtained by adding the redshift and continuum uncertainty errors in quadrature. All the proximity zone measurements with their errors are shown in Table 5.2. Out of these, proximity zones were previously measured for ten quasars of our present sample. We have updated proximity zone measurements for the quasars PSOJ060+24, SDSSJ0100+2802, SDSSJ0818+1722, PSOJ036+03 (Eilers et al., 2017), PSOJ323+12 (Mazzucchelli et al., 2017), PSOJ158-14, PSOJ359-06 (Eilers et al., 2020) and SDSSJ0927+2001, ULASJ1319+0950, SDSSJ1030+0524 (Ishimoto et al., 2020) with the latest redshifts and X-SHOOTER spectra. The newer measurements differ from the older measurements by $\sim 1\%$ to not more than 5%. The minor differences are expected to be due to difference in redshifts. One quasar ULASJ1319+0950 is reported to have a proximity zone size of 4.99 pMpc from Ishimoto et al. (2020). Our updated measurement of 3.87 pMpc is closer to the value of 3.84 pMpc measured by Eilers et al. (2017).

Figure 5.3 shows the distribution of the proximity zone sizes of the enlarged XQR-30 sample. The largest and smallest proximity zones we measure are 7.22 and 1.95 pMpc, with a median around 5 pMpc. Also shown in blue is the distribution of all previously measured proximity zone sizes, not scaled to a fiducial quasar luminosity and excluding the ten quasars that have been updated in this thesis. Our proximity zone sizes are consistent with previous measurements, and add to the number of small proximity zone sizes ($R_\mathrm{p} < 2$ pMpc) measured in the literature recently.

82

| | Object | $R_p$ (pMpc) | Continuum error ($\Delta R_p$) | | Redshift error ($\Delta R_p$) | | Total error ($\Delta R_p$) | |
|---|---|---|---|---|---|---|---|---|
| | | | Lower $1\sigma$ (pMpc) | Upper $1\sigma$ (pMpc) | Lower $1\sigma$ (pMpc) | Upper $1\sigma$ (pMpc) | Lower $1\sigma$ (pMpc) | Upper $1\sigma$ (pMpc) |
| 1 | J0408-5632 | 3.00 | 0.03 | 0.10 | 0.36 | 0.65 | 0.37 | 0.66 |
| 2 | PSOJ029-29 | 4.91 | 0.01 | 0.01 | 0.37 | 0.66 | 0.37 | 0.66 |
| 3 | ATLASJ029-36 | 4.33 | 0.01 | 0.01 | 0.37 | 0.65 | 0.37 | 0.65 |
| 4 | VDESJ0224-4711 | 6.45 | 0.10 | 0.01 | 0.33 | 0.59 | 0.35 | 0.59 |
| 5 | PSOJ060+24* | 4.13 | 0.01 | 0.08 | 0.35 | 0.63 | 0.35 | 0.63 |
| 6 | PSOJ108+08 | 1.99 | 0.43 | 0.01 | 0.14 | 0.14 | 0.46 | 0.15 |
| 7 | SDSSJ0842+1218 | 6.89 | 0.01 | 0.01 | 0.14 | 0.14 | 0.14 | 0.14 |
| 8 | PSOJ158-14* | 1.95 | 0.01 | 0.01 | 0.14 | 0.14 | 0.14 | 0.14 |
| 9 | PSOJ183-12 | 3.09 | 0.01 | 0.03 | 0.38 | 0.67 | 0.38 | 0.67 |
| 10 | PSOJ217-16 | 2.88 | 0.58 | 0.03 | 0.14 | 0.14 | 0.6 | 0.14 |
| 11 | PSOJ242-12 | 4.87 | 0.01 | 0.01 | 0.15 | 0.15 | 0.15 | 0.15 |
| 12 | PSOJ308-27 | 2.95 | 0.01 | 0.01 | 0.38 | 0.68 | 0.38 | 0.68 |
| 13 | PSOJ323+12* | 6.20 | 0.01 | 0.01 | 0.13 | 0.13 | 0.13 | 0.13 |
| 14 | PSOJ359-06* | 2.71 | 0.01 | 0.01 | 0.14 | 0.14 | 0.14 | 0.14 |
| 15 | SDSSJ0927+2001* | 4.70 | 0.02 | 0.02 | 0.15 | 0.15 | 0.15 | 0.15 |
| 16 | SDSSJ0818+1722* | 5.13 | 1.86 | 0.01 | 1.43 | 1.43 | 2.35 | 1.43 |
| 17 | SDSSJ1306+0356 | 6.43 | 0.03 | 0.01 | 0.14 | 0.14 | 0.15 | 0.14 |
| 18 | ULASJ1319+0950* | 3.87 | 0.03 | 1.07 | 0.14 | 0.14 | 0.14 | 1.08 |
| 19 | SDSSJ1030+0524* | 5.47 | 0.01 | 0.01 | 0.34 | 0.61 | 0.34 | 0.61 |
| 20 | SDSSJ0100+2802* | 7.22 | 0.04 | 0.01 | 0.13 | 0.13 | 0.14 | 0.13 |
| 21 | ATLASJ025-33 | 6.50 | 1.13 | 0.03 | 0.13 | 0.13 | 1.13 | 0.14 |
| 22 | PSOJ036+03* | 3.70 | 0.01 | 0.01 | 0.13 | 0.13 | 0.13 | 0.13 |

\* Previously available measurements that have been updated in this thesis.

**Table 5.2:** Our proximity zone size measurements. Columns show the serial number, the name of the quasar, proximity zone size in proper Mpc with the continuum, redshift and total uncertainties. The minimum error on $R_p$ due to continuum uncertainties is the spatial resolution of the spectra, which is $\sim 0.01$ pMpc. Total error is obtained by adding the continuum and redshift errors in quadrature.
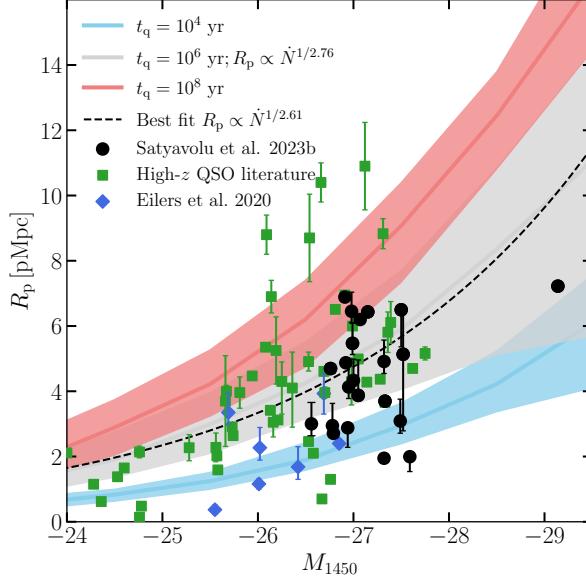


**Figure 5.3:** Distribution of proximity zone sizes reported in this thesis. The blue histogram shows the distribution of all previously available proximity zone sizes (Carilli et al., 2010; Eilers et al., 2017; Mazzucchelli et al., 2017; Reed et al., 2017; Bañados et al., 2018; Eilers et al., 2020; Ishimoto et al., 2020; Bañados et al., 2021), except those only available as values scaled to a fiducial luminosity, or that have been updated in this thesis. The yellow histogram shows the distribution of the 22 proximity zone sizes presented in this thesis.

### 5.2.1 Correlation with quasar luminosity

Figure 5.4 shows the distribution of $R_p$ as a function of quasar magnitude. It can be seen that although the quasars in our sample have very similar magnitudes, with mostly $-26.5 < M_{1450} < -27.5$, the proximity zone size distribution can vary considerably. The smallest proximity zone is found at a magnitude of $-27.51$ and the largest proximity zone at a magnitude of $-29.14$, both at similar redshifts of 6.06 and 6.32, respectively. Most of the measured values are consistent with earlier measurements at similar redshifts ($z \sim 6$).

   Also shown are the median proximity zone sizes and the $1\sigma$ distribution around the median values from our simulations (see Chapter 3), for a lightbulb quasar. The median proximity zone size increases with increase in quasar lifetime, as the longer the quasar is active, the farther its ionisation front can travel before reaching the equilibrium value. For brighter quasars, there is also an increase in the spread of the proximity zone size distribution before the quasar lifetime reaches the equilibration timescale. This can be understood as a consequence of the ionisation fronts traveling farther in a small enough time, and encountering neutral hydrogen islands along random directions. For a fainter quasar, the quasar will need more time for its ionisation front to travel farther and encounter such neutral islands. Therefore, fainter quasars see only their immediate surroundings, which are almost uniformly ionised at these redshifts and lifetimes, leading to a narrower spread. The $1\sigma$ spread is also the largest for $t_q \sim 10^6$ yr for similar reasons, as a younger quasar and an older quasar see a mostly ionised medium. The large proximity zones in our sample are consistent with the models of lightbulb quasars of age $\geq 1$ Myr. The smaller proximity zones with $R_p \lesssim 2$ pMpc appear to indicate a young lifetime of $\lesssim 10^4$ yr for a lightbulb quasar at a redshift $z \sim 6$. The fraction of such quasars with small proximity zones is 2 out of 22 or about $\sim 9$ percent in our sample, consistent with the fraction of 5–10% estimated by Eilers et al. (2020). We discuss the two smallest $R_p$ values in greater detail in Section 5.2.5 below.

   In order to study the correlation of proximity zone sizes with quasar magnitudes without being influenced by the redshift of the quasars, we obtain a best-fit curve to all measured proximity zone sizes (excluding the targeted sample of Eilers et al., 2020) including ours against their magnitudes for quasars with redshifts between $6 < z < 6.2$ assuming a power law between $R_p$ and $\dot{N}$. The redshift range was chosen such that the number of quasars for which proximity
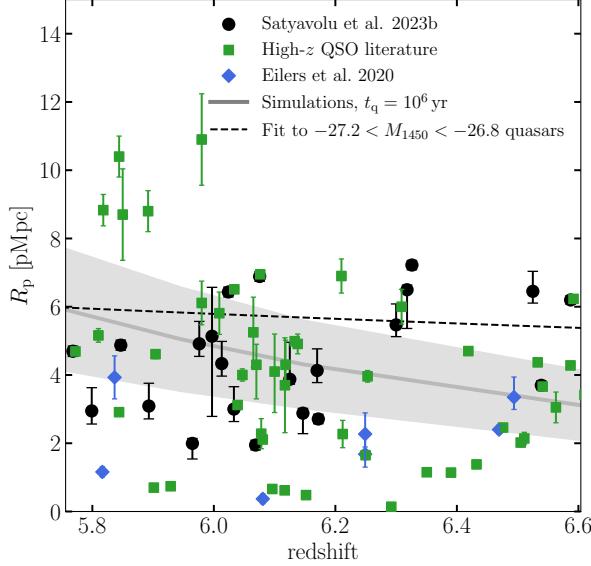
**Figure 5.4:** Proximity zone sizes as a function of quasar magnitude. Previous measurements are shown in green. The targeted sample of Eilers et al. (2020) is shown in blue. Our measurements are shown in black. The errors on our proximity zone sizes are due to both continuum and redshift uncertainties. The blue, grey and red curves are from our simulations for quasar ages of $10^4$, $10^6$ and $10^8$ yr at a redshift of 5.95. Shaded regions show 68% scatter across 500 sightlines from our simulations. The black dotted line shows the best fit curve to a relatively homogeneous subset of quasars with $6 < z < 6.2$, except quasars from the targeted sample of Eilers et al. (2020).

zone sizes are measured is maximized (see Figure 5.1). In a mostly uniform medium, the scaling follows $R_\mathrm{p} \propto \dot{N}^{1/3}$ while in a mostly ionised medium, $R_\mathrm{p} \propto \dot{N}^{1/2}$ (Bolton & Haehnelt, 2007a). Since at the redshifts of our quasars, the universe is most likely to be partly ionised and partly uniform, one could expect the scaling to fall between $R_\mathrm{p} \propto \dot{N}^{1/2}$ and $R_\mathrm{p} \propto \dot{N}^{1/3}$, depending on the redshift of the quasar. Our simulations find an evolution of $R_\mathrm{p} \propto \dot{N}^{1/2.76}$, for a quasar lifetime of 1 Myr and redshift 5.95. The best fit to all data within the redshift range $6 < z < 6.2$ shows an evolution of $R_\mathrm{p} \propto \dot{N}^{1/2.61}$, slightly steeper than the scaling inferred from our simulations, but consistent within the expected range for the scaling at this redshift.

### 5.2.2 Correlation with quasar redshift

The evolution of proximity zone sizes as a function of redshift encodes information about reionization (Satyavolu et al., 2023a). Models in which reionization ends later cause a 30% reduction in proximity zone sizes and increase the scatter in their distribution by 10%, as the growth of ionization fronts is impeded by neutral parts of the IGM.
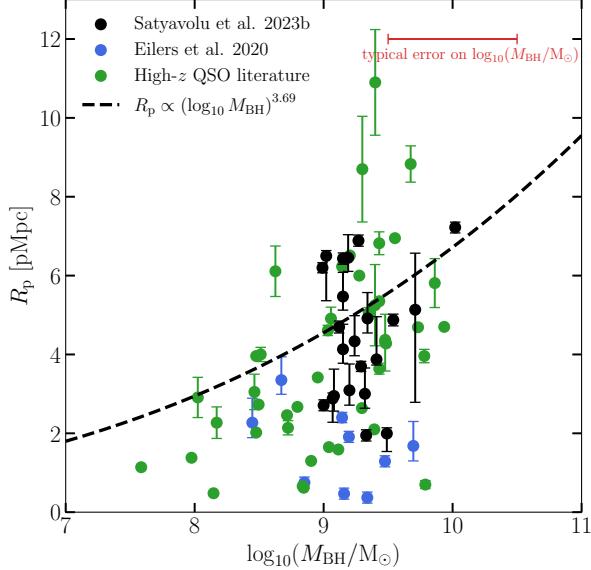
Figure 5.5 shows the proximity zone sizes from all measurements including those presented in this thesis. In order to study the evolution of proximity zone size with redshift, we fit to all measured proximity zone sizes (excluding the targeted sample of Eilers et al., 2020) including

**Figure 5.5:** Evolution of proximity zone sizes. Older measurements are shown in green. The targeted sample of Eilers et al. (2020) is shown in blue. Our measurements are shown in black. Also shown are the simulated proximity zones for a quasar of magnitude $-27$ and age of 1 Myr across different redshifts. The shaded region shows 68% scatter across 500 sightlines in our simulation. The black dotted line shows the best fit curve $R_\mathrm{p} \propto (1+z)^{-0.89}$ to our measurements and previous measurements excluding Eilers et al. (2020). For obtaining the best fit, only a relatively homogeneous subset of quasars, with $-26.8 < M_{1450} < -27.2$ was used.

ours for a relatively homogeneous subsample of quasars with magnitudes between $-26.8$ and $-27.2$, assuming a power law between $R_\mathrm{p}$ and $(1+z)$. Unlike previous analyses, we do not correct the proximity zones to a common luminosity to get a best-fit. This is because the scaling between proximity zone sizes and magnitude is strongly dependent on the redshift of the quasar, and the same scaling cannot be applied to all quasars. Moreover, different measurements use a different scaling to obtain the luminosity-corrected proximity zones, which makes them unsuitable for comparison.

We find a very shallow trend of $R_\mathrm{p} \propto (1+z)^{-0.89}$, shallower than the previous inferences that were made through the luminosity-scaled proximity zones. This trend suggests that the scatter in the proximity zone sizes for similar magnitude quasars, as seen in Figure 5.4, is more likely due to the differences in their lifetimes. Indeed, one can notice that two of the farthest quasars with $z > 6.5$ have larger than average proximity zone sizes, with an average luminosity. Although the universe is more neutral at higher redshifts, such large proximity zones can be explained by longer quasar lifetimes. Smaller proximity zones are in fact found close to the smallest redshifts in the sample, which could have suggested either large scatter in the ionization state between sightline to sightline or smaller quasar lifetimes, although the latter seems to be favoured by our simulations.
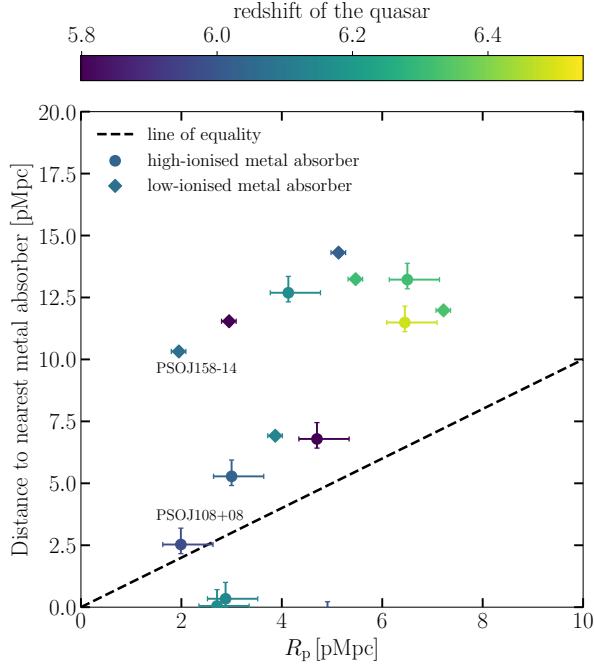
**Figure 5.6:** Proximity zone sizes as a function of supermassive black hole mass. Previous measurements for which black hole masses were available are shown in green and the targeted sample from Eilers et al. (2020) is shown in blue. Our measurements are shown in black. Our black hole masses are from Mazzucchelli et al. (in preparation). The typical error on the black hole masses is represented by the errorbar at the top right in red. All black hole masses are based on Mg II linewidths. The black dotted line shows the best fit curve to our measurements and previous measurements excluding Eilers et al. (2020). For obtaining the best fit, a relatively homogeneous subset of quasars with $-26.8 < M_{1450} < -27.2$ and $6.0 < z < 6.2$ was used. A power-law relationship was assumed between the quasar proximity zone size and logarithm of the black hole mass, as motivated in the text.

### 5.2.3 Correlation with SMBH mass

Proximity zone sizes are sensitive to the quasar lifetime (Eilers et al., 2017; Davies et al., 2020; Eilers et al., 2021; Morey et al., 2021). As a result, combining proximity zone sizes with black hole mass measurements can potentially constrain the growth history of black holes (Satyavolu et al., 2023a). With this in mind, Figure 5.6 shows the proximity zone sizes of quasars in our sample against the masses of their central SMBHs. The black hole masses for XQR-30 quasars were measured by Mazzucchelli et al. (in preparation), based on the Mg II and C IV linewidths, which can be used to derive the velocity of the gas clouds in the broad-line region and thereby the dynamical mass of the black hole, otherwise called the single-epoch viral black hole mass. The black hole masses have a typical total uncertainty of 1 dex (Vestergaard & Osmer, 2009). The black hole masses of our sample are of the order $\sim 10^9 M_\odot$, consistent with the other measurements at this redshift for comparable UV magnitudes (Shen et al., 2019; Yang et al., 2021; Farina et al., 2022).

For comparable UV magnitudes and redshifts, we expect the proximity zone sizes to increase with quasar lifetime, as in Equation (2.3.6). In an exponential growth model for the supermassive black hole, the black hole mass $M_{BH}$ would be proportional to $\exp(t_q)$. We therefore try to fit

**Figure 5.7:** Distance to the nearest metal absorber as a function of proximity zone size. High-ionised metal absorbers are shown as circles while low-ionised metal absorbers are shown as diamonds. Colours represent the redshift of the metal absorber. PSOJ108+08 is the only quasar in our sample with a metal-line absorber close to the edge of the proximity zone. We also highlight PSOJ158–14 on this figure; this quasar is discussed in greater detail in Figure 5.8.

a power law relation between the proximity zone size $R_p$ and the logarithm of the black hole mass, $\log_{10} M_{BH}$, for a relatively homogeneous subset of quasars, with magnitudes $-26.8 < M_{1450} < -27.2$ and redshifts $6 < z < 6.2$. We find a strong correlation of the proximity zone size with the black hole mass as $R_p \propto \log_{10}(M_{BH})^{3.69}$, stronger than what is expected from Equation (2.3.6), which is valid only for lifetimes less than the equilibration timescale. This correlation is also stronger than what was inferred by Ishimoto et al. (2020). We plan to look for black hole growth models that are consistent with both the proximity zone sizes and black hole masses using simulations in a future work (Satyavolu et al., in preparation).

### 5.2.4 Correlation of proximity zones with closeness to metal absorption systems

Figure 5.7 shows the quasars in our sample for which the distance to the nearest metal absorber is within 20 pMpc. Highly ionized absorbers are shown as circles while low-ionized systems are shown as diamonds. Quasars with pDLAs and BALs are excluded from this sample. The ionised absorbers were identified by looking for absorption in additional transition lines corresponding

to each ion through an automated search and visual inspection (Davies et al., 2023). It can be seen that high and low ionization absorbers are found at both high and low redshifts in our sample.

We find that quasar proximity zones fall into three categories. At the bottom of the plot, there are two quasars with relatively small proximity zones (2–3 pMpc) that house high ionization metal absorbers. For most quasars as seen in the top half of the plot, the closest metal absorption system sits beyond 10 pMpc from the quasar, well outside proximity zone boundary. For the quasars in our sample, there appears to be a strong correlation between proximity zone size and the presence of metal absorbers. This could potentially be an effect of the quasar's ionizing radiation on the metal-line chemistry around it. Low ionization metal absorbers, which may have more potential to truncate proximate zones, are found to cover the whole range of proximity zone sizes from 2 to 7 pMpc. There are three proximity zones from 2 to 5 pMpc whose quasar lines of sight contain metal absorbers just outside the boundary of their proximity zones at a distance of 2.5 to 7 pMpc. Only one quasar, PSOJ108+08, contains a metal absorber right at the edge of the proximity zone. The lifetime of this quasar could be potentially underestimated as the proximity zone appears to be prematurely truncated.

