Constraining the neutral fraction of the IGM using quasar absorption spectroscopy

M.Sc Thesis

By
Manpreet Singh

Supervisor: Dr. Prakash Gaikwad



DEPARTMENT OF ASTRONOMY, ASTROPHYSICS & SPACE ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

November 2025

Constraining the neutral fraction of the IGM using quasar absorption spectroscopy

A THESIS

Submitted in partial fulfillment of the requirements for the awards of the degree
of
Master of Science

by

Manpreet Singh



DEPARTMENT OF ASTRONOMY, ASTROPHYSICS & SPACE ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

November 2025

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **Constraining the Neutral Fraction in the IGM using Quasar Absorption Spectroscopy** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF SCIENCE** and submitted in the **DEPARTMENT OF Astronomy, Astrophysics and Space Engineering, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from **July 2024** to **May 2026** under the supervision of **Dr Prakash Gaikwad, Assistant Professor, Indian Institute of Technology Indore** .

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Manpreet Singh 07/05/26

Signature of the student with date
(MANPREET SINGH)

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

(07 May 2026)

Signature of the Supervisor of

M.Sc. thesis (with date)

(NAME OF SUPERVISOR)

Manpreet Singh has successfully given his M.Sc. Oral Examination held on **05/05/26**.

Signature(s) of Supervisor(s) of MSc thesis
Date: 07 May 2026

Manoneeta Chakrabarty
Programme Coordinator, M.Sc.
Date: 08/05/2026

Manoneeta Chakrabarty
Convenor, DPGC
Date: 08/05/2026

HoD, DAASE
Date: 08/05/2026

Acknowledgements

I acknowledge my supervisor, Dr Prakash Gaikwad, for his constant guiding throughout my thesis work. His insights in the field are unmatched to anyone from my perspective. He explains even the most complex topics in a simplistic manner which is even understandable for me. The progress of this work has been at quite a steady pace due to his guidance. His supervision has truly made the research work feel like something meaningful and unique.

I also thank Dr Soumavo Ghosh and Dr Deepika Bollimpali for teaching galaxy formation and radiative processes, their insights have also cleared a lot of things on my mind. I thank Dr Narendranath Patra for his insights on fourier transforms, radio interferometry and his passion for astronomy has always motivated me to do more better. I thank teachers like Dr Biman Nath and Dr Suman Majumdar who have taught me the basics of cosmology, ignited a spark of interest in what I will seem to dedicate my life to. I also thank Dr Bhargav Vaidya, Dr Amit Shukla and Dr Manoneeta Chakraboty for teaching astrophysical fluids, plasma and gravitational physics. I am indebted to the professors for teaching me basics of astronomy and astrophysics.

I always love to have discussions with my labmate Kartik and his insights on the topic has always been helpful. Our discussions always leave both of us with open questions at the end of the day. Nityananda has always been like a second brother to me with whom I have been enjoying the coffee discussions for most evenings.

Most of them all, I am indebted to my parents who gave me the luxury to do research, and didn't stop me from chasing my passion. No words can describe my love for them. This whole thesis is dedicated to them.

ABSTRACT

The Epoch of Reionization (EoR) connects the observables to the Early Universe, providing important insights about ionization and thermal state of the Universe. Intergalactic medium (IGM) is ionized by the first luminous sources such as young galaxies and its properties can reveal timing, topology and physics of EoR. Direct measurements of observables during EoR becomes difficult primarily due to the IGM being opaque to Lyman α transmission due to high absorption,the patchiness of the reionization and even other faint 21 cm observations being dominated by the foreground effects. We aim to find whether the saturated regions arise from the UVB fluctuations or truly neutral islands.

In this work, we investigate the ionization state of the IGM using dark gaps present in the quasar spectra. We consider the dark gaps are defined as the continuous regions between two successive transmission peaks lying below normalized flux $F_{\rm norm} < 0.05$ arising due to strong absorption by neutral hydrogen present along line of sight. We consider the minimum length of the dark gap as $1\,{\rm h}^{-1}$Mpc and also rebin the spectra to $1\,{\rm h}^{-1}$Mpc. Using EXCITE simulations, we explore three models: Early, Late, Ultra-Late Model and we also define a scaling parameter which scales the mean photoionization rate to explore larger parameter space of neutral fraction i.e. $\Gamma_{12}(x) = f_\Gamma \langle \Gamma_{12} \rangle$. We investigate the dark gap distributions using Dark Gap Statistics such as Cumulative Distribution (CDF), Probability Distribution Function (PDF), and F10 Statistics(the cumulative distribution of the dark gaps longer $10\,{\rm h}^{-1}$Mpc). We show that the long dark gaps are primarily sensitive to the mean photoionization rate rather than the reionization models whereas short dark gaps are affected by the fluctuations highlighting a degeneracy between the UVB fluctuations and neutral islands.

By fitting the PDF using a χ^2 minimization, finding optimized f_Γ and marginalizing over reionization models, we obtain 1σ upper limits on volume-averaged neutral fraction as: $\langle x_{\rm HI} \rangle \leq 1.4 \times 10^{-4}$, 2.76×10^{-3}, 4.705×10^{-2} (Late); 4.0×10^{-5}, 9.0×10^{-5}, 3.55×10^{-3} (Early); and 7.9×10^{-4}, 1.423×10^{-2}, 8.398×10^{-2} (Ultra-Late), at $\bar{z} = 5.177$, 5.455, 5.769 respectively. Although we get consistent $\langle x_{\rm HI} \rangle$ values with the observations, these model-dependent upper bound constrains indicate a strong degeneracy between the neutral fraction and reionization histories.