### 5.2.5 Anomalously small proximity zones

Two quasars in our sample show particularly small proximity zones, with $R_p < 2$ pMpc. These quasars are also at the brighter magnitude and lower redshift end of the range spanned by our sample, making it hard to explain the small proximity zone sizes without invoking a young quasar age. While we leave a deeper investigation of these proximity zones for future work (Satyavolu et al., in preparation), we make some preliminary remarks here.
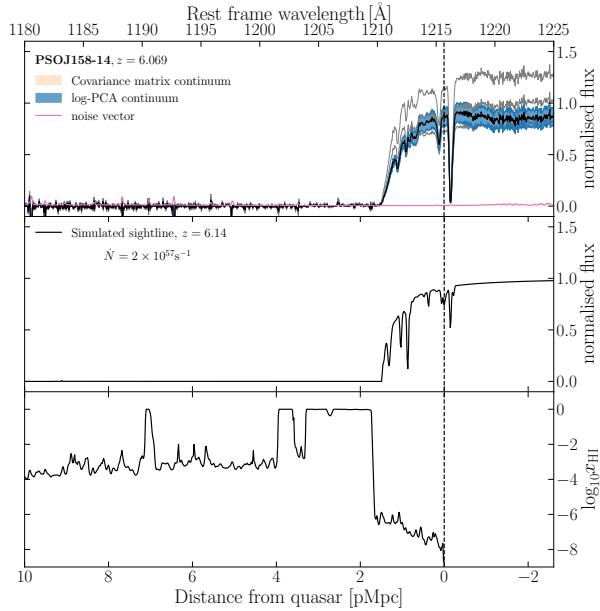
**PSOJ158-14**

The quasar PSOJ158-14 is at a redshift of 6.0687 with a magnitude of −27.32. The proximity zone size of this quasar is 1.95 pMpc. Eilers et al. (2020) have investigated this quasar and reported that it has a large star formation rate ($\sim 1420 M_\odot \, \text{yr}^{-1}$), large bolometric luminosity ($\sim 10^{47}$ erg s$^{-1}$), high Eddington ratio ($\lambda_{\text{edd}} \sim 1$), and shows signs of strong internal motions within the broad line region. They also point out the dust continuum emission of this quasar is

very strong ($F_{\mathrm{cont}} \sim 3.46$ mJy).

Figure 5.8 shows the continuum normalised spectrum of this quasar close to its Ly$\alpha$ line. We see that the spectrum blueward of the Ly$\alpha$ line resembles a damping wing. Additionally, the flux redward of the Ly$\alpha$ line shows attenuation, as one would expect in the presence of a damping wing. The flux continues to remain attenuated till 1233 Å. Interestingly, there is no evidence of a compact high-column-density absorber ahead of the quasar. The nearest metal absorber is at a redshift of 5.89874 (Davies et al., 2023), which is well outside the edge of the proximity zone (at $\sim 10$ pMpc from the quasar; see Figure 5.7). This suggests that if the feature around the Ly$\alpha$ line of PSOJ158-14 is indeed a damping wing, it is likely to be caused by a neutral hydrogen 'island' in the IGM. Indeed, we do find similar sightlines in our simulation for comparable redshift and quasar brightness. An example for $z = 6.14$ and $M_{1450} = -27$ is shown in the middle panel of Figure 5.8. This simulated sightline has a clearly visible damping wing, caused by a large neutral hydrogen patch in the IGM, which can be seen in the bottom panel of Figure 5.8. For a quasar age of 1 Myr, only 1 of 500 simulated sightlines shows a damping wing. For larger quasar lifetimes, this incidence drops. For a quasar lifetime of 10 Myr, none of the simulated sightlines show a damping wing. For a flickering lightcurve quasar, this number could be larger (Satyavolu et al., 2023a).

However, an IGM damping wing interpretation of the spectrum of PSOJ158-14 is less than certain. Several aspects of this spectrum complicate its analysis. For example, Figure 5.8 also shows the continuum normalised flux for this quasar for a different continuum reconstruction, based on the covariance matrix method of Greig et al. (2017a). We see that with this continuum, although the spectral shape still resembles a damping wing, the flux redward of the Ly$\alpha$ does not appear to be attenuated. Furthermore, when compared to the noise vector shown in Figure 5.8, the spectrum of PSOJ158-14 reveals flux just blueward of the proximity zone, suggesting that the damping-wing-like absorption might be not caused by the IGM. While the evidence for this extended flux is relatively weak, the spectrum appears to have a statistically significant spike in flux at around $\sim 2$ pMpc from the edge of the proximity zone. These considerations suggest that perhaps the spectrum is a result of absorption by a metal-poor absorber instead of the IGM. In this scenario, the proximity zone size could be the result of a small quasar lifetime of $< 10^4$ yr, and the flux blueward of the proximity zone could be explained by residual flux from partial covering of the quasar continuum, or weak Ly$\alpha$ emission from the absorber.

**Figure 5.8:** *Top panel*: Continuum-normalised spectrum of PSOJ158-14, for two continuum reconstruction methods, the log-PCA method (Chen et al., 2022) and the covariance matrix method (Greig et al., 2017a), shown in blue and orange, respectively. Shaded regions show the $1\sigma$ spread around the median value. (We use the log-PCA method for all quasars in this thesis.) *Middle panel*: A simulated spectrum showing an IGM damping wing at $z = 6.14$ for a quasar with magnitude $-27$ and age 1 Myr. *Bottom panel*: The ionised hydrogen fraction along the same simulated sightline. This reveals the neutral hydrogen regions that create the damping wing seen in the middle panel. At redshift 6.14, only one of 500 sightlines in our simulation shows this feature.

More data seem to be necessary to rule out an IGM damping wing for this quasar. But if confirmed, PSOJ158-14 would be an interesting exception to the finding by Fan et al. (2023) that a quasar with both small proximity zone and damping wing has not be found below redshift 7 so far.

**PSOJ108+08**

The quasar PSOJ108+08 is at a relatively lower redshift of 5.9647 with a magnitude of $-27.59$. This quasar has the second smallest proximity zone size in our sample, with $R_{\rm p} = 1.99$ pMpc. As we see in Figure 5.2, the spectrum of this quasar does not show a damping wing. Although the proximity zone size is small, the flux blueward of the Ly$\alpha$ line extends all the way up to $\sim 6$ pMpc i.e., nearly 3 times the proximity zone size and immediately increases above our 10% threshold beyond the proximity zone. We find a high-ionization metal absorber at 2.53 pMpc from this quasar, indicating that the proximity zone might be prematurely truncated due to absorption of the quasar flux by this absorber.

PSOJ108+08 suggests that to better estimate lifetimes in such quasars, it might be worthwhile to explore alternate definitions for the proximity zone, such as defining the proximity zones as

points where the flux transmission is at 5% as well as 20% and changing the smoothing length, which we will explore in future work (Satyavolu et al. in preparation).

## 5.3 Final remarks about the XQR-30 quasar proximity zones

We measured proximity zone sizes of 22 quasars at redshifts between 5.8 and 6.5 and UV magnitudes $M_{1450}$ between $-26$ and $-29$ using high-SNR spectra obtained with the X-SHOOTER instrument on the VLT telescope. Of the 22 quasar spectra that we study, 14 were obtained as part of the XQR-30 survey. The other eight quasars were obtained with X-SHOOTER from previous programs and were chosen to have similar resolution and SNR to the XQR-30 spectra. We summarize our results below:

- The proximity zone sizes of our quasars range from 1.95 to 7.22 pMpc. This roughly corresponds to quasar lifetimes of $10^4$ to $10^8$ yr in the lightbulb model. About 9% of our measured proximity zones are small, requiring lifetimes of less than $10^4$ yr. This distribution of proximity zone sizes is consistent with previous measurements of quasars with similar magnitudes and redshifts. This thesis increases the number of available proximity zone size measurements at $z > 5.7$ to 87.

- We update the proximity zone size measurements of 10 quasars previously studied in the literature, with the help of updated spectra and redshifts. The new measurements are consistent with previous measurements within 1–5%.

- We infer a scaling of proximity zone size with UV magnitude based on all measurements for quasars within the redshift range $6 < z < 6.2$ and find it to be consistent with our expectations from simulations. This scaling is shallower than what was measured previously (Ishimoto et al., 2020).

- We infer a scaling of proximity zone size with redshift based on all measurements for quasars with magnitudes $-27.2 < M_{1450} < -26.8$ and find it to be shallower than what was measured from previous analyses (Eilers et al., 2017; Mazzucchelli et al., 2017; Ishimoto et al., 2020). The shallowness of this scaling suggests that the scatter in the proximity zone sizes for quasars of similar UV magnitudes is a result of variation in quasar lifetimes.

92

- We infer a scaling of proximity zone size with black hole mass and find it to be steeper than what is expected from theory. Previously, Ishimoto et al. (2020) reported little to no correlation between $R_{\rm p}$ and black hole mass.

- Two of our quasars have exceptionally small $R_{\rm p}$ of less than 2 pMpc. One of these quasars shows possible signatures of a damping wing produced by the intergalactic medium or an extremely metal-poor foreground galaxy. Another has a high-ionized metal absorber close to the edge of the proximity zone.

Our measurements of proximity zone sizes, and their correlations with quasar brightness, redshift, and black hole mass point towards a diverse range of quasar lifetimes. The overall picture remains consistent with our previous finding that proximity zone size measurements seem to support a scenario in which supermasive black holes at high redshifts undergo obscured growth (Satyavolu et al., 2023a).

# Chapter 6

# Effect of Quasar Proximity Zones on the Direct Measurement of the LyC Mean Free Path at $z \sim 6$

The epoch of reionization marks the era during which the neutral hydrogen in the Universe became largely ionized, most likely due to UV photons from stars in galaxies across the cosmic volume, with potentially an additional minor contribution from quasars. To understand how reionization happened, we need to know how the photons that escaped from the various sources of radiation interacted with the intergalactic medium (IGM) to ionize the universe. An important characteristic of the propagation of these ionizing photons is determined by their mean free path (MFP), defined as the average distance a photon travels before getting absorbed (Rybicki & Lightman, 1985). In a homogeneous IGM, the MFP increases with the background UV photoionization rate, as this lowers the IGM opacity. More generally, the MFP $\lambda_{\mathrm{MFP}}$ varies with the background photoionization rate $\Gamma_{\mathrm{bg}}$ as a power law (Miralda-Escudé et al., 2000), and, furthermore, both of these quantities can be expressed as functions of the cosmological emissivity $\epsilon$ of ionizing radiation. Knowing the MFP therefore puts constraints on the ionizing sources as well as the sinks of ionizing photons. At higher redshifts, the rate of increase of the MFP with redshift indicates how rapid the progress of reionization is. Measurement of the MFP of hydrogen-ionizing photons is thus important for characterizing reionization.

During reionization, the sources of radiation carve out regions of ionized hydrogen in the

IGM around themselves. These ionized regions later coalesce, at which point a background of ionizing radiation gets established throughout the Universe (Gnedin & Kaurov, 2014; Bauer et al., 2015). This is considered to be the end of reionization and is thought to occur at $z \sim 5.3$ (Kulkarni et al., 2019; Bosman et al., 2022; Zhu et al., 2022). At lower redshifts, in the post-reionization Universe, the MFP is dominated by the residual neutral hydrogen systems in the otherwise fully ionized IGM. These systems retain neutral hydrogen due to self-shielding thanks to their high density. These appear as high-column-density absorbers in the spectra of background sources. The corresponding MFP is then set by the average spacing between such absorbers, obtained by counting the number of optically-thick absorbers or Lyman-limit systems (LLS) along the lines of sight in quasar spectra (Storrie-Lombardi et al., 1994; O'Meara et al., 2007, 2013). At higher redshifts, before reionization has ended, however, the boundaries of the ionized regions also play an important role in setting the MFP. At these times, the MFP is the average distance between LLSs or the typical size of ionized regions, depending on which is smaller (Shukla et al., 2016; Madau, 2017).

Observationally, measurements of the MFP have been inferred using direct and indirect means. One way to obtain a direct measurement of the MFP is to use stacked quasar spectra bluewards of rest-frame 912Å. The MFP is then computed as the distance at which the effective Lyman-continuum optical depth becomes unity (Prochaska et al., 2009; Fumagalli et al., 2013; O'Meara et al., 2013; Worseck et al., 2014). This has resulted in a measurement of $\lambda_{\mathrm{mfp}} = 10.3 \pm 1.6$ pMpc at $z = 5.16$, with the MFP increasing as $(1+z)^{-5.4}$ down to $z = 4.56$. At higher redshifts, the MFP can however become comparable or smaller than the size of the proximity zone of the quasars whose spectra are used in the stacks. This biases the MFP towards lower values (Worseck et al., 2014; D'Aloisio et al., 2018). Becker et al. (2021, B21) reported a measurement of the MFP in which they sought to correct the bias due to the quasar proximity zones by modifying the direct measurement method to account for the excess flux due to the quasar. The resultant value of the MFP inferred by B21 is $0.75^{+0.65}_{-0.45}$ pMpc at $z = 6$. The MFP thus shows a sharp drop at $z > 5$ relative to the $(1+z)^{-5.4}$ behaviour seen at redshifts lower than 5. Moreover, the value of the MFP at $z = 6$ is smaller than the value in several of the latest reionization simulations by a factor of 2 or more (Keating et al., 2020a). Davies et al. (2021) discussed the implications of this tension for the ionizing budget of galaxies to argue that a shorter MFP that is consistent with the observed neutral hydrogen fraction from dark

pixels of the Lyα forest requires an ionizing emissivity that is up to six times larger than the typically assumed values. Cain et al. (2021) found that the short MFP reported by B21 is consistent with a late reionization scenario powered by fainter galaxies with a high ionizing photon production efficiency (Lewis et al., 2022; Garaldi et al., 2022). More recently, Zhu et al. (2023b) have measured the MFP for quasars at redshifts between $5.1 < z < 6$ using the B21 method and found the value at redshift 6 to agree with that measured by B21. They also find the decrease in the MFP to be steeper with increasing redshift, with a sharp drop in MFP by nearly 75% between $z \sim 5.6$ and $5.9$. Such a sharp decline in the MFP points towards a rapid end to reionization. Another direct measurement technique is to define the MFP by averaging free paths measured from the distribution of absorbers along individual quasar sightlines (Romano et al., 2019; Songaila & Cowie, 2010). This approach suggested a milder evolution of the MFP with redshift, between $3 < z < 6$, when compared to the direct stacking method. Bosman (2021b) have extended this approach to detect absorption due to LLS in the six lowest-order Lyman-series transitions and put a lower limit based on the average of individual free paths defined this way to be $\lambda_{\mathrm{MFP}} > 0.31$ pMpc at redshift 6, consistent with the measurement of B21.

Indirect measurements of the MFP have been obtained by comparing the observed Lyα opacity of the IGM with that in numerical simulations. The mean and the scatter of the cumulative distribution function (CDF) of the effective Lyα opacity has been used to simultaneously constrain both the MFP and the photoionization rate for a given emissivity (e.g., Gaikwad et al., 2023; Wolfson et al., 2023, and Davies et al. 2023, in preparation). Overall, the indirect MFP measurements based on the Lyα opacity (Gaikwad et al., 2023) are consistent with the direct MFP measured from the quasar stack beyond the Lyman limit, within the error bars (Becker et al., 2021; Zhu et al., 2023b). However, the slope of evolution of the MFP with redshift differs beyond redshift $z \sim 5.5$, with the direct measurements yielding a steeper slope than the indirect measurements. A gradual slope agrees better with the reionization simulations of D'Aloisio et al. (2018), Keating et al. (2020a), and Cain et al. (2021), as compared to the steeper slope seen in the models of Lewis et al. (2022) and Garaldi et al. (2022). Possible reasons for the mismatch between the indirect and direct measurements of the MFP could be the assumptions made by the two methods. B21 assume an analytic model to measure the effective optical depth, where the opacity $\kappa$ is proportional to the photoionization rate as $\Gamma^{-\xi}$. In the QSO proximity zone, the total $\Gamma$ is a sum of both $\Gamma_{\mathrm{bg}}$ and $\Gamma_{\mathrm{qso}}$, the latter predominantly varying with distance
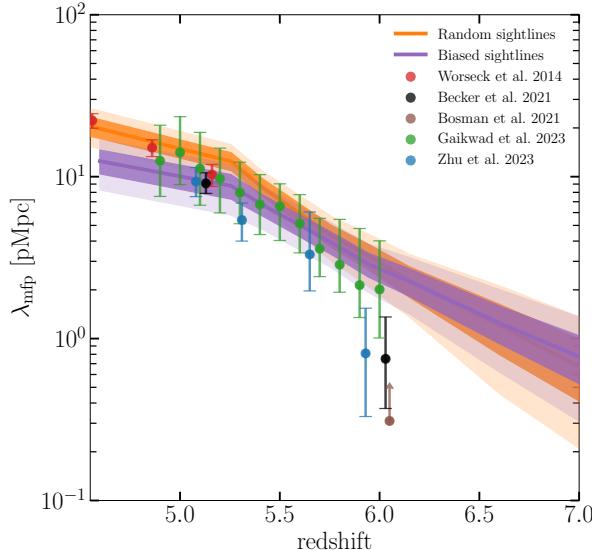
from the quasar due to geometric dilution. B21 solve for $\Gamma_{\mathrm{qso}}$ numerically, assigning average parameters for their QSO stack. They thereby keep the background value of the photoionization rate, $\Gamma_{\mathrm{bg}}$ fixed while keeping the MFP as a free parameter. In reality, the photoionization rate and MFP will both depend on the ionizing emissivity and will co-evolve. B21 also fix the opacity due to higher-order Lyman series absorption to the values obtained using an optically thin simulation while fitting their model. Roth et al. (2024) discuss the effect of this on the inference of B21. The slope of the variation of the MFP with redshift is also different between the counting LLS method, indirect inference method and the direct stacking method, the latter showing the steepest evolution. These differences reflect the dependence of the MFP on the nature of absorbing sources, each sensitive to a different measurement technique (Inoue et al., 2014).

In this thesis, we critically examine the direct measurement method of B21 for possible biases due to (*a*) higher cosmological densities around high-redshift quasars, (*b*) incomplete reionization at $z > 5.3$, and (*c*) differences between the structure of quasar proximity zones as computed using the analytical model of B21 and that obtained via radiative transfer calculations. We also investigate the challenge of defining the MFP during the epoch of reionization, when a cosmological UV radiation background is not yet established uniformly.

## 6.1   Data and Models

Our strategy in this thesis is to use a cosmological radiative transfer (RT) simulation of reionization (see Section 3.1 for details of the simulation used here). Such a simulation provides us with a realistic model for a partially ionized IGM at $z > 5.3$, on top of which we model quasar proximity zones using one-dimensional radiative transfer (see Section 3.2 for more details). We then use the resultant one-dimensional spectra to examine the MFP.

For our measurement of the MFP, we use the observed data from B21. This is in the form of a composite of quasar spectra with mean redshift $z_{\mathrm{qso}} = 5.97$, constructed with 13 quasar spectra obtained using the Keck/ESI and VLT/XSHOOTER spectrographs. The redshifts used for the quasars in the stack were obtained using the C II line, from the Ly$\alpha$ halo emission, or from the apparent start of the Ly$\alpha$ forest absorption. While selecting quasars in the stack, a minimum cut-off of $S/N \gtrsim 20$ per 30 km s$^{-1}$ was applied for the signal-to-noise ratio near the rest-frame
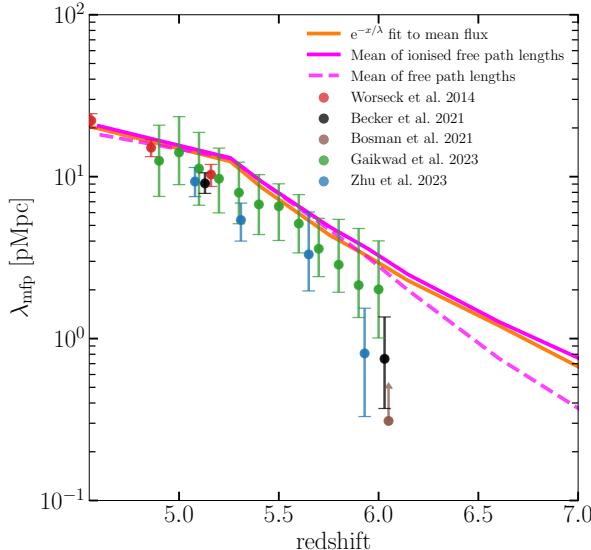
**Figure 6.1:** Effect of higher overdensity around high-redshift quasars on the MFP in simulations. The orange curve and shaded regions show the median value at the one- and two-sigma scatter of the MFP in our reionization model, when the MFP is defined using Equation (6.2.1). The purple curve and shaded regions show the same but when only sightlines originating in the highest mass halos in the simulation volume are used. The data points and error bars show various measurements of the MFP. Measurements by Becker et al. (2021) and Bosman (2021b) have been displaced in redshift by $\delta z = 0.03$ and 0.05 respectively, for legibility.

1285Å in the continuum. The mean brightness of the quasars in the sample is $M_{1450} = -27$, with a range of $-27.8 < M_{1450} < -25.7$. The typical resolution for the quasars in the stack is FWHM $\sim 45\,\mathrm{km\,s^{-1}}$ for spectra obtained using ESI and FWHM $\sim 25\,\mathrm{km\,s^{-1}}$ for spectra obtained using XSHOOTER. While stacking, the quasar flux is normalised by the median flux between 1270–1380 Å, although the normalisation with respect to the intrinsic continuum is kept a free parameter for our analysis, similar to B21. In order to study the Lyman-continuum MFP, we focus on the wavelength range of 912–825Å.

## 6.2 MFP of hydrogen-ionizing photons before the end of reionization

The post-reionization ($z \lesssim 5.5$) MFP measurements are in good agreement with several reionization simulations (Keating et al., 2020b; D'Aloisio et al., 2020; Cain et al., 2021; Lewis et al., 2022). At higher redshifts $z \gtrsim 5.5$, when reionization is still believed to be ongoing, both observations and simulations differ in the measured MFP values. While some simulations agree better with the MFP measured by B21 (Lewis et al., 2022; Garaldi et al., 2022), some models (e.g., Cain et al., 2021), including ours (Kulkarni et al., 2019), agree better with the

**Figure 6.2:** A comparison of three definitions of the MFP in an incompletely reionized IGM. The solid orange curve shows the MFP obtained from our simulation using Equation (6.2.1). This curve is also shown in Figure 6.1. The solid pink curve shows the the value of the MFP computed as the mean of the higher-value subpopulation of the bimodal distribution of free-path lengths in the simulation box. The dashed pink curve shows the MFP when computed as the mean of the full bimodal distribution of the free-path lengths.

MFP measured by Gaikwad et al. (2023). Keeping this in mind, we look at possible biases in the computation of the MFP from the simulations.

We use the standard definition for the MFP (Kulkarni et al., 2016) where the average Lyman-continuum transmission across random sightlines in the comoving frame is fit by an exponential with an $e$-folding length scale of $\lambda_{\mathrm{MFP}}$,

$$\langle F \rangle = F_0 \exp\left(-\frac{x}{\lambda_{\mathrm{MFP}}}\right). \tag{6.2.1}$$
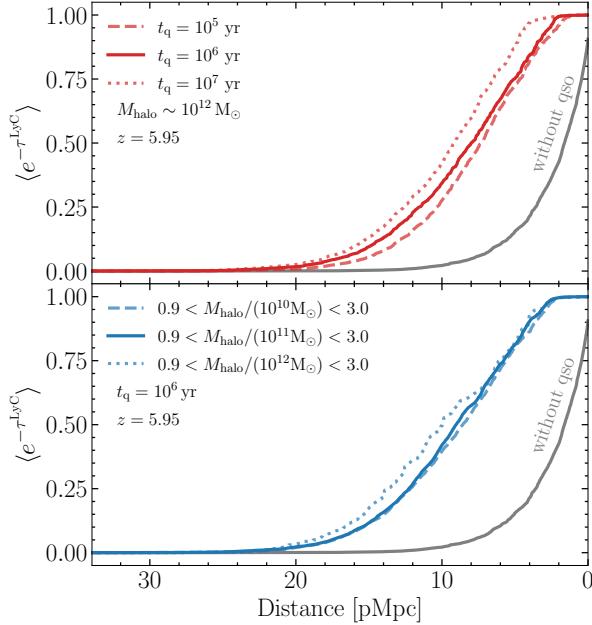
Figure 6.1 shows the evolution in our model of the MFP defined in this manner. The orange curve shows the MFP obtained by fitting Equation (6.2.1) to composite spectra of 13 randomly drawn sightlines from our simulation, to match the number of sightlines used in B21. In order to obtain the sample variance, we repeat this computation for 10,000 randomly drawn samples. This method for measuring the MFP is analogous to the approach of Prochaska et al. (2009)[1]. The solid curves in Figure 6.1 show the median of the resultant distribution of MFP values, with the shaded regions showing the one- and two-sigma scatter. We see that the disagreement between the MFP measured in this manner from the measured value of B21 and Zhu et al. (2023b) is more than $2\sigma$ at $z = 6$. Also missing in the simulations is the steep decline in the

---

[1]We have verified this by fitting our mock quasar stack discussed in Section 6.3 with the model of Prochaska et al. (2009)

MFP observed by B21 and Zhu et al. (2023b) between $z \sim 5.5$ and $z \sim 6$. We also see that the evolution of the MFP in our model becomes even more gradual once reionization is complete at $z \sim 5.3$. At these post-reionization redshifts, the optical depth is controlled by the overdense regions responsible for LLSs.

Direct measurements infer MFP from sightlines towards QSOs, which tend to reside in overdensities (e.g. Wang et al., 2023). To check for the bias due to large scale structure, we draw random sightlines originating only from halos with the highest masses sampled by our simulation volume, between $10^{12}$ and $10^{13} M_\odot$, otherwise applying Equation (6.2.1) as above. The resulting median MFP with the one- and two-sigma scatter is shown in Figure 6.1 in purple. The MFP along biased sightlines shows qualitatively similar behavior with redshift when compared to the MFP along random sightlines, but with a slightly shallower slope. The difference between the two MFPs is insignificant at $z = 6$. The random sightlines include sightlines that start from any location, including halos. Any difference seen between the two cases in Figure 6.1 is because sightlines originating from massive halos are ionized earlier than other regions in the IGM. Once reionization is complete, the over-density along sightlines originating from halos will also lead to a higher optical depth or lower MFP when compared to the measurement of random sightlines. We see in Figure 6.1 that this biased MFP is in better agreement with the measurement by Zhu et al. (2023b) and B21 than that by Worseck et al. (2014) at $z \lesssim 5.5$. We discuss this further in Section 6.3 below.

Another potential source of bias in the MFP as formally defined by Equation (6.2.1) is that this exponential attenuation assumes a spatially constant opacity. In reality, reionization is not yet complete at $z = 6$. Consequently, the Lyman-continuum opacity of the IGM has large spatial variations. An alternative definition of the MFP that addresses this complexity is one that obtains a distribution of the free-path lengths of photons in a simulation box and computes its mean (Rahmati & Schaye, 2018). We implement this in our simulation by following Rahmati & Schaye (2018) and measuring the free path as the distance at which the Lyman-continuum optical depth along the sightline becomes unity. The distribution of such free-path lengths is bimodal (see Appendix D.1). We measure the MFP as the average of the dominant part of the bimodal distribution, which represents the ionized IGM. The MFP defined in this manner is shown by the solid pink curve in Figure 6.2. We see that this results in a value of the MFP that is very close to the value obtained by using Equation (6.2.1), shown by the orange curve in
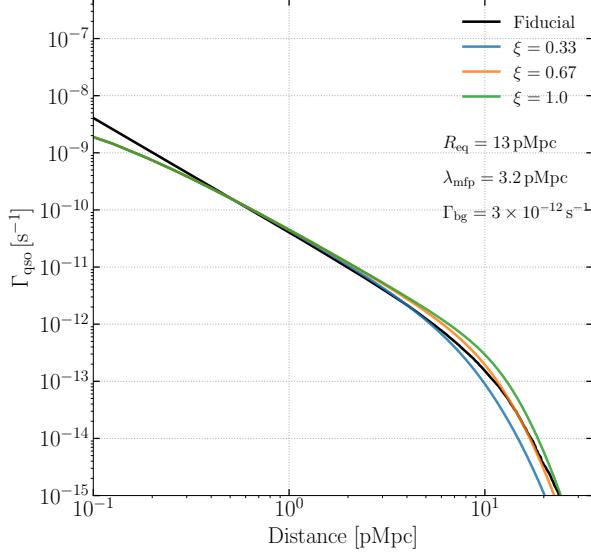
**Figure 6.3:** Effect of quasar lifetime (top panel) and host halo mass (bottom panel) on the Lyman-continuum composite spectrum of 1000 sightlines in our simulation. In both cases, the grey curve shows the transmission without the effect of the quasar proximity zone.

Figure 6.2.

In their CoDa III simulation, Lewis et al. (2022) compute the MFP using a similar free-path length method. The free-path lengths are defined in this case by doing an exponential fit to the flux along individual sightlines, which they find to be similar to the free paths measured following Rahmati & Schaye (2018). The dashed pink curve in Figure 6.2 shows the result from our simulation when we average over all free paths, similar to one of the approaches used in Lewis et al. (2022) to measure the MFP. As expected from the shape of the distribution of the free-path lengths, the MFP is now biased towards lower values. We find that this bias is small at $z \lesssim 6$. This could be a reflection of the inadequate spatial resolution of our simulation, due to which a large number of free paths become smaller than our cell size.

## 6.3 Effect of incomplete reionization on the MFP

We have seen above that neither the large-scale-structure bias nor the variations in the definition of the MFP cause a significant change in the value of the MFP in our simulation at $z = 6$. We now investigate the effect of the residual neutral hydrogen 'islands' in the IGM at this redshift on the B21 measurement of the MFP. Given the good agreement of our simulations with the measurements of the MFP by Worseck et al. (2014), Gaikwad et al. (2023) and Bosman

**Figure 6.4:** HI photoionization rate due to QSO: The black curve shows the average $\Gamma_{qso}$ for QSOs with lifetimes ranging between $10^5 < t_q < 10^7$ yr with magnitude $M_{1450} = -27$ at redshift $z = 5.95$, computed using our 1D radiative transfer simulation. The green, orange and blue curves show the average $\Gamma_{qso}$ computed analytically for the same QSO magnitude at redshift 6, for $\xi = 0.33, 0.67$, and 1.0 respectively. The value of $R_{eq}$ is computed using $\Gamma_{bg} = 3 \times 10^{-13} s^{-1}$.

(2021b), it is important to consider if the absence of patchy reionization in the models of B21 and Zhu et al. (2023b) could potentially bias their measurements. In order to do this, we construct mock data from our simulation and apply the B21 method to it. We then compare the resultant measurement of the MFP with the 'true' value of the MFP in the simulation given by Equation 6.2.1. A similar test was also performed by B21 themselves. However, the mock spectra used in their test were created from numerical simulations where reionization is assumed as instantaneous from a uniform ionizing background and post-processed to include fluctuations following the approach of Davies & Furlanetto (2016).

### 6.3.1 The B21 method

We begin by briefly reviewing the B21 method for measuring the MFP. The effective Lyman-continuum optical depth for a photon that is emitted at $z_{qso}$ and redshifts to the Lyman limit at $z_{912}$ is given by

$$\tau_{eff}^{LL}(z_{912}) = \frac{c}{H_0 \Omega_m^{1/2}}(1+z)^{2.75} \int_{z_{912}}^{z_{qso}} \kappa_{912}(z')(1+z')^{-5.25} dz', \qquad (6.3.1)$$

where $\kappa_{912}$ is the opacity to 912 Å photons. The dependence on redshift is as follows: the exponent 2.75 is a result of the dependence of cross-section on frequency as $\sigma_\nu^{\mathrm{LL}} \propto \nu^{-2.75}$, while the exponent $-5.25$ comes from the conversion of comoving distance to redshift as $r \propto aH \propto (1+z)^{-1} \times (1+z)^{-3/2}$ in the matter-dominated era. The opacity $\kappa_{912}$ is assumed to scale with the photoionization rate as a power law, so that at a distance $r$ from the quasar,

$$\kappa_{912}(r) = \kappa_{912}^{\mathrm{bg}} \left( \frac{\Gamma_{\mathrm{tot}}(r)}{\Gamma_{\mathrm{bg}}} \right)^{-\xi}, \tag{6.3.2}$$
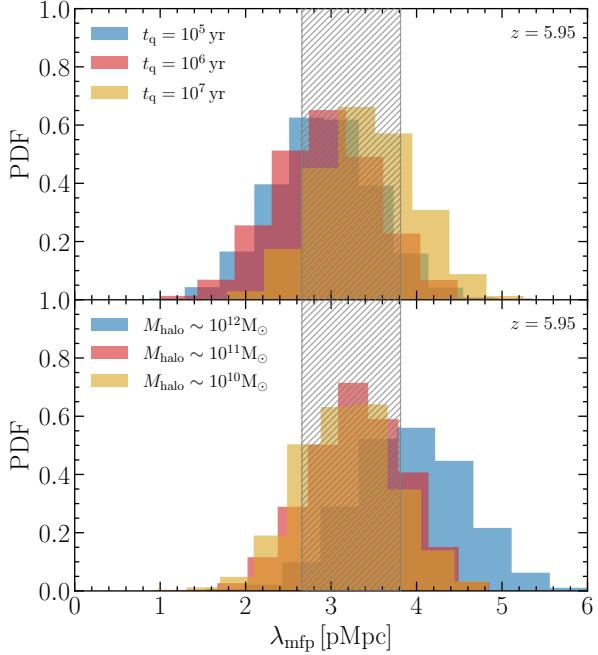
where $\kappa_{912}^{\mathrm{bg}}$ is the opacity to the ionized background and $\Gamma_{\mathrm{tot}}(r) = \Gamma_{\mathrm{qso}}(r) + \Gamma_{\mathrm{bg}}(r)$ is the sum of the photoionization rate due to the QSO and the local background photoionization rate. The opacity and photoionization rate are related to each other through their mutual dependence on the shape and number density of neutral gas absorbers. This information is parameterised using the power-law index $\xi$, which has been studied using analytic arguments as well as numerical simulations (Muñoz et al., 2016; McQuinn et al., 2011). We discuss this parameter in detail in Section 6.3. Equation 6.3.3 is applied to a stack of QSO spectra, so the terms in the equation are all averaged quantities, and the average of the local background photoionization is assumed to be uniform and equal to $\Gamma_{\mathrm{bg}}$. The average photoionization rate due to all quasars at a location $r$ is computed by iteratively solving for

$$\Gamma_{\mathrm{qso}}(r + \delta r) = \Gamma_{\mathrm{qso}}(r) \left( \frac{r + \delta r}{r} \right)^{-2} e^{-\kappa_{912}\delta r}, \tag{6.3.3}$$

with the initial condition

$$\Gamma_{\mathrm{qso}}(\delta r) = \Gamma_{\mathrm{bg}} \left( \frac{\delta r}{R_{\mathrm{eq}}} \right)^{-2}. \tag{6.3.4}$$

Here, $R_{\mathrm{eq}}$ is the distance at which the photoionization rate due to the quasar's radiation is equal to the background photoionization rate in the absence of any absorption. $R_{\mathrm{eq}}$ depends on the QSO magnitude and spectrum in addition to the background photoionization rate. The free parameters in this model are therefore, $\kappa_{912}^{\mathrm{bg}}$, $R_{\mathrm{eq}}$, $\xi$ and $\Gamma_{\mathrm{bg}}$. The mean flux of the stack is fit by the above model of opacity and the MFP $\lambda_{\mathrm{MFP}}$ is inferred as the distance from the quasar at which $\tau_{\mathrm{eff}}^{\mathrm{LL}}(z_{912})$ becomes equal to 1.

**Figure 6.5:** Distribution of the MFP obtained using 10,000 draws of 13-sightline stacks from our simulation for different quasar ages (top panel), and different quasar host halo mass (bottom panel). In each case, the M FP is determined using the B21 method, where $R_{\mathrm{eq}}$ and $\Gamma_{\mathrm{bg}}$ are chosen from our simulation and $\xi = 0.67$. As a comparison, the hatched region shows the 68% scatter in the MFP in the simulation, obtained using Equation (6.2.1).

### 6.3.2 Mock data

To construct stacks of mock spectra at redshift 6, we consider QSOs with magnitude of $M_{1450} = -27$ with a broken power law spectral profile for the specific luminosity as in Lusso et al. (2015),

$$
L_\nu \propto \begin{cases} \nu^{-0.61} & \text{if } \lambda \geq 912 \text{ Å}, \\ \nu^{-1.70} & \text{if } \lambda < 912 \text{ Å}. \end{cases} \tag{6.3.5}
$$

Although it is not a deciding factor, we choose the magnitude of our mock stack to be of the value same as the mean magnitude of quasars in the B21 stack at redshift 6. The power law index values are slightly different from those assumed in B21 ($-0.5$ and $-1.5$ for wavelengths greater than and less than 912 Å respectively).

After placing QSOs in halos, we post-process the sightlines with our 1D radiative transfer code as described in Section 6.1 to obtain the distribution of the neutral hydrogen density. Using this, we compute the ionizing optical depth along a sightline as,

$$
\tau^{\mathrm{LL}}(r) = \sum_r n_{\mathrm{HI}} \sigma_{\mathrm{HI}}^{\mathrm{LL}} \mathrm{d}r, \tag{6.3.6}
$$

where $\sigma_{\mathrm{HI}}^{\mathrm{LL}} = 6.3 \times 10^{-22} \mathrm{m}^2$ is the hydrogen ionization cross-section. Since Equation 6.3.3 does not explicitly assume the QSO lifetime or the host halo mass, we randomize over them in our stack, with halo masses between $10^{10} \mathrm{M}_{\odot} \lesssim M_{\mathrm{halo}} \lesssim 10^{12} \mathrm{M}_{\odot}$ and lifetimes between $10^5 < t_{\mathrm{q}} < 10^7$ Myr. Each of our composite stacks comprises 13 sightlines.
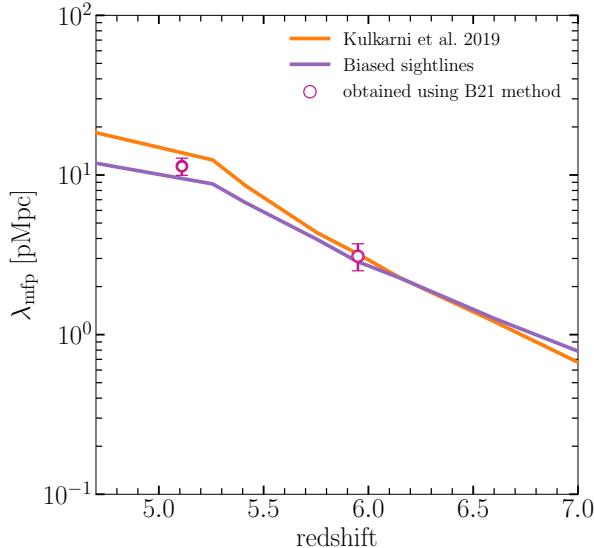
Figure 6.3 shows composite stacks obtained in this manner with and without quasars for the above range of quasar lifetimes and host halo masses. The top panel shows the variation of the mean continuum flux along quasar sightlines originating from $\sim 10^{12} \mathrm{M}_{\odot}$ halos and having lifetimes between $10^5 < t_{\mathrm{q}} < 10^7$ Myr. The mean flux computed along sightlines drawn at random directions is shown in grey. Once the quasar is on, the opacity near the quasar decreases as the quasar emits ionizing photons that reduce the neutral hydrogen density. The mean flux thus increases with increase in the quasar lifetime. While the decrease in flux is gradual when the quasar turns on in a uniformly ionized medium, we notice that the flux fall off is steeper in our patchy model. The bottom panel shows the variation of the mean continuum flux when the quasar lifetime is fixed to be $t_{\mathrm{q}} = 10^6$ yr and the quasar host halo masses are varied to be in three ranges between $10^{10} \mathrm{M}_{\odot} \lesssim M_{\mathrm{halo}} \lesssim 10^{12} \mathrm{M}_{\odot}$. In our simulation, more massive halos are ionized earlier and therefore the mean flux is slightly higher along these sightlines. Figure 6.3 shows that the mean flux changes almost doubles in the presence of the quasar, while changing the lifetime or host halo masses of the quasars by an order of magnitude changes the mean flux by less than around 10%. For the purpose of recovery of the MFP using the B21 method, we do not include the absorption due to higher order Lyman series transitions in the mock spectra. We correspondingly fit the mock spectra without including the contribution from the Lyman series opacity, as shown in Equation 6.3.7. We also do not add a zero-point correction to our mock stack and hence do not include this parameter while fitting using Equation 6.3.7.

### 6.3.3 Choice of $R_{\mathrm{eq}}$, $\Gamma_{\mathrm{bg}}$ and $\xi$

Following B21 we fit the mock spectra by an exponential

$$\langle F \rangle = F_0 \exp\left(-\tau_{\mathrm{eff}}^{\mathrm{LL}}\right), \tag{6.3.7}$$

with $\kappa_{912}^{\mathrm{bg}}$ as free parameter. The other parameters are $R_{\mathrm{eq}}$, $\Gamma_{\mathrm{bg}}$ and $\xi$, which we discuss below. We chose to keep these parameters fixed, as B21 did for their nominal measurement. We have

**Figure 6.6:** Open circles show the MFP inferred using the B21 method from a mock composite spectrum drawn from our simulation. The curves shows the MFP in the simulation, obtained using Equation (6.2.1), using sightlines starting from randomly chosen halos (orange curve), and sightlines starting from only the most massive halos (purple curve). When we use the B21 method for measuring the MFP, we use $\xi = 0.67$ and use values of $R_{\rm eq}$ and $\Gamma_{\rm bg}$ derived from the ionized regions in the simulation.

inspected the results while keeping $\xi$ as a free parameter, and found them to be consistent with what was reported in Zhu et al. (2023b). We leave self-consistent parametrization of $\Gamma_{\rm bg}$ with $\kappa_{912}^{\rm bg}$ for future work.