Contents

List of Figures i

1 Introduction 1

2 Observational Probes of the EoR 6

3 Modeling Quasar Absorption Spectra 11
3.1 Ionization State: 13
3.2 Generating Lyman α Forest Code: 14

4 Observations & Simulations 17
4.1 Observation Data-Set 17
4.2 Simulations & Model Suites 18

5 Methods 23

6 Results 34

7 Conclusion and Discussion 42

8 Future Prospects 44

9 Appendix A: Effect of the threshold on the binned PDFs 45

10 Appendix B: Finding True Mean Free Path 47

List of Figures

1.1 Historical Timeline of the Universe, starting from last scattering surface to present day. Credits: Robertson (2022) 1

1.2 Schematic illustration taken from (Chakraborty and Choudhury, 2026) representing the three phases of reionization: Pre-Overlap, Overlap and Post -overlap phase. The ionized bubbles are believed to percolate and coalesce with other bubbles until the IGM is reionized. The final phase represents the tail end of reionization. 3

1.3 Schematic Illustration taken from McQuinn (2016) explaining the thermal and ionization history of the universe at a multi-wavelength scale. Rapid evolution of the temperature and neutral fraction suggests a rapid evolution of the parameters due to drastic change from cold dark ages to Epoch of Reionization. . . 4

3.1 Shape of absorption profile for a single pixel absorber. Right to Left: As the neutral hydrogen density increases, the flux absorption profile becomes more broader and we get a clear strong damping wing for $n_H = 2.5 \times 10^{-2}$. The Voigt profile is dominated by Gaussian profile for low nH values, whereas in high nH values,damping wings are visible i.e., Lorentzian profile dominates. 12

3.2 Growth Curve: Equivalent Width plotted as a function of column density N_H with changing thermal parameter. The strength of absorptioin line depends on oscillator stength,column density and thermal parameter,b. There are different regimes of the column density where the strength of absorption behaves differently. It is affected by the thermal parameter only in saturated regions. 13

3.3 Neutral fraction evolution with lookback time, illustrating the reionization process; taken from Kulkarni et al. (2019) 14

3.4 Synthetic Ly α forest generated using LOS optical depth 3.2. This is compared with ex-cite simulations (Gaikwad et al., 2023) and the match makes sure that code is working properly. 15

4.1 The slices of the simulation models plotted for photoionization fluctuations and neutral fraction at $z = 5.95$. Left to Right: Early Model (1σ lower limit of τ_{eff}), Best Fit (Fiducial) Model, Ultra-Late Model (1σ upper limit of τ_{eff}). 20

4.2 Evolution of neutral fraction with redshift across different models. The red dotted curve shows the Ultra-Late Model, orange one shows the Late Model (fiducial), and green shows Early Model. The observations include from damping wing analysis (Greig et al., 2017; Bañados et al., 2018; Greig et al., 2019; Wang et al., 2020; Yang et al., 2020), Lyman-α forest analysis (Spina et al., 2024; Zhu et al., 2024), and dark pixel fraction (McGreer et al., 2011; Mason et al., 2018; Jin et al., 2023; Davies et al., 2026). 21

5.1 Dark gap detection steps illustrated in an example sightline,, at $z = 5.11 \rightarrow 5.26$. The steps involve matching the resolution with the instrument and then the orange curve represents the continuum normalized flux obtained from the synthetic spectra, and it is smoothed to $1\,h^{-1}$Mpc. Shaded regions represent the dark gaps obtained in the sightline, and the red dashed line represents the threshold $F = 0.05$. 24

5.2 Dark gap detection is illustrated in an example sightline, PSO J183-12, at $z = 5.97$ with a signal-to-noise ratio of 61.8. The orange curve represents the continuum normalized flux obtained from the synthetic spectra, and it is smoothed to $1\,h^{-1}$Mpc. Shaded regions represent the dark gaps obtained in the sightline, and the red dashed line represents the threshold $F = 0.05$. 25

5.3 Dark gap lengths as a function of redshift z for the Ultra-Late reionization scenario. Each point represents an individual dark gap identified across all sightlines. The black dashed line marks the threshold $L = 10\,h^{-1}$ Mpc used to define long dark gaps. Longer dark gaps become increasingly prominent toward higher redshift. 26

5.4 Cumulative Distribution Function of the dark gaps plotted as a function of their lengths. The dotted lines represents the data points for the fiducial model and solid lines are the observations (Zhu et al., 2021). This plot serves an χ by eye-test to check whether the models are able to produce the observations. 27

5.5 The global PDF of different reionization histories across redshift $5.0 \leq z \leq 6.0$ as a function of photoionization rate.Left to Right: The scaling values change from $f_\Gamma = 0.3 \rightarrow 3.0$. The mean photoionization rate for this redshift range is taken to be $\Gamma_{12} = 0.35$. The dashed lines represent the observations (Zhu et al., 2021) and the solid lines indicate different models based on z_{50} (Ultra Late: Blue, Late(Best-Fit) Model: Orange, Early Model: Green Solid lines respectively) with error bars representing the 68% uncertainties on the PDF. 28