The distance $R_{\rm eq}$ is given by Calverley et al. (2011)

$$R_{\rm eq} = \left[ \frac{L_{912}\sigma_0}{8\pi h^2 \Gamma_{\rm bg}(\alpha_\nu^{\rm ion} + 2.75)} \right]^{1/2}. \qquad (6.3.8)$$

The analytic expression for $R_{\rm eq}$ is computed under the approximation that the quasar photoionization rate falls as $1/r^2$, where $r$ is the distance from the source. This equation can be used to further compute $\Gamma_{\rm qso}$ in the presence of absorption by iterating over Equation 6.3.3. For the mock stacks, we chose a value of $R_{\rm eq}$ = 13 pMpc, for $L_{912}$ corresponding to magnitude of $M_{1450} = -27.0$ and $\Gamma_{\rm bg}$ of $3 \times 10^{-13}\,{\rm s}^{-1}$ as discussed below.
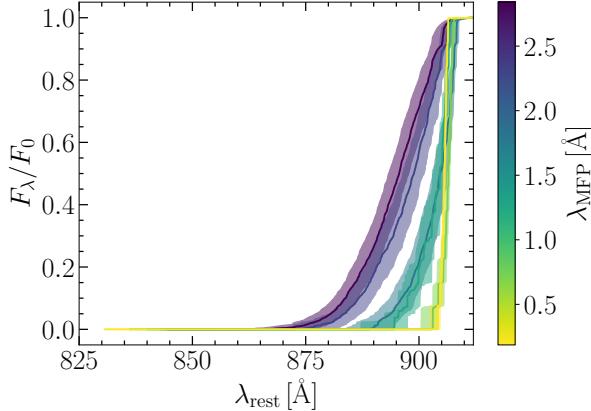
B21 chose a value of $3 \times 10^{-13}\,{\rm s}^{-1}$ for $\Gamma_{\rm bg}$ in their analysis. The background photoionization rate in our simulations is inhomogeneous. In ionized regions, we find the average $\Gamma_{\rm bg}$ to be around $3 \times 10^{-13}\,{\rm s}^{-1}$ at redshift $z \sim 6$. However, the average value of $\Gamma_{\rm bg}$ is $10^{-13}\,{\rm s}^{-1}$, almost 3 times smaller than the value assumed by B21, as there are large neutral regions where $\Gamma_{\rm bg}$ is as small as $10^{-22}\,{\rm s}^{-1}$. Given that at redshift $z \sim 6$, the volume averaged ionized hydrogen fraction is close to 80%, it would suffice to use this value of background photoionization rate to compute

$\Gamma_{qso}$. We note that Zhu et al. (2023b) use a different $\Gamma_{bg}$ of $1.5 \times 10^{-13}\,\text{s}^{-1}$ at $z \sim 6$ from Gaikwad et al. (2023). They find that at $z \sim 5.93$, assuming a value of $3 \times 10^{-13}\,\text{s}^{-1}$ for $\Gamma_{bg}$ increases their mean free path from the measured value of 0.81 pMpc to ~1 pMpc. The parameter $\xi$ encodes the nature of the density of the absorbers that set the local MFP (Furlanetto & Oh, 2005). For an isothermal absorber, the theoretical prediction is $\xi = 2/3$, while the scaling inferred from simulations is dependent on the self-shielding systems and can range between 0.33 and 1.0 (McQuinn et al., 2011; D'Aloisio et al., 2020). For $\xi$, we use the nominal parameter from B21 where it is assumed to be 0.67 based on the arguments presented in Miralda-Escudé et al. (2000). In Figure 6.4, we compute the photoionization rate using Equation 6.3.3 for a quasar at redshift 6 and magnitude $M_{1450} = -27$. $R_{eq}$ was computed using Equation 6.3.8, assuming $\Gamma_{bg}$ of $\sim 3 \times 10^{-13}\,\text{s}^{-1}$. The resulting $\Gamma_{qso}$ is compared to the $\Gamma_{qso}$ from our simulations averaged over 1000 lines of sight to the QSO. The analytic $\Gamma_{qso}$ is in agreement with our simulated value within 25%, for a $\xi$ value of 0.67. The analytic method however, under-estimates $\Gamma_{qso}$ by almost two orders of magnitude close to the quasar for all values of $\xi$. Keeping $\xi$ as a free parameter can change the MFP measurement by up to a factor of 2, as has been discussed in Zhu et al. (2023b). Roth et al. (2024) discuss that keeping $\xi$ as a free parameter yields a better fit to their mock spectra in the presence of neutral islands. They however find that such a fit might result in a MFP that is biased high with respect to the true value in their simulation.

### 6.3.4 Recovery of the 'true' MFP using the B21 method

We use the B21 method to measure the MFP in our simulation using the mock data generated using the methods discussed in Section 6.3. We find that the resulting MFP agrees with the MFP of our simulation measured in Kulkarni et al. (2019). This shows that the MFP measured using the B21 method is not biased due to the residual neutral hydrogen islands, quasar lifetimes or due to the overdensities around quasars.
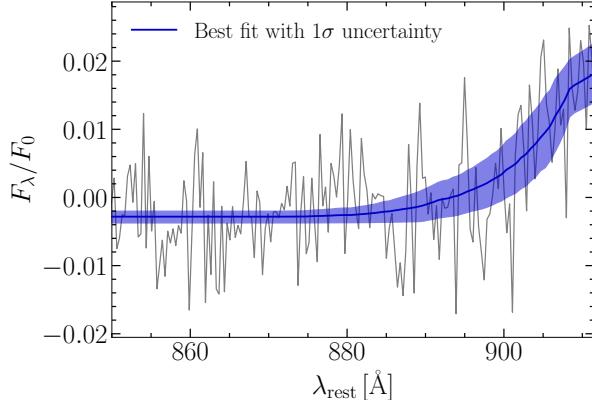
In Figure 6.5, we show the distribution of MFP measured in our simulations along 10,000 stacks of 13 QSOs each. In the fiducial model, the MFP is measured along random sightlines without a QSO by simply fitting an exponential to the average flux as discussed in Section 6.2. To compute MFP from mock stacks including a QSO, we use the B21 method with parameters discussed in the previous subsection. We consider mock stacks with different QSO properties,

**Figure 6.7:** Stacked quasar spectra blueward of rest-frame 912Å in our model for a range of values of the MFP at $z = 6$. In each case, the solid curve shows the composite of 13 randomly chosen sightlines from the simulation box. The shaded region shows the 68% scatter across 10,000 such samples.

varying QSO lifetimes and varying host halo masses. The resulting distribution of MFP for 10,000 stacks each of 13 QSOs for different QSO lifetimes and host halo masses in shown in the top and bottom panels of Figure 6.5, respectively. We find that irrespective of the quasar lifetime, we are able to recover the MFP from the mock stacks using the B21 method reasonably well, up to within $\sim 10\%$ of our fiducial value. Similarly, the MFP computed from mock stacks with lower mass halos is in good agreement with the fiducial MFP. If we use a quasar stack constructed along sightlines originating in the heaviest mass halos with $10^{11}\,\mathrm{M_\odot} \lesssim M_{\mathrm{halo}} \lesssim 10^{12}\,\mathrm{M_\odot}$, the measured MFP distribution is offset from the fiducial value by $\sim 25\%$.

We show the MFP measured from the 10,000 mock stacks for nominal values of quasar lifetime ($10^6\,\mathrm{yr}$) and host halo mass ($\sim 10^{12}\,\mathrm{M_\odot}$) at redshifts $z = 5.1$ and 6 in Figure 6.6. Also shown are the MFPs along random sightlines and biased sightlines originating in halos but without a QSO, as discussed in Section 6.2. The MFP recovered from our mock stacks using the B21 method matches better with the MFP measured along biased sightlines, which is particularly evident at post-reionization redshifts. This would suggest that the analytic computation of $\Gamma_{\mathrm{qso}}$ is robust in accounting for the QSO ionization flux, but is biased to measure the MFP along sightlines originating in overdensities. Note however that at higher redshifts, this bias becomes negligible. Another concern with the analytic model is the uncertainties introduced by the assumptions about the free parameters in the model, $R_{\mathrm{eq}}$, $\xi$ and $\Gamma_{\mathrm{bg}}$. Zhu et al. (2023b) have shown that increasing or decreasing the value of $\xi$ can half or double the MFP estimates relative to the estimates with nominal values. This raises a need for a better constrained value of the MFP independent of the choice of $\xi$. For this purpose, we use simulated models as discussed

**Figure 6.8:** Our best-fit model for the B21 composite spectrum, with the $1\sigma$ uncertainty indicated by the shaded region.

in the following section.

## 6.4 A direct measurement using radiative transfer models

We now proceed to analyse the composite spectrum of B21 using our radiative transfer models of patchy reionization and quasar proximity zones. Figure 6.7 shows the composite spectra from 13 randomly drawn sightlines along a QSO with magnitude $M_{1450} = -27.0$ and lifetime of 1 Myr in our simulation for a range of values of the mean free path. As our reionization simulation is calibrated to the observed Ly$\alpha$ mean transmission, there is no freedom to change the mean free path at a given redshift. Instead, to construct Figure 6.7, we use the values of the ionized hydrogen fraction, the gas density, and temperature from different redshifts, in each case scaling the densities by $(1+z)^3$ to $z = 6$.

In order to fit the model stacks in Figure 6.7 to the data, two modifications need to be made. First, we need to account for the absorption due to higher-order Lyman-series transitions that will cause absorption blueward of the rest-frame quasar Lyman-limit. Second, to match the observed flux, we need to consider the absorption due to the Lyman-series and continuum photons of the intrinsic QSO spectrum. Following B21 we assume the intrinsic quasar spectrum is a power law with $f_\lambda^{\text{SED}} = f_{912}(\lambda/912\,\text{Å})^{-\alpha_{\text{ion}}}$, and keep $f_{912}$ as a free parameter.

We include the foreground absorption due to higher-order Lyman-series transitions as follows. At a given $\lambda_{\text{obs}}$, the absorption due to transition of the jth Lyman series line will happen due to an absorber at a redshift $z_j = \lambda_{\text{obs}}/\lambda_j - 1$. The effective Lyman series optical depth due

109

to all transitions is then,

$$\tau_{\text{eff}}^{\text{Lyman}}(\lambda_{\text{obs}}) = \sum_j \tau_{\text{eff}}^j(z_j) \tag{6.4.1}$$

To obtain the optical depth due to the $j$th transition, we rescale the corresponding Ly$\alpha$ optical depths at $z_j$ by the product of their oscillator strength and rest-frame wavelength such that $\tau_j = \tau_\alpha f_j \lambda_j / f_\alpha \lambda_\alpha$. This scaling is valid in the optically thin regime corresponding to most of the redshifts of our interest, while in the damping wings the optical depths scale roughly as square of oscillator strengths (Malloy & Lidz, 2015). We use the post-processed sightlines to compute the Ly$\alpha$ optical depths in the presence of a quasar. The optical depths outside the proximity zone were assumed to be from the 40-2048 Sherwood simulation (Bolton et al., 2017), rescaled to the match the observed mean flux at low $z$ as discussed in B21. From the mean flux at $z_j$, computed as negative exponential of $\tau_j$, the corresponding effective optical depth is obtained as $\tau_{\text{eff}}^j = -\log\langle F \rangle$. Including the absorption from Lyman-limit and 38 higher order Lyman-series terms, the observed flux is

$$F_\lambda = f_{912}\left(\frac{\lambda}{912\,\text{Å}}\right)^{-\alpha_{\text{ion}}} \exp(-\tau_{\text{eff}}^{\text{Lyman}}) \exp(-\tau_{\text{eff}}^{\text{LyC}}) + f_0 \tag{6.4.2}$$

where $f_0$ is the zero-point correction factor, that accounts for the uncertainty in the estimated zero-point of the data in B21. We fit our models to the data by sampling the posterior distribution in a Bayesian manner. We use the likelihood

$$\mathcal{L}(\mathbf{d}|\theta) = \frac{1}{\sqrt{(2\pi)^{\mathbf{n}}\det\mathbf{C}}} \exp\left[-\frac{1}{2}(\mathbf{m}-\mathbf{d})^{\text{T}}\mathbf{C}^{-1}(\mathbf{m}-\mathbf{d})\right], \tag{6.4.3}$$

where the column matrices $\mathbf{m}$ and $\mathbf{d}$ denote the model and data vectors, respectively, both with $n$ elements corresponding to the number of wavelength bins used. The covariance matrix $\mathbf{C}$ is obtained as follows. The data covariance matrix $\mathbf{C}_{\text{data}}$ over the wavelength range of our interest (912–850 Å) is computed from the data by B21, using bootstrap realisations of the mean flux. Due to the limited data set of 13 quasars, this matrix is noisy and may underestimate the sample variance. We therefore compute a separate bootstrapped model covariance matrix for each of the model parameters using 30,000 simulated stacks and obtain a model correlation matrix $r$. We then regularise the data covariance matrix $\mathbf{C}_{\text{data}}$ using the model correlation matrix $r$ to

obtain the covariance matrix $\mathbf{C}$ used in Equation (6.4.3) at each point in the parameter space as

$$\mathbf{C}^{ij} = \frac{r^{ij}_{\mathrm{sim}}}{\mathbf{C}^{ii}_{\mathrm{data}} \mathbf{C}^{jj}_{\mathrm{data}}}, \tag{6.4.4}$$

where

$$r^{ij}_{\mathrm{sim}} = \frac{\mathbf{C}^{ij}_{\mathrm{sim}}}{\mathbf{C}^{ii}_{\mathrm{sim}} \mathbf{C}^{jj}_{\mathrm{sim}}}. \tag{6.4.5}$$

See applications by Lidz et al. (2006a); Viel et al. (2013); Iršič et al. (2017) for a discussion of this approach.
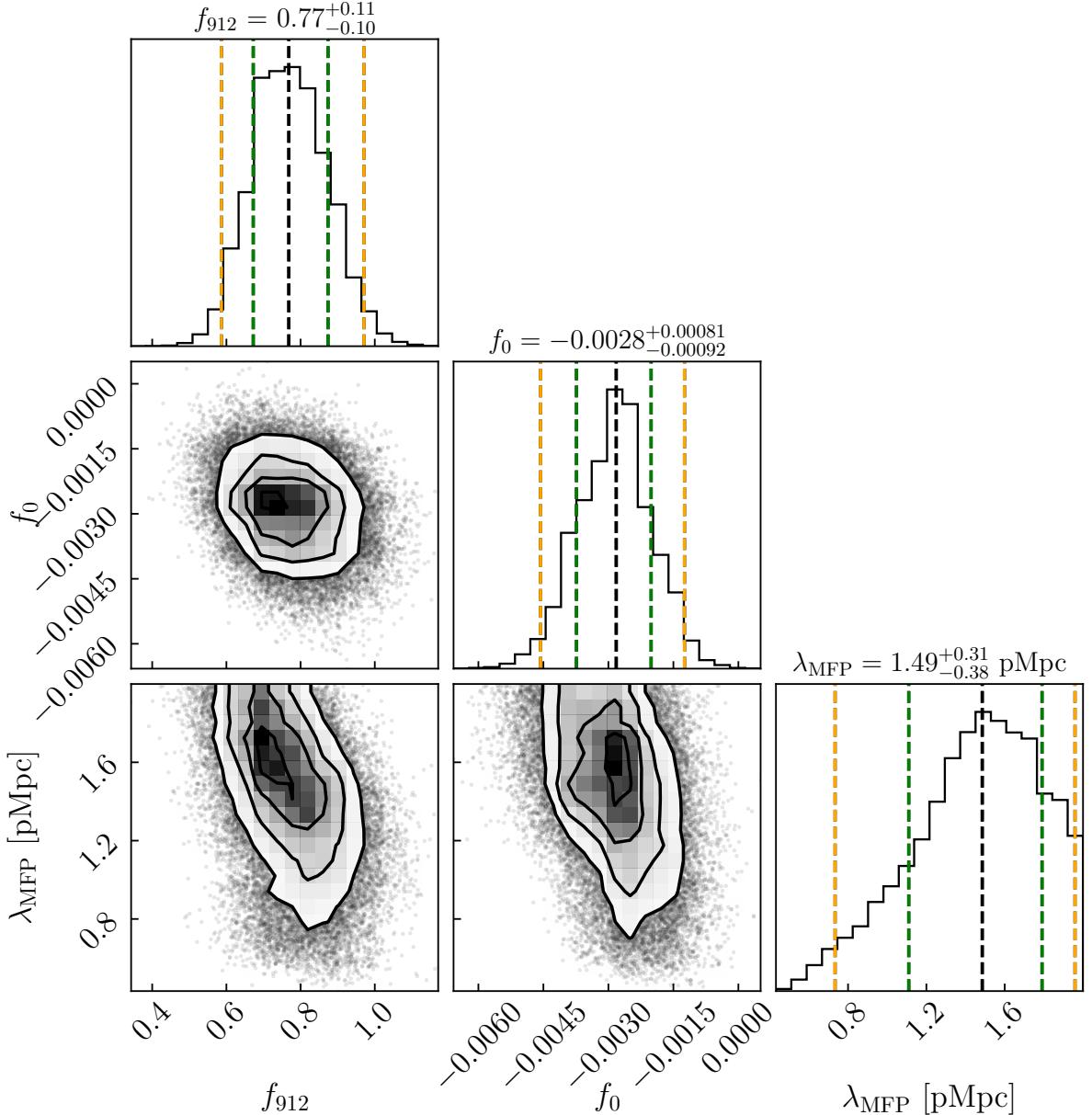
We assumed uniform priors in the ranges $0.3 < \lambda_{\mathrm{MFP}} < 2$, $0 < f_{912} < 1$ and $-0.001 < f_0 < 0.01$. The model stacks and the likelihood function were linearly interpolated over a grid of models wherever necessary. We used the `emcee` package (Foreman-Mackey et al., 2013) to perform the inference by sampling the posterior distribution using MCMC via Bayes' theorem,

$$p(\theta|\mathbf{d}) \propto \mathcal{L}(\mathbf{d}|\theta)\mathbf{p}(\theta). \tag{6.4.6}$$

The resultant best-fit curve, given by the mean of the posterior, is shown in Figure 6.8, along with the $1\sigma$ uncertainty. The marginalised posterior distributions of our model parameters is shown in Figure 6.9.

We expect our assumptions for quasar lifetimes and host halo masses to introduce additional uncertainty on these parameters. We approximate these uncertainties by adding them in quadrature to the uncertainty derived from the posterior distributions. We saw in Figure 6.5 that varying quasar lifetimes and host halo masses leads to a $\sim$10% and $\sim$25% change, respectively, in the mean free path. With this additional error included, our inferred value of MFP is $\lambda_{\mathrm{MFP}} = 1.49^{+0.47}_{-0.52}$ pMpc.

Figure 6.10 shows the comparison of our new measurement with previous measurements and our simulation. While our new measurement is slightly higher than the MFP measured by B21, it is consistent with their result. Our measurement is also consistent with the other estimates by Gaikwad et al. (2023) and Zhu et al. (2023b). Our new measurement remains lower than the MFP of our simulations by more than $2\sigma$. Nonetheless, the consistency between our inference and that of B21 suggests that the B21 method is robust with respecive to the quasar proximity zone modelling.
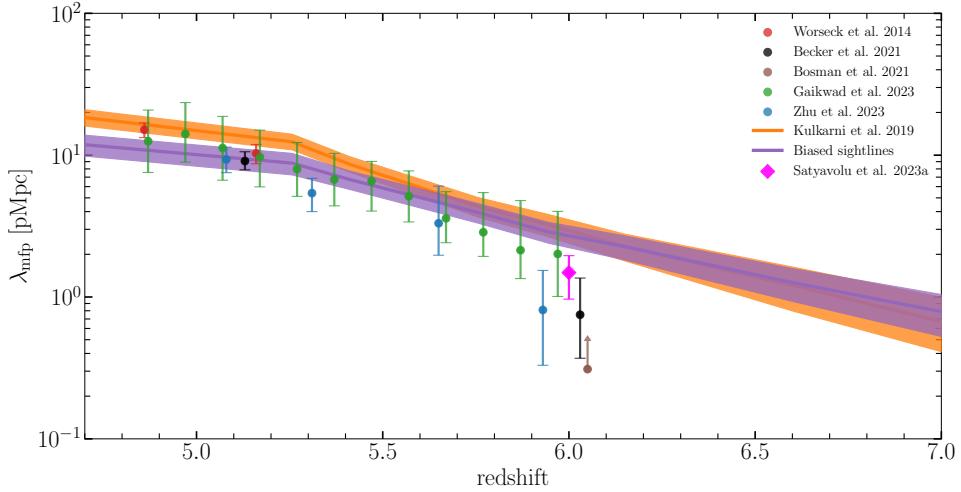
**Figure 6.9:** Corner plot showing the posterior distributions of the normalization parameters ($f_{912}$, $f_0$) and the mean free path ($\lambda_{\mathrm{MFP}}$). The rightmost panels in each row indicate the marginalized posteriors. The black, green, and yellow dotted lines represent the median, $1\sigma$, and $2\sigma$ levels of the distribution, respectively.

## 6.5 MFP at redshift $z \sim 6$

In this Chapter, we have taken a closer look at the mean free path of hydrogen-ionizing photons at $z = 6$ in order to examine potential sources of bias in recent measurements. Our findings are as follows,

- At least for $z \lesssim 6.5$, there is no significant difference in the value of the mean free path obtained by using any of the multiple definitions of this quantity that have been used in

**Figure 6.10:** The magenta point shows our measurement of the MFP using the B21 data with our model. The error bar on this point shows the $1\sigma$ uncertainty. We see that this measurement is consistent with that of B21 which is shown by the black point. This suggests that the B21 measurement is robust against the sources of potential bias that we have considered in this thesis. The orange curve with shaded regions shows the mean free path in the reionization simulations of Kulkarni et al. (2019). We see that our new measurement prefers a factor-of-two smaller MFP than this simulation. We also show other measurements of the MFP, by Worseck et al. (2014), Bosman (2021b), Gaikwad et al. (2023), and Zhu et al. (2023b). Measurements by Becker et al. (2021); Zhu et al. (2023b) and Bosman (2021b) have been displaced in redshift by $\delta z = \pm 0.03$ and 0.05 respectively for legibility.

the recent literature for redshifts at which reionization is still incomplete.

- The bias in the mean free path due to the overdensities around quasars is also minimal at $z \sim 6$. At lower redshifts, with $z \lesssim 5$, the overdensities can bias the mean free path towards lower values by about 50%.

- Due to the short mean free path at $z \sim 6$, the Lyman-continuum composite spectra used in the literature for a direct MFP measurement are affected by the quasar proximity zones. The dependence of the inferred mean free path on the variations in quasar lifetime and host halo masses appears, however, only to be small and less than 25%.

- The patchiness of reionization also has minimal effect on the direct measurement of the MFP reported by B21.

- Using radiative transfer modelling of patchy reionization and quasar proximity zones, we reanalyze the data obtained by B21 to find a MFP of $\lambda_{\mathrm{MFP}} = 0.90^{+0.66}_{-0.40}$ pMpc at $z = 6$. This is consistent with the B21 measurement, and is smaller than the MFP in reionization simulations by a factor of $\sim 2$.

We fit to the data while self-consistently modelling quasar proximity effect on the MFP in our simulations. We include the quasars through post-processing, thereby neglecting the

response of gas densities to the photoheating caused by the quasar ionization. However, the dynamical timescale for the gas to respond ($\sim 100$ Myr) is usually much larger than the average episodic lifetime of the quasar, which is around $\sim 1$ Myr (D'Aloisio et al., 2020; Morey et al., 2021; Satyavolu et al., 2023a). We also do not include feedback from AGN which could result in gas heating as well as AGN driven clumpy outflows, although the former effect is shown to be not major (Trebitsch et al., 2018). The response of gas densities to heating from reionization might nonetheless be not captured in our main cosmological RT simulation. This would require running either hybrid or fully coupled cosmological hydrodynamical simulations (e.g. Puchwein et al., 2023; Garaldi et al., 2022). The error on MFP due to keeping the quasar spectral profile fixed is expected to be small (B21).

Cain et al. (2021) have modeled sinks in a subgrid fashion and found that the small-scale clumping reduces the MFP. There has also been evidence for presence of excess LLSs along the line-of-sight to quasars (Hennawi et al., 2006). We furthermore do not resolve the dense gas within mini-halos in our simulation. Recent works show that the Ly$\alpha$ flux is suppressed by 10% on average due to mini-halos, with the suppression being enhanced along lines of sight in the vicinity of large halos, above redshifts $z \gtrsim 5.5$ (Park et al., 2024). Inclusion of dense absorbing gas such as LLSs or mini-halos in our simulations therefore will not only reduce the MFP along the random sightlines shown in orange in Figure 6.1, but also further reduce the continuum flux along quasar sightlines. This would push our measured MFP in Figure 6.10 to higher than its current value.

In summary, the shortness of the mean free path relative to radiative transfer simulation could be due to unresolved excess Lyman-limit systems (LLSs) along quasar sightlines. Consequently, the overall picture that emerges is consistent with one in which reionization ends much later than redshift 6, with the photon budget dominated by faint galaxies with high ionizing efficiency (Davies et al., 2021; Cain et al., 2021).

# Chapter 7

# Summary and Conclusion

A primary goal of observational cosmology is to bridge the observational gap between the CMB and the post- HI reionization era. Thanks to several ground-based optical and sub-mm telescopes, and more recently JWST, we are now shedding light on the Cosmic Dawn, the latter half of the aforementioned observational gap. For example, in the last five years, the number of quasars identified from the reionization epoch has increased five-fold. On the theoretical side, the objective is to develop a comprehensive simulation of the Universe that accurately represents the co-evolution of dark matter, baryons, and radiation during reionization. We now have a good understanding of the ionization state of the intergalactic gas during the late stages of reionization, as well as the timing of these stages, thanks to quasar absorption spectroscopy and cosmological hydrodynamical radiative transfer simulations. However, it remains crucial to constrain the remaining unknowns of reionization, such as the nature of ionizing sources and sinks, by leveraging our existing knowledge.

We now know that the formation and evolution of the first galaxies are closely linked to their central supermassive black holes (SMBHs). These early SMBHs, visible to us as quasars when they are radiatively accreting, have masses up to $10^9$ $M_\odot$, comparable to the heaviest SMBHs in the local Universe. While several pathways have been proposed to achieve such early growth of massive black holes, it is still unclear how their contributions play out.

Evidently, from the above discussion, high-redshift quasars stand at the intersection of several open questions in cosmology and extragalactic astrophysics. In particular, quasar proximity zones, as discussed in this thesis, have been crucial for estimating quasar lifetimes and the mean free path of hydrogen ionizing photons in the reionization epoch. The goal of this

thesis was to update the models of proximity zones based on realistic reionization simulations and test their consistency with data, while also adding to their measurements using the latest observations. The findings of this thesis are summarized below, in the form of the following three questions.

## 1. Are residual neutral islands towards the late end of reionization responsible for the fraction of small quasar proximity zones at high redshifts?

Proximity zone sizes of about 10% of the brightest quasars at $z > 6$ have been surprisingly found to be 2–3 times smaller than the average (Eilers et al., 2021, 2017). Such small proximity zones could result from episodic activity of quasars, or from the premature truncation of proximity zones due to absorption systems such as DLAs, LLSs or residual neutral islands in the reionized IGM around the quasar. In our work (Satyavolu et al., 2023a), we showed, using the latest reionization simulations post-processed with 1D radiative transfer simulation, that the truncation of proximity zones due to the residual neutral hydrogen islands at the tail end of reionization can lead up to 30% reduction in their sizes at redshifts $z \sim 6$, with the percentage being higher at higher redshifts (Figure 3.3). This demonstrated that neutral islands from patchy reionization cannot alone explain the fraction of small proximity zones or the scatter in the observed distribution of proximity zone sizes at high redshifts (Figure 3.6).

## 2. How can observed proximity zone sizes be made consistent with SMBH growth?

A crucial factor in understanding SMBH growth is the duration for which they underwent accretion. Quasar proximity zones were used to estimate individual episodic lifetimes for quasars (e.g., Eilers et al., 2021). In our work (Satyavolu et al., 2023a), we showed, for the first time, that the observed distribution of both small and large proximity zones at $z \sim 6$ are best explained by a model where the quasars have small duty cycle of 0.1 and an episodic lifetime of 1 Myr (Figure 4.3). Using an exponential SMBH growth model with these parameters, we show that it is possible to reach a billion solar mass black hole by $z \sim 6$, only when quasars were UV-obscured for more than 60% of their active phase. Our findings on the obscuration fraction are consistent with several theoretical predictions from quasars and black hole growth simulations (Ni et al., 2020; Davies et al., 2019; Worseck et al., 2021).

To improve statistical reliability, we measured the proximity zone sizes for 22 quasars (Satyavolu et al., 2023b) using high-quality XQR-30 spectra (D'Odorico et al., 2023b) at redshifts between 5 and 7 (Figure 5.4). This effort has increased the total number of proximity zone measurements in the epoch of reionization to 87, which will strengthen our confidence in determining quasar lifetimes in future studies.

### 3. How to measure the mean free path of hydrogen-ionizing photons during reionization?

The mean free path (MFP) of ionizing photons is an important parameter for understanding the sources and sinks of reionization. MFP can be measured using stacked quasar spectra by computing the distance at which the mean flux beyond the rest-frame Lyman-limit gets attenuated by a factor of $1/e$ (Prochaska et al., 2009). However, at higher redshifts, the measurement of MFP within the proximity zone is biased high due to the quasar ionization. Becker et al. 2021 (B21) measured the MFP at $z \sim 6$ by analytically modelling the quasar proximity effect and found it to be $\lambda_{\mathrm{MFP}} = 0.75^{+0.65}_{-0.45}$ pMpc, which is around two times smaller than several of the reionization simulations. In our study (Satyavolu et al., 2023c), we investigate potential biases in the B21 direct measurement method due to higher cosmological densities around high-redshift quasars, incomplete reionization at $z > 5.3$, and differences in proximity zones between B21's analytical model and radiative transfer calculations. We find that the analysis of B21 is robust with respect to these biases, but the systematic uncertainties remain. Further, we devise a new technique where we forward model the quasar proximity zones using simulations to directly infer the MFP from the data of B21. Our measurement, $\lambda_{\mathrm{MFP}} = 1.49^{+0.42}_{-0.57}$ pMpc, reconciles the differences between the direct and indirect measurements of the MFP at redshift $z \sim 6$ (Figure 6.10), establishing confidence in the measurements of the $\lambda_{\mathrm{MFP}}$ at high-redshifts.

## 7.1   Future Outlook

Looking ahead, upcoming large scale surveys using Rubin-LSST, Euclid and Roman hold promise for detecting hundreds more quasars at high redshifts. Subsequent follow-up observations with ground-based telescopes such as VLT, MMT and KECK will be crucial for probing the proximity zones of these quasars. In this context, the findings of this thesis raise several important questions, which we will discuss now. Our flickering quasar model for quasar prox-

imity zones discussed in Chapter 4 implies that a significant portion of SMBH growth must have occurred under obscuration. This prediction can be tested using JWST by looking for dust emission from the AGN torus using MIRI. JWST has already been detecting AGN-like objects called the Little Red Dots (LRDs), which can potentially be obscured quasars. Confirming their nature therefore is highly relevant for measuring the obscuration fraction (Kakiichi et al., 2024). Obscuration related to geometry can also be probed, for example, through the transverse proximity effect (e.g. Schmidt et al., 2018) and this needs to be extended to higher redshifts (Eilers et al., 2023). Further, determining the obscured QLFs will allow us to measure the radiative efficiency via the Soltan argument. It will also derive an upper limit on the contribution of quasars to reionization.

We also need to revisit the constraints on quasar lifetimes during the epoch of reionization. We demonstrated in Chapter 3 that quasar lifetimes inferred from their proximity zone sizes in a patchily ionized medium can exhibit a large scatter. Ideally, fitting to the full spectrum would extract more comprehensive information than relying solely on a single parameter, such as proximity zone size. However, large data covariances could introduce significant challenges in this approach. In an ongoing work (Satyavolu et al. in prep), we have tested the idea of using multiple definitions for quasar proximity zone sizes to infer quasar lifetimes by jointly fitting models corresponding to all definitions. We find that the joint posterior probability of quasar lifetimes with multiple $R_p$ definitions shows a decreased scatter on the individual lifetimes, but not in all cases. It however, remains, that most of the quasar lifetimes fall below the Salpeter timescale for an SMBH for varying radiative efficiencies and Eddington ratios. In the future, it is necessary to study proximity zones of quasars with variability consistent with the SMBH growth history as predicted by simulations, considering an environment around quasars that aligns with reionization observations. Given the high computational cost of performing black hole growth simulations for different seed masses and redshifts (e.g., Zhou et al., 2023), we believe in developing semi-analytic models for this purpose. Apart from the measurements using quasar proximity zone sizes, the lifetime measurements from clustering will be crucial, as we expect to detect thousands more quasars at high redshifts with the upcoming Rubin-LSST telescope.

Another interesting question that this thesis raised is the steep evolution of the mean free path (MFP) of ionizing photons. We find that the directs measurement of MFP are biased by the

correlations in the data. Hence, there is a need for a larger number of high-quality quasar data to accurately measure the MFP at high-$z$ in the future. On the theoretical side, it would be intriguing to investigate the role and nature of sinks responsible for setting the mean free path after the post-overlap phase of reionization, as some ongoing works are exploring through analytic or sub-grid modeling. In the future, it would be interesting to find solutions to bridge the gap between high-resolution, low-volume and low-resolution, high-volume simulations. Finally, a crucial piece of information missing about the IGM is its thermal state during reionization. While there have been several measurements of the volume-averaged hydrogen neutral fraction, constraints on the temperature at mean density are missing, as the usual probe for measuring it, the Ly$\alpha$ forest, is saturated during reionization. While we are working towards obtaining 21cm observations, one could already potentially investigate this problem using quasar proximity zones. For example, the transmitted flux within quasar proximity zones can be forward modeled for different thermal histories and compared to observations (see Davies & Hennawi (2023) for a proof of concept). This is because, in addition to natural line-broadening, the absorption lines within the proximity zones undergo thermal (Doppler) broadening due to line-of-sight velocities, as well as pressure smoothing. The latter depends on the integrated thermal history of the IGM, including heating due to the X-ray background, reionization, and quasar photoheating (if observed after long enough time scales), as well as dark matter free-streaming. Finally, the recent detection of potential damping wings around dark gaps in the IGM provides an opportunity for future work on them using simulations, offering new constraints on reionization.

In conclusion, given the rise in the number of high-redshift quasars and the availability of accurate reionization simulations, proximity zones will continue to be an invaluable probe for studying reionization and SMBH growth. The methodologies developed in this thesis, which include both simulations and the analysis of quasar proximity zones, will be crucial for these future endeavors.

# Bibliography

Abazajian K., et al., 2019, arXiv:1907.04473 [astro-ph.IM]

Abbott R., et al., 2020, ApJ, 900, L13

Acharya S. K., Majumdar S., 2023, MNRAS, 521, 2149

Ade P., et al., 2019, J. Cosmology Astropart. Phys., 2019, 056

Adelberger K. L., 2004, ApJ, 612, 706

Andika I. T., et al., 2020, ApJ, 903, 34

Anglés-Alcázar D., Davé R., Faucher-Giguère C.-A., Özel F., Hopkins P. F., 2017a, MNRAS, 464, 2840

Anglés-Alcázar D., Faucher-Giguère C.-A., Quataert E., Hopkins P. F., Feldmann R., Torrey P., Wetzel A., Kereš D., 2017b, MNRAS, 472, L109

Antonucci R., 1993, ARA&A, 31, 473

Arita J., et al., 2023, ApJ, 954, 210

Aubert D., Teyssier R., 2008, MNRAS, 387, 295

Aubert D., Teyssier R., 2010, ApJ, 724, 244

Bañados E., et al., 2016, ApJS, 227, 11

Bañados E., et al., 2018, Nature, 553, 473

Bañados E., et al., 2019, ApJ, 885, 59

Bañados E., et al., 2021, ApJ, 909, 80

Bañados E., et al., 2023, ApJS, 265, 29

Bahcall J. N., et al., 1993, ApJS, 87, 1

Bartos I., Kocsis B., Haiman Z., Márka S., 2017, ApJ, 835, 165

Bauer A., Springel V., Vogelsberger M., Genel S., Torrey P., Sijacki D., Nelson D., Hernquist L., 2015, MNRAS, 453, 3593

Baxter E. J., Weinberger L., Haehnelt M., Iršič V., Kulkarni G., Pandey S., Roy A., 2021, MNRAS, 501, 6215

Becker G. D., Bolton J. S., Haehnelt M. G., Sargent W. L. W., 2011, MNRAS, 410, 1096

Becker G. D., Bolton J. S., Lidz A., 2015a, PASA, 32, e045

Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015b, MNRAS, 447, 3402

Becker G. D., Davies F. B., Furlanetto S. R., Malkan M. A., Boera E., Douglass C., 2018, ApJ, 863, 92

Becker G. D., et al., 2019, ApJ, 883, 163

Becker G. D., D'Aloisio A., Christenson H. M., Zhu Y., Worseck G., Bolton J. S., 2021, MNRAS, 508, 1853

Beckmann R. S., Devriendt J., Slyz A., 2019, MNRAS, 483, 3488

Begelman M. C., 1979, MNRAS, 187, 237

Beifiori A., Courteau S., Corsini E. M., Zhu Y., 2012, MNRAS, 419, 2497

Bird S., Ni Y., Di Matteo T., Croft R., Feng Y., Chen N., 2022, MNRAS, 512, 3703

Bischetti M., et al., 2022, Nature, 605, 244

Blandford R. D., Begelman M. C., 2004, MNRAS, 349, 68

Boera E., Becker G. D., Bolton J. S., Nasir F., 2019, ApJ, 872, 101

Bolan P., et al., 2022, MNRAS, 517, 3263

Bolgar F., Eames E., Hottier C., Semelin B., 2018, MNRAS, 478, 5564

Bolton J. S., Becker G. D., 2009, MNRAS, 398, L26

Bolton J. S., Haehnelt M. G., 2007a, MNRAS, 374, 493

Bolton J. S., Haehnelt M. G., 2007b, MNRAS, 381, L35

Bolton J. S., Oh S. P., Furlanetto S. R., 2009, MNRAS, 395, 736

Bolton J. S., Haehnelt M. G., Warren S. J., Hewett P. C., Mortlock D. J., Venemans B. P., McMahon R. G., Simpson C., 2011, MNRAS, 416, L70

Bolton J. S., Becker G. D., Raskutti S., Wyithe J. S. B., Haehnelt M. G., Sargent W. L. W., 2012, MNRAS, 419, 2880

Bolton J. S., Puchwein E., Sijacki D., Haehnelt M. G., Kim T.-S., Meiksin A., Regan J. A., Viel M., 2017, MNRAS, 464, 897

Bondi H., 1952, MNRAS, 112, 195

121

Borisova E., Lilly S. J., Cantalupo S., Prochaska J. X., Rakic O., Worseck G., 2016, ApJ, 830, 120

Bosman S. E. I., 2021a, "All $z > 5.7$ quasars currently known", updated February 10, 2022, https://doi.org/10.5281/zenodo.5510200

Bosman S. E. I., 2021b, arXiv:2108.12446 [astro-ph.CO]

Bosman S. E. I., Kakiichi K., Meyer R. A., Gronke M., Laporte N., Ellis R. S., 2020, ApJ, 896, 49

Bosman S. E. I., et al., 2022, MNRAS, 514, 55

Bouwens R. J., et al., 2012, ApJ, 752, L5

Bromm V., 2013, Reports on Progress in Physics, 76, 112901

Bromm V., Yoshida N., 2011, ARA&A, 49, 373

Bruton S., Lin Y.-H., Scarlata C., Hayes M. J., 2023, ApJ, 949, L40

Buchner J., et al., 2015, ApJ, 802, 89

Bustamante S., Springel V., 2019, MNRAS, 490, 4133

Cain C., D'Aloisio A., Gangolli N., Becker G. D., 2021, ApJ, 917, L37

Calverley A. P., Becker G. D., Haehnelt M. G., Bolton J. S., 2011, MNRAS, 412, 2543

Cantalupo S., Arrigoni-Battaia F., Prochaska J. X., Hennawi J. F., Madau P., 2014, Nature, 506, 63

Carilli C. L., et al., 2010, ApJ, 714, 834

Carnall A. C., et al., 2015, MNRAS, 451, L16

Cen R., 1992, ApJS, 78, 341

Cen R., Haiman Z., 2000, ApJ, 542, L75

Chardin J., Haehnelt M. G., Aubert D., Puchwein E., 2015, MNRAS, 453, 2943

Chen H., Gnedin N. Y., 2018, ApJ, 868, 126

Chen H., Gnedin N. Y., 2021, ApJ, 911, 60

Chen X., Miralda-Escudé J., 2004, ApJ, 602, 1

Chen H., et al., 2022, ApJ, 931, 29

Chen N., Motohara K., Spitler L., Nakajima K., Terao Y., 2024, ApJ, 968, 32

Choquette J., Cline J. M., Cornell J. M., 2019, J. Cosmology Astropart. Phys., 2019, 036

Christenson H. M., Becker G. D., Furlanetto S. R., Davies F. B., Malkan M. A., Zhu Y., Boera E., Trapp A., 2021, ApJ, 923, 87

Ciardi B., Stoehr F., White S. D. M., 2003, MNRAS, 343, 1101

Coatman L., Hewett P. C., Banerji M., Richards G. T., 2016, MNRAS, 461, 647

Coil A. L., Hennawi J. F., Newman J. A., Cooper M. C., Davis M., 2007, ApJ, 654, 115

Cooper T. J., Simcoe R. A., Cooksey K. L., Bordoloi R., Miller D. R., Furesz G., Turner M. L., Bañados E., 2019, ApJ, 882, 77

Costa T., Sijacki D., Trenti M., Haehnelt M. G., 2014, MNRAS, 439, 2146

Croom S., et al., "AGN Physics from QSO Clustering", in *Astronomical Society of the Pacific Conference Series Vol. 311*, eds., Richards G. T., Hall P. B., San Francisco: Astronomical Society of the Pacific, 2004:457

Cyburt R. H., Fields B. D., Olive K. A., Yeh T.-H., 2016, Reviews of Modern Physics, 88, 015004

D'Aloisio A., McQuinn M., Trac H., 2015, ApJ, 813, L38

D'Aloisio A., McQuinn M., Davies F. B., Furlanetto S. R., 2018, MNRAS, 473, 560

D'Aloisio A., McQuinn M., Trac H., Cain C., Mesinger A., 2020, ApJ, 898, 149

D'Odorico V., et al., 2023b, MNRAS, 523, 1399

Davies R. L., et al., 2023b, MNRAS, 521, 314

Davies F. B., Furlanetto S. R., 2016, MNRAS, 460, 1328

Davies F. B., Hennawi J. F., 2023, arXiv:2312.06763 [astro-ph.CO]

Davies F. B., Furlanetto S. R., McQuinn M., 2016, MNRAS, 457, 3006

Davies F. B., Becker G. D., Furlanetto S. R., 2018b, ApJ, 860, 155

Davies F. B., et al., 2018c, ApJ, 864, 142

Davies F. B., et al., 2018d, ApJ, 864, 143

Davies F. B., Hennawi J. F., Eilers A.-C., 2019, ApJ, 884, L19

Davies F. B., Hennawi J. F., Eilers A.-C., 2020, MNRAS, 493, 1330

Davies F. B., Bosman S. E. I., Furlanetto S. R., Becker G. D., D'Aloisio A., 2021, ApJ, 918, L35

Davies R. L., et al., 2023, MNRAS, 521, 289

Davies F. B., et al., 2024, ApJ, 965, 134

Davis S. W., Laor A., 2011, ApJ, 728, 98

De Rosa A., et al., 2019, New A Rev., 86, 101525

Decarli R., et al., 2018, ApJ, 854, 97

Del Moro A., Alexander D. M., Bauer F. E., Daddi E., Kocevski D. D., Stanley F., McIntosh D. H., 2017, Frontiers in Astronomy and Space Sciences, 4, 67

Di Matteo T., Croft R. A. C., Feng Y., Waters D., Wilkins S., 2017, MNRAS, 467, 4243

Di Matteo T., Angles-Alcazar D., Shankar F., "Massive black holes in galactic nuclei: Theory and simulations", in *The Encyclopedia of Cosmology (Set 2): Black Holes*, ed., Z. Haiman, New Jersey: World Scientific, 2023.