5.6 The KDE plotted for the distribution of the dark gaps of different models across different redshift models. The dotted line represent the observations (Zhu et al., 2021) whereas the solid lines represent the simulation KDE curves. We get suppressed f_Γ in second redshift bin due to averaging of the photoionization rate due to high opacity fluctuations via rapid evolution of neutral fraction in this redshift bin. 29

5.7 Cumulative F30 Statistics Distribution for different reionization models and mean photoionization rate (Left to Right: Ultra Late Model, Late Model, Early Model). The long dark gaps are sensitive to the mean photoionization rate and the observation is well within the extreme values of the mean photoionization rates. All the reionization histories show similar distribution, strengthening the argument that longer dark gaps are not affected much by the reionization histories. 30

5.8 F10 Statistics evolving with the redshift for different reionization histories and different mean photoionization rates(Left to Right: Ultra Late Model, Late Model, Early Model). This acts as an eye-test we see that all the models produce similar shape as the observations (Zhu et al., 2021). 31

6.1 The D-Value Distribution plotted across different mocks for different redshift bins and models. The tightness and low value of KS statistic D are indicators of the closeness of the models to the observations. 34

6.2 Median value of the D distribution across different redshift bins. The preference of the models can be easily inferred: For $5.0 \leq z \leq 5.3$: Early Model; For $5.3 \leq z \leq 5.6$: Early Model; For $5.6 \leq z \leq 6.0$: Late (Best-Fit) Model 35

6.3 Minimum χ^2 curve plotted for different models over $5.0 \leq z \leq 6.0$. The optimized values for f_Γ are as expected for different models. All three different reionization histories are able to reproduce the PDF for a certain value of f_Γ indicating the degeneracy in the dark gap statistics. 36

6.4 Minimum χ^2 curves plotted for different models across different redshifts. The optimized f_Γ values follow similar trend as in global minimization curve. 37

6.5 Using Dark Gap Statistics on ATON simulations (Kulkarni et al., 2019) to check the consistency of our methods against EXCITE simulations (Gaikwad et al., 2023). The blue curve represents EXCITE simulations and red curve represents the ATON simulations. The PDF behaves as expected due to low photoionization rates measured in ATON compared to EXCITE. 39

6.6 Upper Bound Constraints on the neutral fraction evolving with the redshift over different reionization histories. 41

9.1 χ^2 double minimum (Green curve) for the binned PDF due to domination of Poisson Statistics. The masking of ≤ 0.1 gives such a double minimum , but we consider that changing the threshold still keeps the optimized scaling parameter f_Γ within the uncertainties. 45

10.1 Fitting of the stacked flux profile for $\lambda_0 = 1.4$ and a $\langle \Gamma_{\rm HI} \rangle = 3.35 \times 10^{-13}\,{\rm s}^{-1}$. 47

List of Tables

6.1 Constraints on the neutral hydrogen fraction $\langle x_{\rm HI} \rangle$ and the optimised photoionisation-rate rescaling factor f_Γ for the **Early Reionization** model. The mean IGM photoionisation rate $\langle \Gamma_{12,\rm HI} \rangle$ (in $\rm s^{-1}$) is averaged over all simulation snapshots taken from Gaikwad et al. (2023) measurements. All upper bounds on $\langle x_{\rm HI} \rangle$ are quoted at 1σ. 1σ. 39

6.2 Constraints on the neutral hydrogen fraction $\langle x_{\rm HI} \rangle$ and the optimised photoionisation-rate rescaling factor f_Γ for the **Late (Best-Fit)** reionisation model. The mean IGM photoionisation rate $\langle \Gamma_{12,\rm HI} \rangle$ (in $\rm s^{-1}$) is averaged over all simulation snapshots taken from Gaikwad et al. (2023) measurements. All upper bounds on $\langle x_{\rm HI} \rangle$ are quoted at 1σ. 40

6.3 Constraints on the neutral hydrogen fraction $\langle x_{\rm HI} \rangle$ and the optimised photoionisation-rate rescaling factor f_Γ for the **Ultra-Late Reionization** model. The mean IGM photoionisation rate $\langle \Gamma_{12,\rm HI} \rangle$ (in $\rm s^{-1}$) is averaged over all simulation snapshots taken from Gaikwad et al. (2023) measurements. All upper bounds on $\langle x_{\rm HI} \rangle$ are quoted at 1σ. 40

CHAPTER 1

Introduction

There is a fundamental connection between the physical processes that occur in the Early Universe and the large-scale structure of the Universe (Peebles, 1993) . As the Universe cools down, it undergoes various phases that affect the thermal and ionization state of the Universe. This provides us with the motivation to start with the thermal history of the Universe.