DiPompeo M. A., Myers A. D., Hickox R. C., Geach J. E., Hainline K. N., 2014, MNRAS, 442, 3443

Dunlop J. S., et al., 2013, MNRAS, 432, 3520

Eilers A.-C., Davies F. B., Hennawi J. F., Prochaska J. X., Lukić Z., Mazzucchelli C., 2017, ApJ, 840, 24

Eilers A.-C., Davies F. B., Hennawi J. F., 2018a, ApJ, 864, 53

Eilers A.-C., Hennawi J. F., Davies F. B., 2018b, ApJ, 867, 30

Eilers A.-C., et al., 2020, ApJ, 900, 37

Eilers A.-C., Hennawi J. F., Davies F. B., Simcoe R. A., 2021, ApJ, 917, 38

Eilers A.-C., et al., 2023, *Mapping Quasar Light Echoes with Lyman-alpha Forest Tomography during the Epoch of Reionization*, JWST Proposal ID 3117

Eilers A.-C., et al., 2024, arXiv:2403.07986 [astro-ph.GA]

Ellis J., Wands D., 2023, arXiv:2312.13238v1 [astro-ph.CO]

Endsley R., Stark D. P., Whitler L., Topping M. W., Chen Z., Plat A., Chisholm J., Charlot S., 2023, MNRAS, 524, 2312

Fan X., et al., 2000, AJ, 120, 1167

Fan X., et al., 2006a, AJ, 131, 1203

Fan X., et al., 2006b, AJ, 132, 117

Fan X., et al., 2019, BAAS, 51, 121

Fan X., Bañados E., Simcoe R. A., 2023, ARA&A, 61, 373

Fanidakis N., Macciò A. V., Baugh C. M., Lacey C. G., Frenk C. S., 2013, MNRAS, 436, 315

Farina E. P., et al., 2022, ApJ, 941, 106

Faucher-Giguère C.-A., Lidz A., Hernquist L., Zaldarriaga M., 2008, ApJ, 688, 85

Fausey H. M., et al., 2024, arXiv:2403.13126 [astro-ph.HE]

Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, MNRAS, 455, 2778

Fernandez R., Bryan G. L., Haiman Z., Li M., 2014, MNRAS, 439, 3798

Ferrarese L., Merritt D., 2000, ApJ, 539, L9

Ferrarese L., et al., 2006, ApJ, 644, L21

Fields B., Sarkar S., 2006, arXiv:astro-ph/0601514v2

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Friedrich M. M., Mellema G., Iliev I. T., Shapiro P. R., 2012, MNRAS, 421, 2232

Fumagalli M., O'Meara J. M., Prochaska J. X., Worseck G., 2013, ApJ, 775, 78

Furlanetto S. R., 2006, MNRAS, 370, 1867

Furlanetto S. R., Oh S. P., 2005, MNRAS, 363, 1031

Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004, ApJ, 613, 1

Gabor J. M., Bournaud F., 2013, MNRAS, 434, 606

Gaikwad P., Choudhury T. R., Srianand R., Khaire V., 2018, MNRAS, 474, 2233

Gaikwad P., et al., 2020, MNRAS, 494, 5091

Gaikwad P., et al., 2023, MNRAS, 525, 4093

Garaldi E., Gnedin N. Y., Madau P., 2019, ApJ, 876, 31

Garaldi E., Kannan R., Smith A., Springel V., Pakmor R., Vogelsberger M., Hernquist L., 2022, MNRAS, 512, 4909

García-Vergara C., et al., 2022, ApJ, 927, 65

Ghara R., Mellema G., Giri S. K., Choudhury T. R., Datta K. K., Majumdar S., 2018, MNRAS, 476, 1741

Gnedin N. Y., Kaurov A. A., 2014, ApJ, 793, 30

Gnedin N. Y., Madau P., 2022, Living Reviews in Computational Astrophysics, 8, 3

Gnedin N. Y., Becker G. D., Fan X., 2017, ApJ, 841, 26

Gonçalves T. S., Steidel C. C., Pettini M., 2008, ApJ, 676, 816

Goto H., et al., 2021, ApJ, 923, 229

Greenstein J. L., 1963, Nature, 197, 1041

Greig B., Mesinger A., McGreer I. D., Gallerani S., Haiman Z., 2017a, MNRAS, 466, 1814

Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017b, MNRAS, 466, 4239

Greig B., Mesinger A., Bañados E., 2019, MNRAS, 484, 5094

Greig B., Mesinger A., Davies F. B., Wang F., Yang J., Hennawi J. F., 2022, MNRAS, 512, 5390

Greig B., et al., 2024, MNRAS, 530, 3208

Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633

Haardt F., Madau P., 1996, ApJ, 461, 20

Haardt F., Madau P., 2012, ApJ, 746, 125

Habouzit M., Volonteri M., Latif M., Dubois Y., Peirani S., 2016, MNRAS, 463, 529

Habouzit M., Volonteri M., Somerville R. S., Dubois Y., Peirani S., Pichon C., Devriendt J., 2019, MNRAS, 489, 1206

Hada R., Takada M., Inoue A. K., 2024, MNRAS, 531, 2912

Haiman Z., 2004, ApJ, 613, 36

Haiman Z., Quataert E., Bower G. C., 2004, ApJ, 612, 698

Hassan S., Davé R., McQuinn M., Somerville R. S., Keating L. C., Anglés-Alcázar D., Villaescusa-Navarro F., Spergel D. N., 2022, ApJ, 931, 62

Hayes M. J., Scarlata C., 2023, ApJ, 954, L14

Helton J. M., et al., 2024, arXiv:2405.18462 [astro-ph.GA]

Hennawi J. F., et al., 2006, ApJ, 651, 61

Hennawi J. F., Prochaska J. X., Cantalupo S., Arrigoni-Battaia F., 2015, Science, 348, 779

Hennawi J. F., Davies F. B., Wang F., Oñorbe J., 2021, MNRAS, 506, 2963

Hogan C. J., Anderson S. F., Rugers M. H., 1997, AJ, 113, 1495

Huang K.-W., Ni Y., Feng Y., Di Matteo T., 2020, MNRAS, 496, 1

Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27

Hutter A., Dayal P., Yepes G., Gottlöber S., Legrand L., Ucci G., 2021, MNRAS, 503, 3698

Igumenshchev I. V., Narayan R., Abramowicz M. A., 2003, ApJ, 592, 1042

Iliev I. T., Mellema G., Pen U. L., Merz H., Shapiro P. R., Alvarez M. A., 2006a, MNRAS, 369, 1625

Iliev I. T., et al., 2006b, MNRAS, 371, 1057

Iliev I. T., Pen U.-L., Bond J. R., Mellema G., Shapiro P. R., 2007, ApJ, 660, 933

Inayoshi K., Tanaka T. L., 2015, MNRAS, 450, 4350

Inayoshi K., Visbal E., Haiman Z., 2020, ARA&A, 58, 27

Inoue A. K., Shimizu I., Iwata I., Tanaka M., 2014, MNRAS, 442, 1805

Iršič V., et al., 2017, Phys. Rev. D, 96, 023522

Ishimoto R., et al., 2020, ApJ, 903, 60

Jeon J., Bromm V., Liu B., Finkelstein S. L., 2024, arXiv:2402.18773 [astro-ph.GA]

Jiang L., Fan X., Vestergaard M., Kurk J. D., Walter F., Kelly B. C., Strauss M. A., 2007, AJ, 134, 1150

Jiang Y.-F., Stone J. M., Davis S. W., 2014, ApJ, 796, 106

Jin X., et al., 2023, ApJ, 942, 59

Kakiichi K., et al., 2018, MNRAS, 479, 43

Kakiichi K., et al., 2024, *COSMOS-3D: A Legacy Spectroscopic/Imaging Survey of the Early Universe*, JWST Proposal ID 5893

Kauffmann G., Haehnelt M. G., 2002, MNRAS, 332, 529

Kaurov A. A., Gnedin N. Y., 2015, ApJ, 810, 154

Keating L. C., Haehnelt M. G., Cantalupo S., Puchwein E., 2015, MNRAS, 454, 681

Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020a, MNRAS, 491, 1736

Keating L. C., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020b, MNRAS, 497, 906

Keel W. C., et al., 2012, MNRAS, 420, 878

Khrykin I. S., Hennawi J. F., McQuinn M., Worseck G., 2016, ApJ, 824, 133

Khrykin I. S., Hennawi J. F., Worseck G., 2019, MNRAS, 484, 3897

Khrykin I. S., Hennawi J. F., Worseck G., Davies F. B., 2021, MNRAS, 505, 649

Kim S., et al., 2009, ApJ, 695, 809

King A., Nixon C., 2015, MNRAS, 453, L46

Kirkman D., Tytler D., 2008, MNRAS, 391, 1457

Kormendy J., Richstone D., 1995, ARA&A, 33, 581

127

Kulkarni G., Choudhury T. R., Puchwein E., Haehnelt M. G., 2016, MNRAS, 463, 2583

Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, MNRAS, 485, L24

LSST Science Collaboration et al., 2009, arXiv:0912.0201 [astro-ph.IM]

Larson R. L., et al., 2022, ApJ, 930, 104

Larson R. L., et al., 2023, ApJ, 953, L29

Laurent P., et al., 2017, J. Cosmology Astropart. Phys., 2017, 017

Lewis J. S. W., et al., 2022, MNRAS, 516, 3389

Lidz A., Heitmann K., Hui L., Habib S., Rauch M., Sargent W. L. W., 2006a, ApJ, 638, 27

Lidz A., Oh S. P., Furlanetto S. R., 2006b, ApJ, 639, L47

Lidz A., McQuinn M., Zaldarriaga M., Hernquist L., Dutta S., 2007, ApJ, 670, 39

Lusso E., Worseck G., Hennawi J. F., Prochaska J. X., Vignali C., Stern J., O'Meara J. M., 2015, MNRAS, 449, 4204

Ma Z., et al., 2024, arXiv:2406.04617 [astro-ph.GA]

Madau P., 2017, ApJ, 851, 50

Madau P., Rees M. J., 2001, ApJ, 551, L27

Madau P., Haardt F., Rees M. J., 1999, ApJ, 514, 648

Maiolino R., et al., 2023, arXiv:2308.01230 [astro-ph.GA]

Maiolino R., et al., 2024, Nature, 627, 59

Maity B., Choudhury T. R., 2023, MNRAS, 521, 4140

Malloy M., Lidz A., 2015, ApJ, 799, 179

Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, MNRAS, 351, 169

Martini P., Weinberg D. H., 2001, ApJ, 547, 12

Mascia S., et al., 2023, A&A, 672, A155

Maselli A., Gallerani S., Ferrara A., Choudhury T. R., 2007, MNRAS, 376, L34

Masi S., 2002, Progress in Particle and Nuclear Physics, 48, 243

Mason C. A., et al., 2019, MNRAS, 485, 3947

Massonneau W., Volonteri M., Dubois Y., Beckmann R. S., 2023, A&A, 670, A180

Matsuoka Y., et al., 2016, ApJ, 828, 26

Matsuoka Y., et al., 2018a, PASJ, 70, S35

Matsuoka Y., et al., 2018b, ApJS, 237, 5

Matsuoka Y., et al., 2019, ApJ, 883, 183

Matsuoka Y., et al., 2022, ApJS, 259, 18

Matthee J., et al., 2024, ApJ, 963, 129

Matthews T. A., Sandage A. R., 1963, ApJ, 138, 30

Mazzucchelli C., et al., 2017, ApJ, 849, 91

Mazzucchelli C., et al., 2023, A&A, 676, A71

McGreer I. D., Mesinger A., D'Odorico V., 2015, MNRAS, 447, 499

McQuinn M., 2016, ARA&A, 54, 313

McQuinn M., Oh S. P., Faucher-Giguère C.-A., 2011, ApJ, 743, 82

Mellema G., Iliev I. T., Alvarez M. A., Shapiro P. R., 2006, New A, 11, 374

Mesinger A., *Understanding the Epoch of cosmic Reionization: challenges and progress*, Cham: Springer International Publishing, 2016

Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955

Meyer R. A., et al., 2020, MNRAS, 494, 1560

Mignoli M., et al., 2020, A&A, 642, L1

Miralda-Escudé J., 1998, ApJ, 501, 15

Miralda-Escudé J., Rees M. J., 1998, ApJ, 497, 21

Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, ApJ, 530, 1

Mo H., van den Bosch F. C., White S., *Galaxy Formation and Evolution*, Cambridge: Cambridge University Press, 2010

Morales M. F., Wyithe J. S. B., 2010, ARA&A, 48, 127

Morales A. M., Mason C. A., Bruton S., Gronke M., Haardt F., Scarlata C., 2021, ApJ, 919, 120

Morey K. A., Eilers A.-C., Davies F. B., Hennawi J. F., Simcoe R. A., 2021, ApJ, 921, 88

Mortlock D. J., et al., 2009, A&A, 505, 97

Mortlock D. J., et al., 2011, Nature, 474, 616

Mostardi R. E., Shapley A. E., Steidel C. C., Trainor R. F., Reddy N. A., Siana B., 2015, ApJ, 810, 107

Muñoz J. A., Oh S. P., Davies F. B., Furlanetto S. R., 2016, MNRAS, 455, 1385

Mutch S. J., Geil P. M., Poole G. B., Angel P. W., Duffy A. R., Mesinger A., Wyithe J. S. B., 2016, MNRAS, 462, 250

Naidu R. P., et al., 2022, MNRAS, 510, 4582

Nanni R., Vignali C., Gilli R., Moretti A., Brandt W. N., 2017, A&A, 603, A128

Ni Y., Di Matteo T., Gilli R., Croft R. A. C., Feng Y., Norman C., 2020, MNRAS, 495, 2135

Novak G. S., Faber S. M., Dekel A., 2006, ApJ, 637, 96

Novak G. S., Ostriker J. P., Ciotti L., 2011, ApJ, 737, 26

Novikov I. D., Thorne K. S., "Astrophysics of Black Holes" in *Black Holes (Les Astres Occlus)*, eds., DeWitt, C., DeWitt, B. S., New York: Gordon and Breach, 1973:343–450

O'Meara J. M., Prochaska J. X., Burles S., Prochter G., Bernstein R. A., Burgess K. M., 2007, ApJ, 656, 666

O'Meara J. M., Prochaska J. X., Worseck G., Chen H.-W., Madau P., 2013, ApJ, 765, 137

Oh S. P., 2002, MNRAS, 336, 1021

Omukai K., Schneider R., Haiman Z., 2008, ApJ, 686, 801

Ono Y., et al., 2012, ApJ, 744, 83

Oppenheimer B. D., Segers M., Schaye J., Richings A. J., Crain R. A., 2018, MNRAS, 474, 4740

Osterbrock D. E., Ferland G. J., *Astrophysics of gaseous nebulae and active galactic nuclei*, California: University Science Books, 2006

Pâris I., et al., 2011, A&A, 530, A50

Park H., Lukić Z., Sexton J., Alvarez M. A., Shapiro P. R., 2024, ApJ, 969, 46

Peebles P. J. E., *Physical cosmology*, New Jersey: Princeton University Press, 1971

Peebles P. J. E., *Principles of Physical Cosmology*, New Jersey: Princeton University Press, 1993

Peirani S., de Freitas Pacheco J. A., 2008, Phys. Rev. D, 77, 064023

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

Peterson B. M., *An Introduction to Active Galactic Nuclei*, Cambridge: Cambridge University Press, 1997

Planck Collaboration 2020, A&A, 641, A6

Planck Collaboration XVI 2014, A&A, 571, A16

Pritchard J. R., Loeb A., 2012, Reports on Progress in Physics, 75, 086901

Prochaska J. X., Hennawi J. F., Herbert-Fort S., 2008, ApJ, 675, 1002

Prochaska J. X., Worseck G., O'Meara J. M., 2009, ApJ, 705, L113

Prochaska J. X., O'Meara J. M., Worseck G., 2010, ApJ, 718, 392

Puchwein E., et al., 2023, MNRAS, 519, 6162

Quataert E., Gruzinov A., 2000, ApJ, 539, 809

Rahmati A., Schaye J., 2018, MNRAS, 478, 5123

Reed S. L., et al., 2015, MNRAS, 454, 3952

Reed S. L., et al., 2017, MNRAS, 468, 4702

Regan J. A., Wise J. H., O'Shea B. W., Norman M. L., 2020, MNRAS, 492, 3021

Ren K., Trenti M., Marshall M. A., Di Matteo T., Ni Y., 2021, ApJ, 917, 89

Riechers D. A., et al., 2013, Nature, 496, 329

Robertson B. E., et al., 2013, ApJ, 768, 71

Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, ApJ, 802, L19

Romano M., Grazian A., Giallongo E., Cristiani S., Fontanot F., Boutsia K., Fiore F., Menci N.,
2019, A&A, 632, A45

Rosdahl J., Blaizot J., Aubert D., Stranex T., Teyssier R., 2013, MNRAS, 436, 2188

Roth J. T., D'Aloisio A., Cain C., Wilson B., Zhu Y., Becker G. D., 2024, MNRAS, 530, 5209

Rybicki G. B., Lightman A. P., *Radiative processes in astrophysics.*, New Jersey: Wiley, 1985

Salpeter E. E., 1964, ApJ, 140, 796

Satyavolu S., Kulkarni G., Keating L. C., Haehnelt M. G., 2023c, arXiv:2311.06344 [astro-
ph.CO]

Satyavolu S., Kulkarni G., Keating L. C., Haehnelt M. G., 2023a, MNRAS, 521, 3108

Satyavolu S., et al., 2023b, MNRAS, 522, 4918

Schawinski K., Koss M., Berney S., Sartori L. F., 2015, MNRAS, 451, 2517

Schindler J.-T., et al., 2020, ApJ, 905, 51

Schmidt M., 1963, Nature, 197, 1040

Schmidt T. M., Worseck G., Hennawi J. F., Prochaska J. X., Crighton N. H. M., 2017, ApJ, 847,
81

Schmidt T. M., Hennawi J. F., Worseck G., Davies F. B., Lukić Z., Oñorbe J., 2018, ApJ, 861, 122

Seljak U., et al., 2005, Phys. Rev. D, 71, 103515

Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337

Shankar F., Salucci P., Granato G. L., De Zotti G., Danese L., 2004, MNRAS, 354, 1020

Shankar F., Weinberg D. H., Miralda-Escudé J., 2009, ApJ, 690, 20

Shankar F., Weinberg D. H., Shen Y., 2010, MNRAS, 406, 1959

Shankar F., Marulli F., Mathur S., Bernardi M., Bournaud F., 2012, A&A, 540, A23

Shankar F., et al., 2016, MNRAS, 460, 3119

Shapiro S. L., Teukolsky S. A., *Black holes, white dwarfs, and neutron stars : the physics of compact objects*, New Jersey: Wiley, 1983

Shen Y., 2021, ApJ, 921, 70

Shen Y., et al., 2007, AJ, 133, 2222

Shen Y., et al., 2011, ApJS, 194, 45

Shen Y., et al., 2019, ApJ, 873, 35

Shukla H., Mellema G., Iliev I. T., Shapiro P. R., 2016, MNRAS, 458, 135

Siana B., et al., 2007, ApJ, 668, 62

Simcoe R. A., 2006, ApJ, 653, 977

Songaila A., Cowie L. L., 2010, ApJ, 721, 1448

Spina B., Bosman S. E. I., Davies F. B., Gaikwad P., Zhu Y., 2024, arXiv:2405.12273 [astro-ph.CO]

Springel V., 2005, MNRAS, 364, 1105

Springel V., et al., 2005, Nature, 435, 629

Stark D. P., 2016, ARA&A, 54, 761

Steidel C. C., Adelberger K. L., Shapley A. E., Pettini M., Dickinson M., Giavalisco M., 2000, ApJ, 532, 170

Stone M. A., Lyu J., Rieke G. H., Alberts S., Hainline K. N., 2024, ApJ, 964, 90

Storrie-Lombardi L. J., McMahon R. G., Irwin M. J., Hazard C., 1994, ApJ, 427, L13

Suzuki N., 2005, arXiv e-prints, arXiv:astro-ph/0503248

Takahashi H. R., Ohsuga K., 2015, PASJ, 67, 60

Tanabashi M., et al., 2018, Phys. Rev. D, 98, 030001

Tang M., Stark D. P., Chevallard J., Charlot S., 2019, MNRAS, 489, 2572

Tenneti A., Di Matteo T., Croft R., Garcia T., Feng Y., 2018, MNRAS, 474, 597

Tepper-García T., 2006, MNRAS, 369, 2025

Thorne K. S.,"Gravitational Waves" in *Particle and Nuclear Astrophysics and Cosmology in the Next Millenium*, eds. Kolb E. W., Peccei R. D., New Jersey: World Scientific, 1995

Trac H., Chen N., Holst I., Alvarez M. A., Cen R., 2022, ApJ, 927, 186

Trainor R., Steidel C. C., 2013, ApJ, 775, L3

Trakhtenbrot B., Volonteri M., Natarajan P., 2017, ApJ, 836, L1

Trebitsch M., Volonteri M., Dubois Y., Madau P., 2018, MNRAS, 478, 5607

Treister E., Krolik J. H., Dullemond C., 2008, ApJ, 679, 140

Uchiyama H., et al., 2018, PASJ, 70, S32

Umeda H., Ouchi M., Nakajima K., Harikane Y., Ono Y., Xu Y., Isobe Y., Zhang Y., 2023, arXiv:2306.00487 [astro-ph.GA]

Venemans B. P., et al., 2015, ApJ, 801, L11

Venemans B. P., et al., 2020, ApJ, 904, 130

Vernet J., et al., 2011, A&A, 536, A105

Vestergaard M., Osmer P. S., 2009, ApJ, 699, 800

Vestergaard M., Peterson B. M., 2006, ApJ, 641, 689

Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013, Phys. Rev. D, 88, 043502

Visbal E., Haiman Z., Bryan G. L., 2014, MNRAS, 442, L100

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, Nature Reviews Physics, 2, 42

Wang R., et al., 2010, ApJ, 714, 699

Wang R., et al., 2013, ApJ, 773, 44

Wang J., Koribalski B. S., Serra P., van der Hulst T., Roychowdhury S., Kamphuis P., Chengalur J. N., 2016, MNRAS, 460, 2143

Wang F., et al., 2020, ApJ, 896, 23

Wang F., et al., 2023, ApJ, 951, L4

Weinberg D. H., Katz N., Hernquist L., "Simulating Cosmic Structure Formation" in *Astronomical Society of the Pacific Conference Series Vol. 148*, eds., Woodward C. E., Shull J. M., Thronson Harley A. J., San Francisco: Astronomical Society of the Pacific, 1998:21

Weinberger L. H., Kulkarni G., Haehnelt M. G., Choudhury T. R., Puchwein E., 2018a, MNRAS, 479, 2564

Weinberger R., et al., 2018b, MNRAS, 479, 4056

Weinberger L. H., Haehnelt M. G., Kulkarni G., 2019, MNRAS, 485, 1350

Wellons S., et al., 2023, MNRAS, 520, 5394

White M., et al., 2012, MNRAS, 424, 933

Willott C. J., et al., 2007, AJ, 134, 2435

Willott C. J., et al., 2010, AJ, 140, 546

Wise J. H., Regan J. A., O'Shea B. W., Norman M. L., Downes T. P., Xu H., 2019, Nature, 566, 85

Wold I. G. B., et al., 2022, ApJ, 927, 36

Wolfson M., Hennawi J. F., Davies F. B., Oñorbe J., Hiss H., Lukić Z., 2021, MNRAS, 508, 5493

Wolfson M., Hennawi J. F., Davies F. B., Oñorbe J., 2023, MNRAS, 521, 4056

Worseck G., Wisotzki L., "The Transverse Proximity Effect in Spectral Hardness", in *Astronomical Society of the Pacific Conference Series Vol. 379, Cosmic Frontiers*, eds., Metcalfe N., Shanks T., San Francisco: Astronomical Society of the Pacific, 2007:235

Worseck G., et al., 2014, MNRAS, 445, 1745

Worseck G., Khrykin I. S., Hennawi J. F., Prochaska J. X., Farina E. P., 2021, MNRAS, 505, 5084

Wu X. B., Han J. L., Liu F. K., Zhang T. Z., " Supermassive Black Hole Masses and Bulge Luminosities in AGNs" in *8th Asian-Pacific Regional Meeting, Volume II.*, eds., Ikeuchi S., Hearnshaw J., Hanawa T., 2002:405–406

Wu X.-B., et al., 2015, Nature, 518, 512

Xu H., Ahn K., Wise J. H., Norman M. L., O'Shea B. W., 2014, ApJ, 791, 110

Yajima H., Ricotti M., Park K., Sugimura K., 2017, ApJ, 846, 3

Yang J., et al., 2020a, ApJ, 897, L14

Yang J., et al., 2020b, ApJ, 904, 26

Yang J., et al., 2021, ApJ, 923, 262

Yang J., et al., 2023, ApJS, 269, 27

Yoshida N., Abel T., Hernquist L., Sugiyama N., 2003, ApJ, 592, 645

Yu Q., Lu Y., 2005, ApJ, 620, 31

Yu Q., Tremaine S., 2002, MNRAS, 335, 965

Yue M., et al., 2024, ApJ, 966, 176

Zhou Y., Chen H., Di Matteo T., Ni Y., Croft R. A. C., Bird S., 2023, arXiv:2309.11571 [astro-ph.CO]

Zhu Y., et al., 2021, ApJ, 923, 223

Zhu Y., et al., 2022, ApJ, 932, 76

Zhu Y., et al., 2023b, ApJ, 955, 115

Zhu Y., et al., 2024, MNRAS, 533, L49

Ďurovčíková D., et al., 2024, ApJ, 969, 162

Šoltinský T., et al., 2021, MNRAS, 506, 5818

# Appendix A

# 1D Radiative Transfer Algorithm
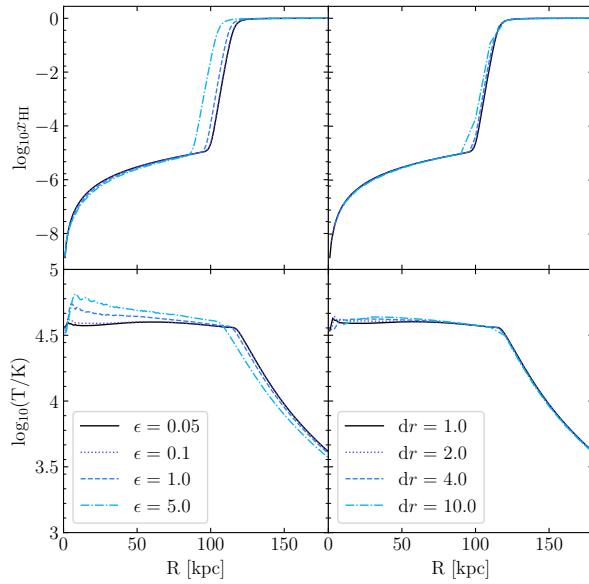
We describe our one-dimensional radiative transfer algorithm to solve the thermochemistry equations given by Equations (3.2.1)–(3.2.3) and (3.2.14). The algorithm employs a fixed user-specified global integration time-step. We exclude the first cell from the computation.

## A.1   Global integration
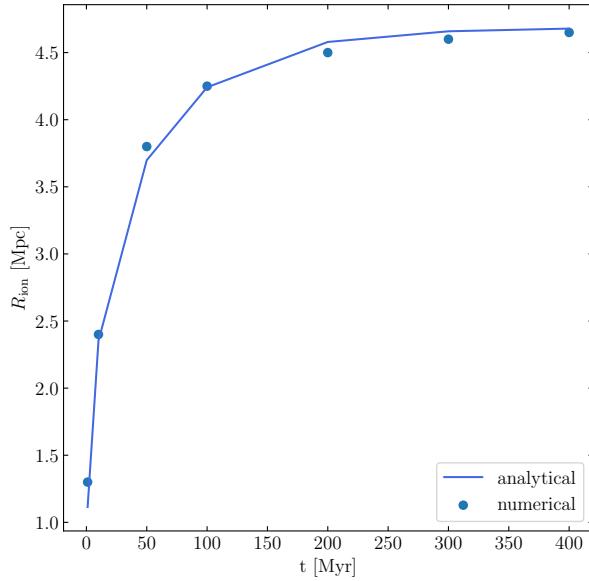
1. [Initialise grid.] Set up spatial grid and initialize all quantities of interest (densities, ionization fractions, and temperature) in all the grid cells.

2. [Fix global timestep.] Choose a global time-step $\Delta t$ based on the Courant criterion that depends on cell-crossing time of the ionization front and therefore cell size (cf. Bolton & Haehnelt, 2007a).

3. [Solve thermochemistry equations for global timestep.] Integrate the discretized thermochemistry equations, Equations (3.2.1)–(3.2.3) and (3.2.14), following algorithm A.2.

4. [Onto next global timestep until final time is reached.] Repeat step **C** with the same global time step until the specified end time is reached.

## A.2   Thermochemistry integration

At a global time step, the following algorithm is used to compute the thermochemistry. This computation is done simultaneously and independently in all cells of the domain (cf. Rosdahl

**Figure A.1:** Spatial and Temporal convergence: Top and bottom panels show the abundance of neutral hydrogen and temperature for different cell sizes on the right and different time steps on the left.



**Figure A.2:** Comparison between the numerical and analytical Stromgren solutions for ionized radius, for the test parameters described in Section B.2. Points represent numerical results obtained using our code; blue curve shows the analytical solution.

et al., 2013). Radiative quantities such as the photoionization rates, the photoheating rate, and the opacities, are computed when required.

1. [Integrate Equation (3.2.14)] Solve for temperature $T(t + \Delta t)$ using explicit Euler integration with a time step $dt = \Delta t$.

2. [Check if sub-cycling is required.] If $T(t + \Delta t) - T(t) > 0.1T(t)$, repeat step **A** with smaller time steps $dt = dt/2$.

3. [Get converged $T(t + \Delta t)$.] Repeat step **A** until the condition in step **B** is satisfied.

4. [Integrate Equation (3.2.1)] Solve for $n_{\text{HII}}(t + \Delta t)$ using an implicit Euler scheme, with a time step $dt = \Delta t$. Use $T(t + \Delta t)$ for computing rate coefficients.

5. [Check if sub-cycling is required.] If $n_{\text{HII}}(t + \Delta t) - n_{\text{HII}}(t) > 0.1n_{\text{HII}}(t)$, repeat step **D** with smaller time steps $dt = dt/2$.

6. [Get converged $n_{\text{HI}}(t + \Delta t)$.] Repeat step **D** until the condition in step **E** is satisfied.

7. [Integrate Equation (3.2.2)] Solve for $n_{\text{HeII}}(t + \Delta t)$ using an implicit Euler scheme, with a time step $dt = \Delta t$. Use $T(t + \Delta t)$ and $n_{\text{HII}}(t + \Delta t)$ for computing rate coefficients.

8. [Check if sub-cycling is required.] If $n_{\text{HeII}}(t + \Delta t) - n_{\text{HeII}}(t) > 0.1n_{\text{HeII}}(t)$, repeat step **G** with smaller time steps $dt = dt/2$.

9. [Get converged $n_{\text{HeII}}(t + \Delta t)$.] Repeat step **G** until the condition in step **H** is satisfied.

10. [Integrate Equation (3.2.3)] Solve for $n_{\text{HeIII}}(t + \Delta t)$ using an implicit Euler scheme, with a time step $dt = \Delta t$. Use the updated temperature, $n_{\text{HI}}$ and $n_{\text{HeII}}$ for computing rate coefficients.

11. [Check if sub-cycling is required.] If $n_{\text{HeIII}}(t + \Delta t) - n_{\text{HeIII}}(t) > 0.1n_{\text{HeIII}}(t)$, repeat step **J** with smaller $dt = dt/2$.

12. [Get converged $n_{\text{HeIII}}(t + \Delta t)$.] Repeat step **J** until the condition in step **K** is satisfied to get converged $n_{\text{HeIII}}$.

13. [Compute $n_{\text{e}}, n_{\text{HeI}}$.] Compute $n_{\text{HeI}} = n_{\text{He}} - n_{\text{HeII}} - n_{\text{HeIII}}$ and $n_{\text{e}} = n_{\text{HII}} + n_{\text{HeII}} + 2n_{\text{HeIII}}$.

## A.3 Discretization

Equations (3.2.1)–(3.2.3) are solved in terms of ionization fractions, $x_{\mathrm{HII}}$, $x_{\mathrm{HeII}}$, and $x_{\mathrm{HeIII}}$, instead of the number densities. For integrating Equation 3.2.1, we follow a semi-implicit numerical scheme as discussed in Rosdahl et al. (2013). The discretized equation looks like

$$x_{\mathrm{HII}}(t + \Delta t) = x_{\mathrm{HII}}(t) + \Delta t \frac{C - x_{\mathrm{HII}}(t)(C + D)}{1 - J\Delta t} \qquad \text{(A.3.1)}$$

where

$$J = \frac{\partial C}{\partial x_{\mathrm{HII}}} - (C + D) - x_{\mathrm{HI}} \left( \frac{\partial C}{\partial x_{\mathrm{HII}}} + \frac{\partial D}{\partial x_{\mathrm{HII}}} \right), \qquad \text{(A.3.2)}$$

and $C$ and $D$ are creation and destruction operators that can be read from Equation (3.2.1) after rearranging in terms of $x_{\mathrm{HII}}$ as

$$\frac{\mathrm{d}x_{\mathrm{HII}}}{\mathrm{d}t} = C - x_{\mathrm{HII}}(C + D). \qquad \text{(A.3.3)}$$

A similar semi-implicit Euler scheme is used for integrating Equations (3.2.2) and (3.2.3). The discretized equations look as follows

$$x_i(t + \Delta t) = \frac{x_i(t) + C\Delta t}{1 + D\Delta t} \qquad \text{(A.3.4)}$$

where i = HeII, HeIII. As before, $C$ and $D$ can be read from Equations (3.2.2) and (3.2.3) by rearranging them in the form

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = C - x_i D \qquad \text{(A.3.5)}$$

For integrating Equation (3.2.14), we use an Euler explicit integration scheme. The discretized equation is

$$T(t + \Delta t) = T(t) + \Delta t \, L \qquad \text{(A.3.6)}$$

where $L$ is the time derivative of temperature evaluated at the previous time step

$$L = \frac{2}{3} \frac{\mu m_{\mathrm{H}}}{\rho k_{\mathrm{B}}} (\mathcal{H} - \Lambda) - 2HT - \frac{T}{n} \frac{\mathrm{d}n}{\mathrm{d}t}, \qquad \text{(A.3.7)}$$

where all the quantities are the same as those in Equation (3.2.14).

139

## A.4 Source spectrum and photoionization rate

We assume the source spectrum to be a broken power law (Lusso et al., 2015) given by

$$
L_\nu \propto
\begin{cases}
\nu^{-0.61} & \text{if } \lambda \geq 912 \text{ Å}, \\
\nu^{-1.70} & \text{if } \lambda < 912 \text{ Å}
\end{cases}
\tag{A.4.1}
$$

For computing the photoionization rates in Equation (3.2.5), we divide the frequencies into 80 bins of equal logarithmic width from $\nu_{\text{HI}}$ to $40\nu_{\text{HI}}$, corresponding to energies between 13.6 eV and 544 eV. Increasing the frequency range or the number of bins does not have any effect on the results.

## A.5 Ly$\alpha$ optical depth

To calculate the Ly$\alpha$ optical depth, we assume a Voigt profile for the absorption cross-section. The optical depth in a given cell of size $dR$ along the line of sight is then calculated as (Weinberger et al., 2018a)

$$
\tau(i) = \frac{\nu_\alpha \sigma_\alpha dR}{\sqrt{\pi}} \sum_{j<i} \frac{n_{\text{HI}}(j)}{\Delta \nu_{\text{D}}(j)} H(a, x(i,j))
\tag{A.5.1}
$$

where $H(a,x)$ is an analytic approximation to the Voigt profile $\phi(\nu)$, related to it as (Tepper-García, 2006)

$$
\phi(\nu) = \nu_{\text{D}}^{-1} \pi^{-1/2} H(a,x),
\tag{A.5.2}
$$

and given by

$$
H(a,x) = \mathrm{e}^{-x^2} - \frac{a}{\sqrt{\pi}x^2}[\mathrm{e}^{-2x^2}(4x^4 + 7x^2 + 4 + 1.5x^{-2})
$$

$$
- 1.5x^{-2} - 1], \quad \text{(A.5.3)}
$$

where $\sigma_\alpha$ is the Lyman alpha cross-section, and $\nu_\alpha$ the corresponding frequency. The parameter $a$ is

$$
a = \frac{\Lambda_\alpha}{4\pi \Delta \nu_{\text{D}}}
\tag{A.5.4}
$$

where

$$\Delta \nu_\mathrm{D} \equiv \frac{\nu_\alpha}{c} \sqrt{\frac{2k_\mathrm{B}T}{m_\mathrm{H}}}, \tag{A.5.5}$$

and $\Lambda_\alpha$ is the hydrogen $2p \to 1s$ decay rate. $x(i,j)$ can be computed given the Hubble velocity $v_\mathrm{H}$ and peculiar velocity $v_\mathrm{pec}$ within the cell as

$$x(i,j) = \sqrt{\frac{m_\mathrm{H}}{2k_\mathrm{B}T}}[v_\mathrm{H}(i) - v_\mathrm{H}(j) - v_\mathrm{pec}(j)]. \tag{A.5.6}$$

# Appendix B

# Code Tests

## B.1   Convergence tests

To check the spatial and temporal convergence of our algorithm, we run a test from Bolton & Haehnelt (2007a). The neutral hydrogen abundance and temperature around a quasar emitting photons at $\dot{N} = 5 \times 10^{53}\,\mathrm{s}^{-1}$ and having a spectral index of 1.5 is computed after time $t = 1$ Myr. The quasar is assumed to be surrounded by a uniform density medium at redshift $z = 6$ with hydrogen and helium in primordial abundance ratio. Figure A.1 shows the results of this test repeated for spatial resolutions $\mathrm{d}r = 1, 2, 4, 10$ kpc and Courant factors $\epsilon = 0.05, 0.1, 1, 5$. The global time step is fixed by the Courant factor and the spatial resolution as (Bolton & Haehnelt, 2007a)

$$dt = 3261.6 \text{ yr } \left(\frac{\epsilon}{0.1}\right)\left(\frac{\mathrm{d}r}{10 \text{ kpc}}\right). \tag{B.1.1}$$

Right column of Figure A.1 shows that the ionization front position and temperature are nearly convergent as cell size becomes less than 2 kpc. Similarly, the left column of Figure A.1 shows that the ionization front position and temperature are nearly convergent for $\epsilon < 0.1$. Substituting these values in Equation (B.1.1) gives a global time step of $6.52 \times 10^{-4}$ Myr. This time step is four orders of magnitude smaller than the total run time ($\sim 1$ Myr). This can limit the computational time required. However, in practice, for cosmological conditions, we found that using a global time step that was smaller than the total run time by two orders of magnitude was sufficient to achieve convergence. Similarly, for the flickering quasar, the global time step was chosen to be one-tenth of the episodic time.

## B.2  Stromgren test

For a monochromatic source in an isothermal and initially neutral uniform hydrogen-only density field, the total number of photons within a shell can be accounted due to ionizations and radiative recombinations as

$$\frac{\mathrm{d}N}{\mathrm{d}t} = 4\pi R^2 n_{\mathrm{HII}} \frac{\mathrm{d}R}{\mathrm{d}t} + \frac{4\pi}{3} R^3 n_{\mathrm{HII}}^2 \alpha \tag{B.2.1}$$

Conversely, the rate of change in the ionized volume can be written as

$$\frac{\mathrm{d}R_{\mathrm{ion}}}{\mathrm{d}t} = \frac{\dot{N} - \frac{4\pi}{3} R^3 n_{\mathrm{HII}}^2 \alpha}{4\pi R^2 n_{\mathrm{HII}}} \tag{B.2.2}$$

Equation (B.2.2) can be solved analytically assuming uniform $n_{\mathrm{HII}} = n_{\mathrm{H}}$ as

$$R(t) = R_{\mathrm{s}} \left[ 1 - \exp\left(\frac{-t}{t_{\mathrm{rec}}}\right) \right]^{1/3} \tag{B.2.3}$$

where

$$R_{\mathrm{s}} \equiv \left( \frac{3\dot{N}}{4\pi \alpha n_{\mathrm{H}}^2} \right)^{1/3}, \tag{B.2.4}$$
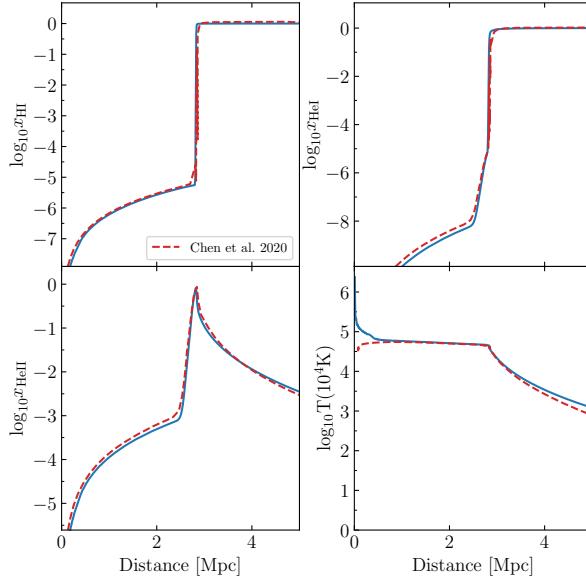
and

$$t_{\mathrm{rec}} = \frac{1}{\alpha n_{\mathrm{H}}}, \tag{B.2.5}$$

and $\alpha$ is the temperature-dependent recombination coefficient. We compare the results from our radiative transfer code with this analytical solution for the test problem 1 from the code comparison project by Iliev et al. (2006b) and find them to be in good agreement. The ionization front increases in size until the number of recombinations balance out the photoionizations, at which point, the ionization front radius saturates to the Stromgren value $R_{\mathrm{s}}$ as shown in Figure A.2.