The earliest stages of the Universe consisted of a very hot and dense plasma where baryons, leptons, neutrinos, etc., were all coupled together in a common thermal plasma (Kolb and Turner, 1990; Weinberg, 2008). The idea of inflation, in which the Universe undergoes exponential acceleration, is proposed to address the horizon and flatness problems (Guth, 1981). This phase occurs occurred around $t \sim 10^{-36} - 10^{-32}$ seconds. In this process, the quantum fluctuations are believed to transform into density perturbations in the Universe at later epochs (Mukhanov and Chibisov, 1981). The Universe entered a radiation-dominated era in which processes such as pair production and neutrino-electron/positron scattering are maintained by thermal equilibrium (Kolb and Turner, 1990). The freezing of weak interaction processes happened below the temperature $T \sim 10^{10}$ K. Since neutrinos have a finite lifetime of $t \sim 880$ seconds. As the temperature fell below $T \sim 10^{10} K$, their abundance is reduced compared to the photons, and we get a ratio of $n/p \sim 1$ roughly (Dolgov, 2002). At the onset of primordial nucleosynthesis, the mismatch of the neutrons and protons resulted in formation of He, giving a mass fraction of $Y \sim 0.25$ where most surviving neutrons got absorbed into this He nuclei (Alpher et al., 1948; Cyburt et al., 2016) . At time $t \sim 380000$ years, the Universe cooled down to atomic levels

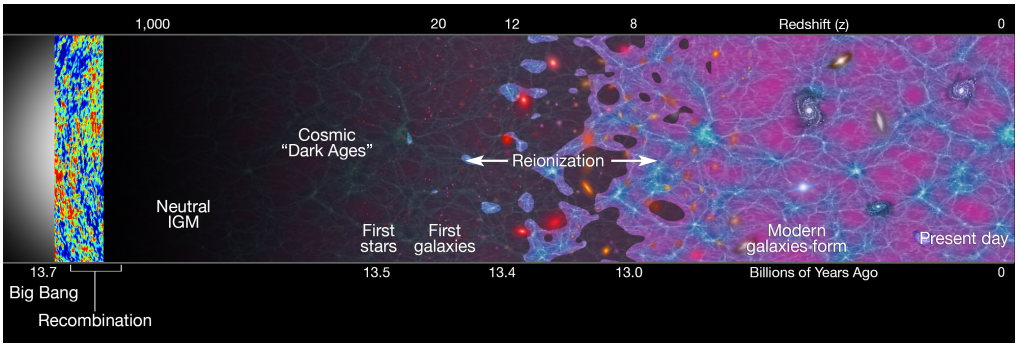


Figure 1.1: Historical Timeline of the Universe, starting from last scattering surface to present day. Credits: Robertson (2022)

where the radiation decouples from the matter since the Thomson scattering rate becomes less than the expansion, H (Zel'Dovich, 1970). This epoch is called Epoch of Recombination and the photons that decoupled from the matter gave rise to cosmic microwave background (CMB) (Seager et al., 1999).

After this epoch of recombination, the CMB photons move freely throughout the Universe without any interaction (Seager et al., 2000). This phase is called "Dark Ages" since there are no luminous sources which can generate light to interact with the matter. Through density perturbations growing under gravitational instabilities and entering in a non-linear regime where over-dense dark matter halo accrete more matter from under-dense regions till the process of halo virialization after collapse(Press and Schechter, 1974; Mo et al., 2010). Process of cooling via molecular hydrogen requires the typical halo masses for the first stars called minihalos to be in range of $10^5 - 10^6 M_\odot$ (Haiman et al., 1996; Barkana and Loeb, 2001). The gas is then able to release its energy and requirements to condense and form the stars are met. These form the first luminous objects such as Population III stars (Loeb and Furlanetto, 2013; Abel and Wandelt, 2002). This marks the beginning of "Cosmic Dawn". The transition between dark ages and cosmic dawn is the phase of the Universe where the Intergalactic Medium is properly defined: most of the baryons lying in the diffused regions. Only a percentage of the baryons fall into the potential of the matter overdensity (Shull et al., 2012). The first luminous sources begin to form and since these sources are really massive, they are short lived (Bromm and Larson, 2004). This makes them emit copius amounts of UV photons capable of ionizing their vicinities, specifically early galaxies formed of such stars become the major sources of reionization. These sources drive the reionization (Meiksin, 2009; McQuinn, 2016), each source produces HII (ionized hydrogen) regions around it i.e. proximity bubbles that can overlap with other proximity bubbles giving onset of HI reionization. The Figure 1.1 represents the historical timeline from neutral IGM to cosmic dark ages and then cosmic dawn giving the onset of the HI reionization.

HI reionization is a process where the neutral islands get ionized by the high energetic UV photons ($> 13.6\,\mathrm{eV}$) to form ionized hydrogen (Choudhury and Ferrara, 2006). The process begins with the star-forming, high-density regions consisting of the ionizing stars and ionizing the vicinity regions (low -density regions). This process is called as inside-out reionization. The reionization can also go through another route: the highly opaque / dense regions without the stars getting penetrated by the outside UV radiation where the low density regions get ionized earlier than the high density regions. This process is called outside in reionization. Most of the studies believe that inside-out reionization is more common at the onset of Epoch of Reionization (EoR) and outside-in reionization at the tails of EoR; making the reionization process an "inside-out-in" type of reionization (Chakraborty and Choudhury, 2026). Figure 1.2 represents the three phases of the reionization: The ionized bubbles percolate the IGM and then coalesce with each other forming a diffused region of the ionized regions until whole IGM is ionized.

We take the definitions of the IGM from McQuinn (2016), an unvirialized medium with density less than 200 times mean cosmic density. On sufficiently large scales, the UV photon radiation field is primarily affected by its propaga-

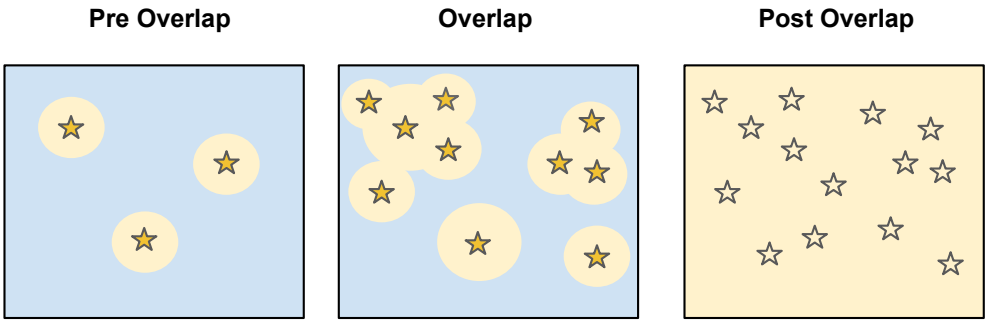
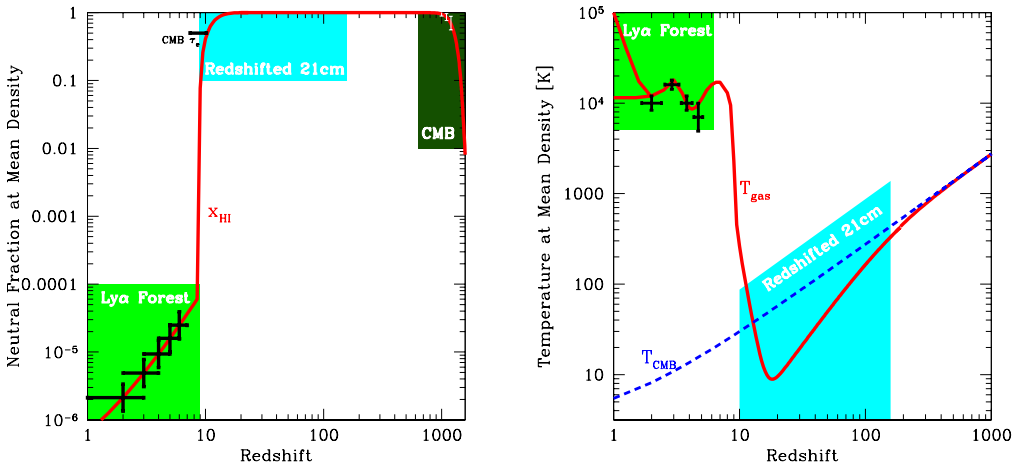


Figure 1.2: Schematic illustration taken from (Chakraborty and Choudhury, 2026) representing the three phases of reionization: Pre-Overlap, Overlap and Post -overlap phase. The ionized bubbles are believed to percolate and coalesce with other bubbles until the IGM is reionized. The final phase represents the tail end of reionization.

tion through the intergalactic medium (IGM). Hence, the structure formation is more related to the properties of the IGM rather than the sources themselves, making it a fruitful prospect in the cosmology community. There are various previous studies that use the properties of the IGM to directly give constraints on the Epoch of Reionization (Fan et al., 2006; Davies et al., 2018; Becker et al., 2021; McGreer et al., 2011). Figure 1.3 represents the evolution of the neutral fraction and temperature at the mean density. It is interesting to see how there is a rapid evolution of the neutral fraction and temperature during EoR. Different phases of the reionization are probed using different wavelengths, making the problem of EoR a multi-wavelength problem. There is a steep decline of the neutral hydrogen due to HI reionization process and it can also be seen in the thermal evolution: the CMB and the gas is coupled via Compton scattering till $z \sim 100$, and when they decouple, they are goverened with different physics i.e. $T_{gas} = (1+z)^2 T_0$ and, $T_{CMB} = (1+z)T_0$ after they decouple. The reionization sources then heat up the gas explaining the rise in the mean temperature of the gas.

Different observations have been done to constrain the EoR, making it a truly a multi-probe problem: Quasar Absorption Spectra(QSO) (Dayal and Ferrara, 2018; McQuinn, 2016), 21 cm based observations (Mesinger et al., 2011), cosmic microwave background (CMB) (Planck Collaboration et al., 2020), Lyman α emitters (LAEs) (Mason et al., 2018), Kinetic Sunyaev-Zel'dovich (kSZ) Effect (George et al., 2015), etc. Quasars act as a powerful probe to constrain the EoR, as their emitted light is absorbed by the IGM, giving imprints in the quasars as total absorption red-ward of the Lyman-α emission line. This happens due to large cross-section of the $1s - 2p$ transition resulting in total absorption of the quasar spectra even by minuscule amounts of neutral hydrogen. This effect is



(a) (b)

Figure 1.3: Schematic Illustration taken from McQuinn (2016) explaining the thermal and ionization history of the universe at a multi-wavelength scale. Rapid evolution of the temperature and neutral fraction suggests a rapid evolution of the parameters due to drastic change from cold dark ages to Epoch of Reionization.

called Gunn-Peterson Effect (Gunn and Peterson, 1965). Lyman-α forest refers to series of absorption features arising from the Lyman-α transition observed in the high-z quasar spectra.