## B.3  Temperature and Helium evolution

Analytical tests are not available for the general case in which we solve for the combined hydrogen and helium thermochemistry. Instead, we test the code by comparing to one-dimensional

**Figure B.1:** Comparison between results from our code and those from Chen & Gnedin (2021) for the test described in Section B.3.1. The panels show neutral hydrogen fraction, single and doubly ionised helium fractions and temperature. Blue curves are from our code. Red curves are from Chen & Gnedin (2021).
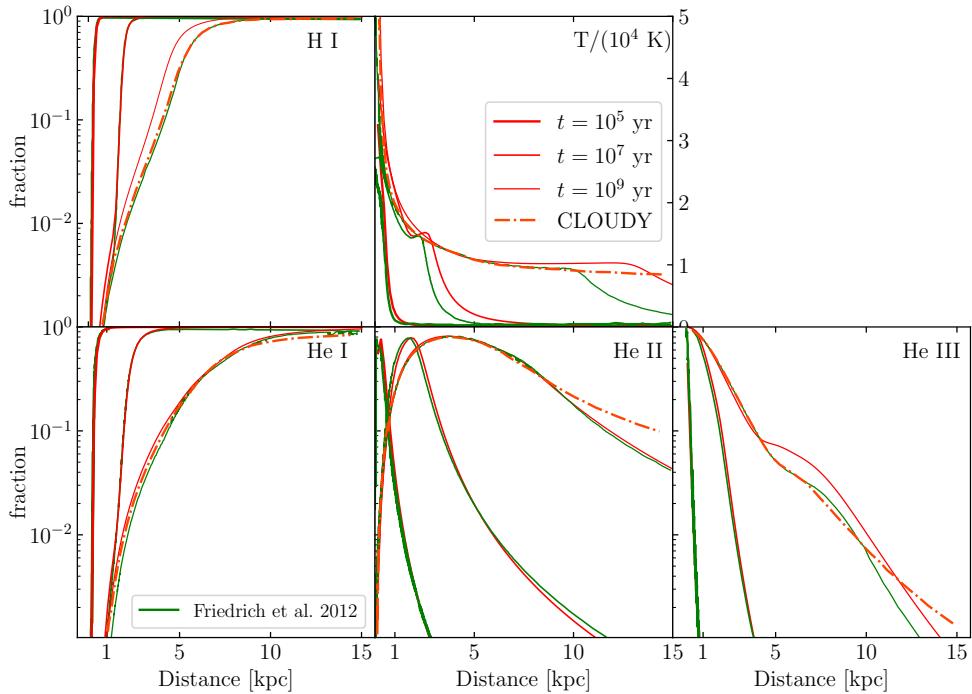
and three-dimensional results from the literature.

## B.3.1  Comparison with one-dimensional results by Chen & Gnedin (2021)

We solve the thermochemistry and temperature evolution equations in one dimension for a test case from Chen & Gnedin (2021) with the following parameters. The quasar is assumed to emit photons at a rate of $\dot{N} = 10^{57}\,\mathrm{s}^{-1}$ in a uniform density medium of hydrogen and helium with $X = 0.76$ and $Y = 0.24$ at a redshift of $z = 7$. The source specific luminosity is assumed to be power law with spectral index 1.5. Hydrogen and helium are assumed to be completely neutral initially and the initial temperature before the quasar is turned on is assumed to be 100 K. Figure B.1 shows the neutral hydrogen and ionized helium fractions, as well as temperature at a quasar age of 10 Myr. Our results shown in blue are in good agreement with the results of Chen & Gnedin (2021), shown in red.

## B.3.2  Comparison with three-dimensional results by Friedrich et al. (2012)

We compare our results from our one-dimensional radiative transfer code with those obtained using the three-dimensional radiative transfer code C2-RAY by Friedrich et al. (2012). The quasar is assumed to emit photons at a rate of $\dot{N} = 5 \times 10^{48}\mathrm{s}^{-1}$ in a uniform density medium of
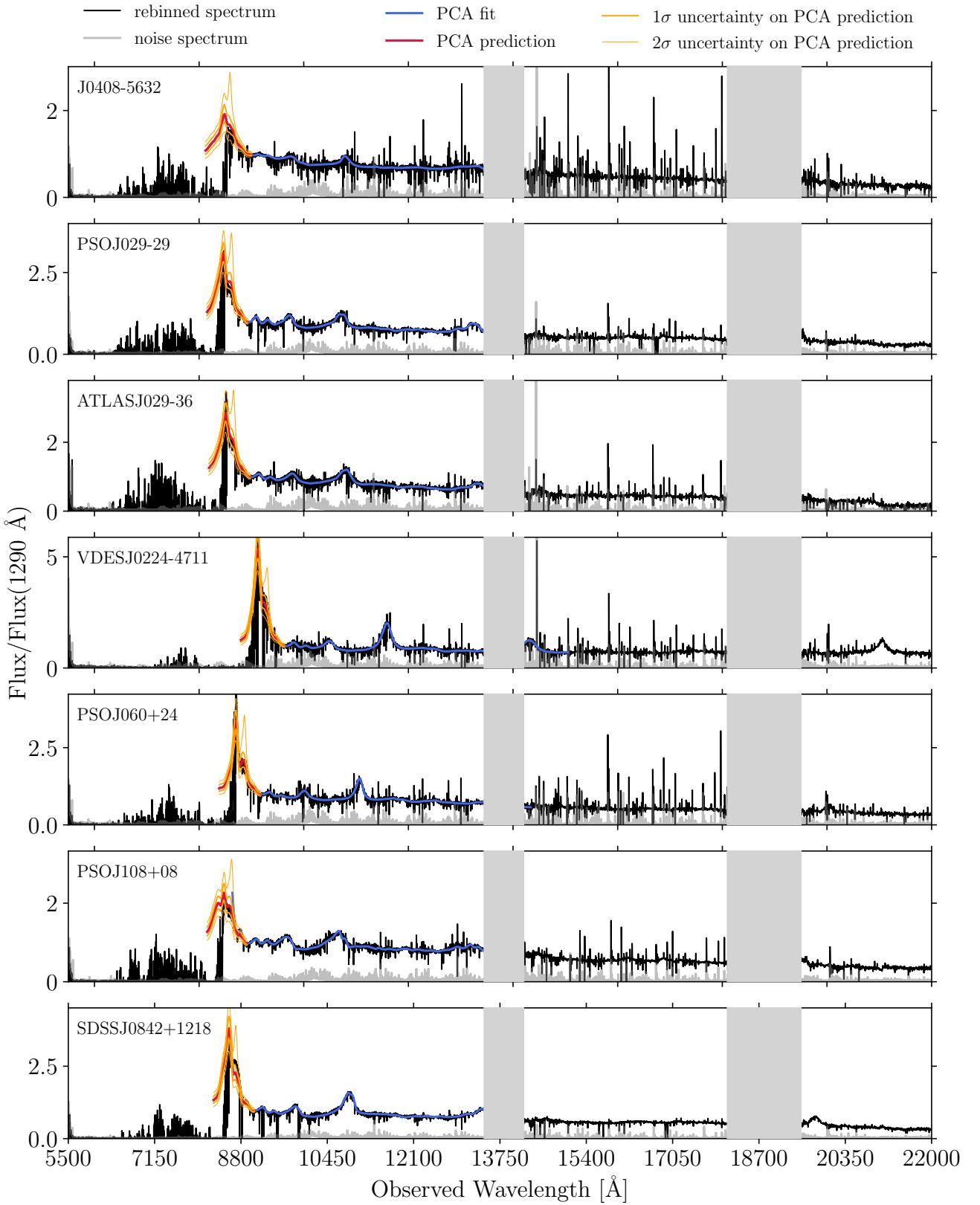
**Figure B.2:** Comparison between results from our code and those from Friedrich et al. (2012) for the test described in Section B.3.2. The panels show neutral hydrogen fraction, single, doubly, and triply ionised helium fractions and temperature. Red curves are from our code, green curves are from Friedrich et al. (2012). The dot-dashed curves are from CLOUDY. The thickness of the curves indicates quasar age, as described in the legend.

hydrogen and helium, with $n_{\mathrm{H}} = 10^{-3}\,\mathrm{cm}^{-3}$ and $n_{\mathrm{He}} = 8.7 \times 10^{-5}\,\mathrm{cm}^{-3}$, and an initial temperature of 100 K. The source specific luminosity is assumed to be power law with spectral index 1. Figure B.2 shows our results (in orange) for the neutral hydrogen fraction, helium ionization fractions, and gas temperature at quasar ages of $10^5, 10^7$, and $10^9$ yr against the results from C2-RAY (in green). Also shown are the equilibrium solutions from CLOUDY (Friedrich et al., 2012). Our results agree very well with the three-dimensional code at $10^5$ and $10^7$ yr. At $10^9$ yr, the helium ionization fractions match very well, but there is a small difference between the hydrogen ionization fronts and temperatures computed by the two codes, potentially due to secondary ionizations that we do not include in our code.
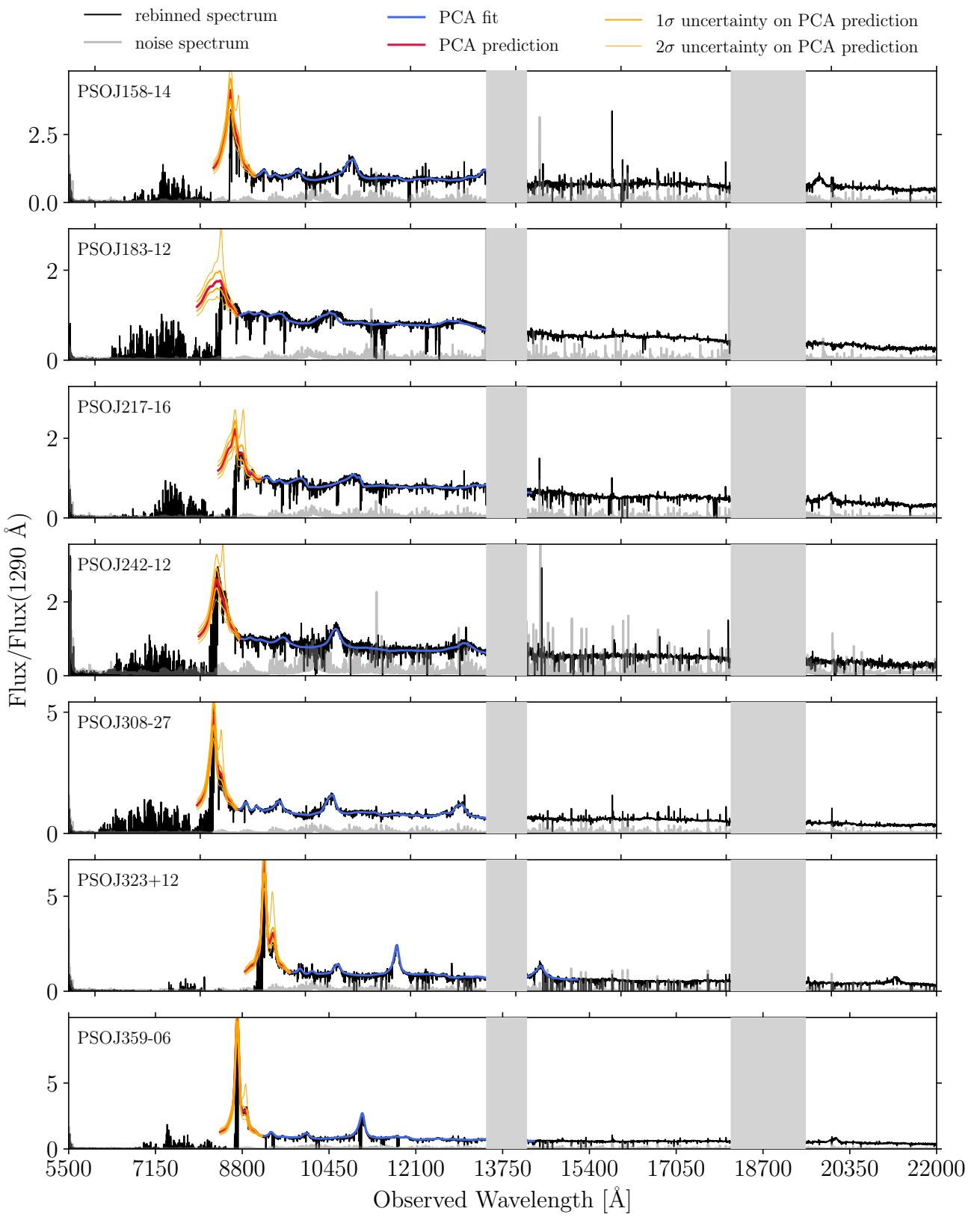
# Appendix C

# XQR-30 Quasar Spectra

Figure C.1 shows all 22 quasars from the extended XQR-30 sample used in this thesis. The flux has been rescaled by the flux at 1290 Åand rebinned to 50 km/s. Also shown is our model for the intrinsic continuum of the blue-side of the Ly$\alpha$ emission line of the quasars (the median predication is shown in red), relevant for proximity zone size measurements. The intrinsic spectrum is reconstructed with the help of low-redshift quasar spectrum. Therefore, the red-side fit to the intrinsic continuum (shown in blue) does not extend to wavalengths beyond the telluric absorption bands (shown in grey), as the training sample of the eBOSS quasars used does not have coverage at such wavelengths. Principal Component Analysis (PCA) is used to reconstruct $\mathcal{F}_{\mathrm{cont}}(\lambda)$. PCA works as follows: given the initial data set, the principal components are constructed by computing the eigenvectors of the covariance matrix of the initial dataset. The principal components together with the feature vector form a basis that corresponds to the least number of uncorrelated variables that can describe the initial data. By training the PCA model on low-$z$ quasar spectra, one can compute linear coefficients of both the red-side and blue-side of the spectrum and determine an optimal mapping between the linear coefficients of the two sides' decomposition. This model can then be used to reconstruct the 'unknown' blue-side of the spectrum ($\lambda < 1220$Å) from the 'known' red-side of the spectrum ($\lambda > 1280$ Å) of our high-redshift quasar sample. More details of the continuum estimation that was used in our analysis is described in Bosman et al. (2022) and D'Odorico et al. (2023b).

**Figure C.1:** All spectra studied in this thesis. Rescaled flux is obtained by dividing with the flux at 1290 Å. This rescaled flux, rebinned to 50 km/s, is shown in black. Uncertainty on the flux is shown in grey. Telluric absorption bands are shown in light grey. Best fit continuum to the redside flux is shown in blue. Predicted blue-side continuum is shown in red. 1 and 2 $\sigma$ errors on the continuum predictions are shown in orange.

147

**Figure C.1**– *continued*

**Figure C.1**– *continued*

149
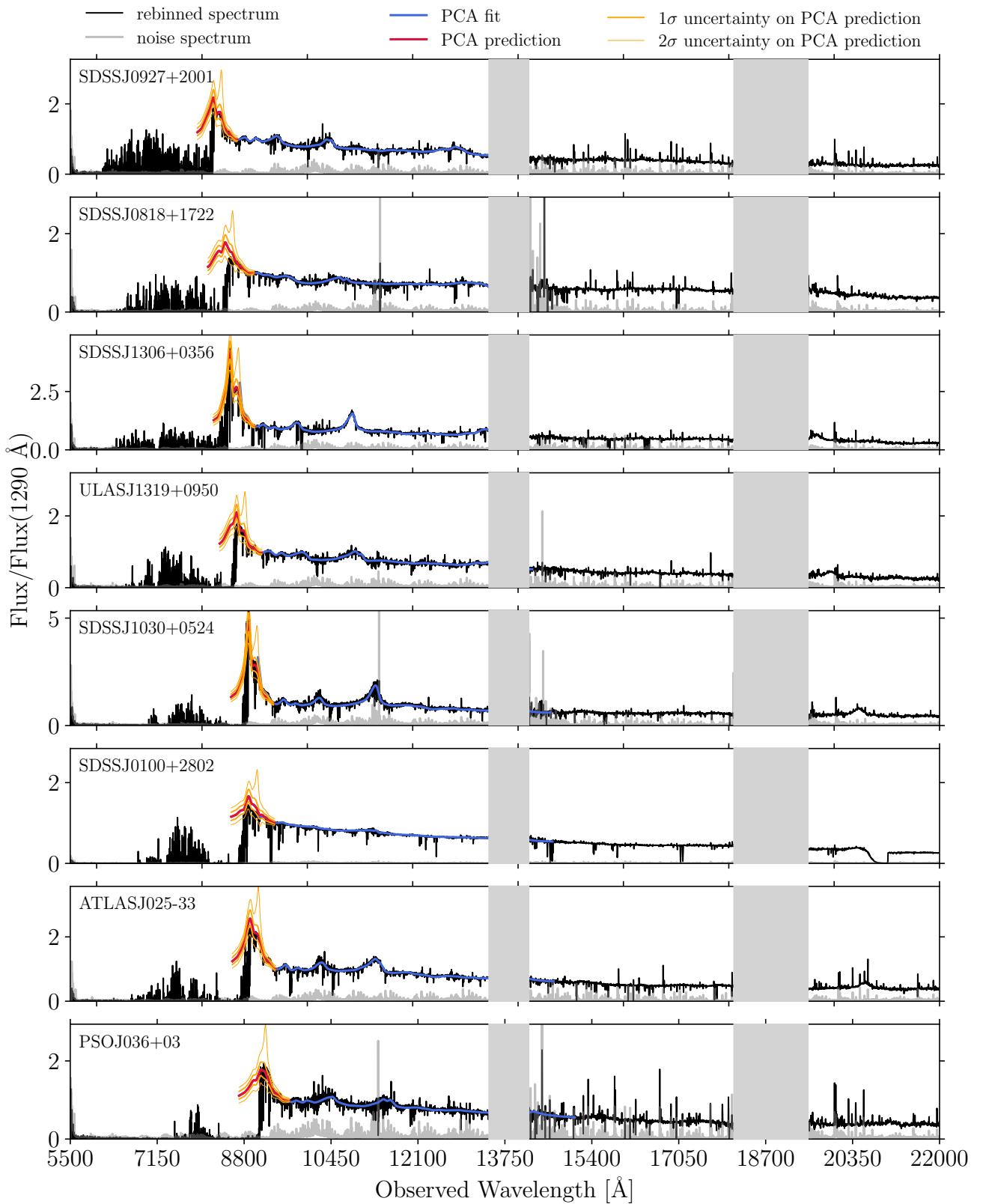
# Appendix D
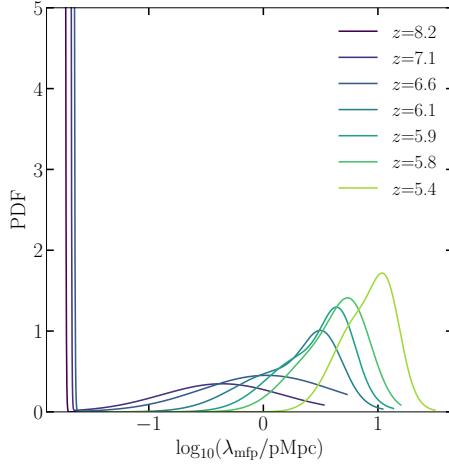
# MFP Distribution and Effect of Covariance Matrix on the Measurements

## D.1  Distribution of free paths
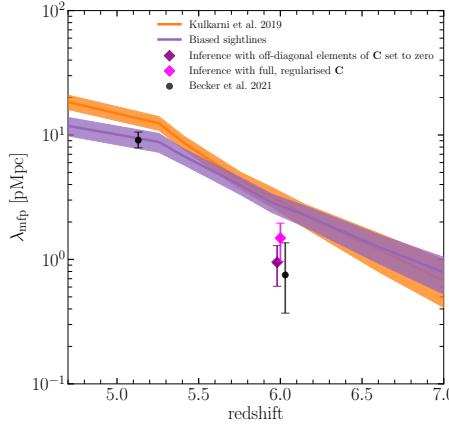
Figure D.1 shows the distribution of free paths in our simulation between redshifts 8 and 5. One part of the distribution describes the mean free path in neutral regions that are of average density. The sightlines with larger free paths on the other hand also include those which encounter neutral islands along the path. The fraction of sightlines with small free paths becomes smaller than those with larger free paths at lower redshifts, until the free path distribution becomes unimodal Gaussian. The divergence seen at redshifts $z < 6$ are numerical artefacts arising from the short free paths limited by the resolution of our simulation.

## D.2  Effect of off-diagonal terms in the covariance matrix

To understand the effect of the off-diagonal terms in the covariance matrix, we repeat our analysis of Section 6.4 by fitting our model to the stacked data of B21 using only the diagonal elements of the data covariance matrix provided by B21. The full, unregularised, data covariance matrix is nearly diagonal, but has some non-zero off-diagonal elements, particularly around $\sim 912$Å. Setting all off-diagonal terms to zero results in an inference shown by the purple diamond in Figure D.2. The magenta diamond in this figure, on the other hand, shows the result of the analysis from Section 6.4. We see that the off-diagonal elements lead to a larger mean

**Figure D.1:** Distribution of free paths across redshifts. At redshifts $z < 6$, the free path length distribution becomes bimodal with a large number of free paths $< 0.01$ pMpc, smaller than the resolution of our base simulation at these redshifts.



**Figure D.2:** Effect of the off-diagonal terms in the covariance matrix $\mathbf{C}$ on our inferred value of the mean free path. Magenta diamond shows the mean of the posterior distribution computed using the full, regularised covariance mastrix as described in the text. Also shown is the one-sigma uncertainty. The purple diamond shows the same for the case in which only the diagonal elements of the covariance matrix are used while computing the likelihood. In this case the off-diagonal terms are set to zero. The black points show the B21 measurements. We see that the off-diagonal elements lead to a larger mean value of the posterior distribution and a narrower posterior distribution, resulting in lower $1\sigma$ uncertainty. The three MFP measurements shown at $z \sim 6$ are at $z = 6$ but we have shown them with small displacements in the redshift directions for better legibility.

value of the posterior distribution and a narrower posterior distribution, resulting in lower $1\sigma$

uncertainty.