Epoch of Reionization (EoR) helps us to connect the early stages of the universe to the present galaxies, stars and all the structures present now in the observable universe Loeb and Furlanetto (2013). The reionization process is fundamental in our understanding of the universe large scale structure formation, galaxies and their formation (Dayal and Ferrara, 2018; McQuinn, 2016; Loeb and Barkana, 2001). The imprints of the reionization process can be clearly seen in the present time and different observables can provide important constraints on the properties of the universe (Fan et al., 2006; Robertson et al., 2015; Dijkstra, 2014; Pritchard and Loeb, 2012). So, we discussed how from the epoch of recombination ($z \sim 1100$) to the Universe was predominantly neutral (Furlanetto et al., 2006), until the first generations of stars (Population III) and early galaxies began to ionize their surroundings (Bromm and Yoshida, 2011; Madau et al., 1999). The reionization process is driven by a variety of ionizing sources (Madau and Dickinson, 2014) — early galaxies (Robertson et al., 2015), faint AGN/quasars (Madau and Haardt, 2015), and high-energy sources such as X-ray binaries (Fragos et al., 2013)— and is affected by feedback, star-formation physics, and large-scale structure formation (Dayal and Ferrara, 2018; Ciardi and Ferrara, 2005)

We have to mention this one more time: the EoR connects the perturbations from the early stages of the Universe to observable structures. This phase of the Universe affects the galaxy formation, large scale structures, evolution of the IGM, clustering, voids, etc (Mo et al., 2010) . Thus, this makes the study of EoR a fruitful research towards understanding the Universe. The evolution of the universe during EoR primarily consists of two natures: Thermal and Ionization Evolution. The thermal evolution of the universe can be understood by the mean

temperature of the IGM (Hui and Gnedin, 1997; Gaikwad et al., 2019; Schaye et al., 1999), and the ionization state of this universe can be understood by the neutral fraction content of the universe (Malloy and Lidz, 2015; Bolton et al., 2011; McGreer et al., 2011).

We can also discuss about the ionization state: the ionized fraction of hydrogen depends on the number density of the ionizing photons that are produced by the young galaxies, clumping factor and the recombination rate (Madau et al., 1999). The ionization state of the IGM is a major indicator of the phases of the reionization, answering the question: how much reionization is completed? The ionization state also indicates the timing of the reionization, distribution of neutral islands, patchiness of the reionization, sometimes even putting constraints on the sources of reionization. This motivates us to look further into methods to constrain the ionization state of the IGM.

We aim to find constraints on the ionization state of the IGM during EoR. We aim to do so by using quasar absorption spectroscopy to place constraints on the neutral fraction of the hydrogen present in the Universe. We shall see how we can use something called as dark gaps and use statistics to place constraints/ upper bounds on the neutral fraction content of the IGM. This study will further help us in understanding the interplay of various parameters like photoionization rate, mean free path and reionization history affecting the EoR.

The thesis is outlined as: Chapter 2 discusses about the observational proves of the EoR, where we talk in detail what are some of the many probes used to constrain the EoR, how the studies of different probes can actually help in validating different results and the implications of each of the probe on our understanding of the EoR. Chapter 3 discusses how can we model the Lyman-α forest by modeling the Lyman-α optical depth, and use them in generating the sightlines for our study. Chapter 4 discusses the observational data-set used for this study as well as the simulations suites that are based on it (Gaikwad et al., 2023). Chapter 5 discusses the dark gaps , their identification and the statistics used in studying the gap length distribution and how they can be related to the neutral fraction constraints. Chapter 6 discusses the main results of this thesis including the χ^2 minimum curves and constraints on the neutral fraction. Chapter 7 summarizes all the methods, results used in this Thesis. Chapter 8 talks about the future implications of this study and how we can connect this study to other probes. We follow standard ΛCDM cosmology with all the measurements from Planck Collaboration et al. (2020) ($\Omega_m = 0.302$, $\Omega_\Lambda = 0.698$,$h = 0.678$) throughout the thesis and the distances are in comoving h^{-1}cMpc until otherwise mentioned so.

CHAPTER 2

Observational Probes of the EoR

Using Damping Wing Method:

The high saturated regions in the IGM or within the proxmity of a quasar and high redshift galaxies from JWST (Carniani et al., 2024) produces a damping wing signal i.e. Extended absorption profile where the Lorenztian part of the absorption dominates (refer to Chapter 3) can be used as an observable to constrain properties of the quasars or the forest. Mainly, quasar damping wings have been used to constraint bounds on neutral fraction and studying the properties of the sources (McGreer et al., 2011; Bañados et al., 2018; Bolton et al., 2011; Becker et al., 2015). The original formalism of the damping wing is given by Miralda-Escudé (1998):

$$\tau_{\mathrm{Ly}\alpha}^{\mathrm{DW}}(\Delta v) \approx \frac{\tau_{\mathrm{GP}}\, R_\alpha\, c}{\pi}\left[\frac{1}{\Delta v} - \frac{1}{\Delta v + v_{\mathrm{ext}}}\right] \tag{2.1}$$

where τ_{GP} is the Gunn Peterson optical depth and $R_\alpha = \frac{\Gamma\lambda_\alpha}{4\pi c}$ and $\Gamma_\alpha = 6.265 \times 10^8 s^{-1}$ is the Lyα decay constant, v_{ext} is the extent of the velocity gap connected to the length of neutral gap and Δv is the separation from HI/HII boundary. Damping wings signatures are tracers of regions with excessive absorption. The velocity separation and velocity extent will act as fitting parameters for the recovered damping wings. Equation 2.1 has a proportional relationship with the neutral hydrogen content of the IGM.

So, even when the core is saturated, the neutral hydrogen present distant to the line center(source) can still be used to constrain the ionization state since it still absorbs photons. This method involves of modeling of the damping wing absorption based on the uncertainties on the intrinsic quasar emission, HII region distribution during reionization, etc. This method usually uses Lyman-α forest to probe into redshifts $z > 7$ (Bañados et al., 2018; Yang et al., 2020) whereas some studies have been done in Lyman- β forest to constrain at redshifts $z < 7$ (Jin et al., 2023).

Malloy and Lidz (2015) proposed method of stacking the quasars spectra to recover the weak damping signal present in Lyman-α forest at $z = 5.5$. Stacking improves the signal-to-noise ratio of the spectra and Spina et al. (2024) recovered the weak damping signal for the first time observationally. Further studies done by Zhu et al. (2024) have also placed constraints on the neutral fraction in Lyman α forest using this method.

Using 21 cm forest:

The 21 cm forest refers to a series of absorption features in the spectra of radio sources in the IGM arising from hyperfine transition of spin flip of neutral hydrogen (Carilli et al., 2002; Furlanetto et al., 2006). This signal has not yet been detected yet but upcoming SKA era telescopes have significantly increased the chances of detection (Koopmans et al., 2015). Techniques like 21 cm line intensity mapping directly probes the neutral hydrogen in the IGM (Pritchard and Loeb, 2012; Bull et al., 2015). Various studies have been done to get a forecast in the detection of 21 cm. We can also trace the spin temperature and ionization state of the IGM using 21 cm transition (Field, 1959; Hirata, 2006). Cross-correlation forecast studies of line intensity maps of different atomic and emission lines from galaxies with 21 cm line can help in not only tracing the neutral hydrogen but also the complex astrophysical processes during galaxy formation (Lidz et al., 2009; Silva et al., 2013)

There is a brightness contrast in the temperature due to interaction of 21 cm with the CMB photons (Madau et al., 1997) During dark ages, the temperature of the gas, 21 cm and CMB photons were all tightly coupled, but at the onset of cosmic dawn, these temperatures decoupled since $T_{\rm gas} \propto (1+z)^2$ and the 21 cm transition depends on the population of the hyperfine levels. So all these three temperature decouple and interactions as CMB coupling , collisional coupling at high densities, and even Lyman α coupling with 21 cm started to compete with each other. So, we get a brightness temperature contrast as :

$$\delta T_b(\mathbf{x}, z) \simeq 27\, x_{\rm HI}(\mathbf{x}, z)\, [1+\delta_b(\mathbf{x}, z)] \left(\frac{1+z}{10}\frac{0.15}{\Omega_m h^2}\right)^{1/2} \left(\frac{\Omega_b h^2}{0.023}\right) \left(1 - \frac{T_\gamma(z)}{T_S(\mathbf{x}, z)}\right) \tag{2.2}$$

where 27 is in mK represents the order of the signal, $x_{\rm HI}$ represents the neutral hydrogen fraction, $\delta_b(\mathrm{x}, \mathrm{z})$ represents the over-density of the baryons , the term containing the $(1 + z)$ factor represents the cosmological scaling factor, coming from the Hubble expansion rate. Factor consisting of Ω_b represents the baryon density normalization, T_γ and T_S represents the CMB and spin temperature respectively.

In Equation 2.2, the most important term is $(1 - \frac{T_\gamma(\mathrm{z})}{T_S(\mathrm{x,z})})$ which determines whether we get emission or absorption. We get positive brightness contrast if $T_S > T_\gamma$ resulting in emission, whereas we get negative contrast if $T_S < T_\gamma$ resulting in absorption. If both the temperatures are same, we get no signal. As we can see that $\delta T_b \propto \bar{x}_{\rm HI}$, gives us constrain on the mean neutral fraction. In order to find the spatial fluctuations in the neutral fraction, we use statistics like power spectrum to see the contributions from density, temperature and neutral fraction fluctuations.

Using kSz (Kinetic Sunyaev Zeldovich Effect):

The kSZ effect arises when the electrons undergoing bulk motion undergo Compton scattering with the CMB photons (Sunyaev and Zeldovich, 1980). The detection of kSZ signal can give us crucial understanding of the reionization physics

(Mueller et al., 2015; McQuinn et al., 2005) like duration of the EoR (Raghunathan et al., 2024) , simulating maps around reionization (Alvarez, 2016),etc. There are two basic forms of kSZ effect that we can observe: Homogeneous /post EoR kSZ effect due to bulk flow of the halos with free electrons at redshifts $z \leq 3$ and inhomogeneous/ reionization kSZ effect $z \geq 6$ due to the motion of ionized bubbles that contain free electrons during EoR. This introduces temperature anisotropies on small angular scales. The temperature fluctuation equation in kSZ effect can be given as :

$$\frac{\Delta T_{kSZ}(\hat{n})}{T_{CMB}} = -\sigma_T \int d\eta e^{-\tau(\eta)} n_e(\eta, \hat{n}) \frac{\mathrm{v_e}.\hat{n}}{c} \tag{2.3}$$

Here, σ_T represents the Thomson cross-section $= 6.65 \times 10^{-29}\,\mathrm{m}^2$ and η represents the conformal time for which the line-of-sight integral is taken. The integral makes kSZ an integrated effect meaning we do not get any direct redshift information. The visibility factor is given by $e^{-\tau(\eta)}$ factor giving the probability that the photons has not yet scattered before η time. In Equation 2.3, the quantity to consider the ionization state of the IGM is i.e $x_e = 1 - x_{\mathrm{HI}}$. $\mathrm{v_e}.\hat{n}$ represents the peculiar velocity of the electrons along the line of sight i.e. contributing Doppler effect. During EoR, the ionization fraction is highly inhomogeneous, hence we get a patchy kSZ at this epoch. The angular power spectrum of kSZ depends strongly on the bubble size distribution and duration of EoR (Lidz et al., 2007; Ferraro et al., 2016).

Using Cosmic Microwave Background Polarization

The cosmic microwave background (CMB) provides a powerful, global constraint on the Epoch of Reionization through its polarization anisotropies. During reionization, the emergence of free electrons leads to additional Thomson scattering of CMB photons, which enhances the large-scale E-mode polarization signal.

This optical depth of this Thomson Scattering is given by:

$$\tau_e = \int_0^{z_{\mathrm{re}}} c\, n_e(z)\, \sigma_T \frac{dt}{dz}\, dz, \tag{2.4}$$

where n_e, σ_T, c represents the free electron number density, the Thomson scattering cross-section, and the speed of light respectively. The optical depth represents the integrated probability that a CMB photon is scattered by free electrons along its path.

As an integrated quantity, τ_e provides a constraint on the total column density of free electrons and hence on the overall timing of reionization. A higher optical depth implies means that the reionization process has started much earlier than the case of low optical depth. But since, τ_{CMB} is due to integrated effect, it does not provide detailed information about the ionized fraction evolution.

Measurements from the updated Planck (Planck Collaboration et al., 2020) give $\tau_e \approx 0.054 \pm 0.007$, favoring a relatively late and extended reionization scenario, which agrees with the quasars measurements with the bulk of the process occurring at $z \sim 6$–8. Overall, CMB polarization measurements provide a

robust normalization constraint on reionization models, that is also used in forward modeling of Ly-α forest. Although the CMB provides global constraints that are used in other methods also, there is no detailed spatial information about the IGM. This can be probed using Ly-α emitters.

Quasar Absorption Spectroscopy

A more direct probe of the IGM is provided by the absorption features observed in the spectra of high-redshift quasars. This probe is only dependent on the intervening neutral hydrogen present in the IGM along the line of sight, providing us with direct measurements on IGM opacity during the Epoch of Reionization.

Quasar absorption spectroscopy solely rely on techniques involving flux received in absorption spectra of quasars at very high redshifts. These quasars have usually high signal-to-noise (SNR) ratio and produce highly energetic photons that can ionize the IGM in their vicinity. The neutral regions in the IGM lying at different redshift will absorb the light redward of the Lyman-α emission line ($\lambda^{\alpha}_{\text{rest}} = 1215.67$ Å), giving a series of discrete absorption features in the quasar spectrum, whereas the light with $\lambda > \lambda^{\alpha}$ will be shifted further away and will not be absorbed by the neutral hydrogen present in the IGM. As a result, we get a sharp flux decrement redwards of the Lyman α emission line. This is called Gunn-Peterson trough (Gunn and Peterson, 1965). The series of absorption lines we get due to Lyman α transition form what is known as Lyman α forest.

The Gunn-Peterson Effect (Gunn and Peterson, 1965) can be used in quasar absorption spectroscopy to constrain neutral hydrogen. The location of the quasar emission line (z_{em}) can be found out as : $\lambda = \lambda_{\alpha}(1+z_{em})$ where $\lambda_{\alpha} = 1215.67$ Å represents the Lyman-α rest wavelength.

This absorption is not uniform throughout the universe as shown by Miralda-Escudé et al. (2000) and later confirmed by observations Becker et al. (2015), strongly pointing to a patchy, inhomogeneous reionization of the universe. This means that along different line of sight, the spectra will look statistically different. The Lyman α forest is modeled by the theoretical formalism of the optical depth in this wavelength regime. The flux is then modeled as $F = e^{-\tau}$. The Gunn-Peterson optical depth can be formulated as (Chakraborty and Choudhury, 2026):

$$\tau_{\alpha}(\nu_{\text{obs}}, z) = \frac{\sigma_{\alpha}\, c}{H(z)}\, x_{\text{HI}}\, n_H (1+z)^3 \tag{2.5}$$

$$\approx 2.3 \times 10^5\, x_{\text{HI}} \left(\frac{n_H}{\bar{n}_H}\right) \left(\frac{\Omega_b h^2}{0.022}\right) \left(\frac{\Omega_m h^2}{0.142}\right)^{-1/2} \left(\frac{1-Y}{0.76}\right) \left(\frac{1+z}{5}\right)^{3/2} \tag{2.6}$$

where $\tau_{\alpha}(\nu_{obs}, z)$ represents the optical depth of the Gunn Peterson trough, and σ_{α} represents the Lyman Alpha absorption cross-section, x_{HI} represents the neutral hydrogen fraction, and $H(z)$ is the Hubble parameter. This formula is further approximated in second equation 2.6 , when we can approximate $H(z)$ in a matter dominated universe.

The thermal evolution of the IGM can be analyzed through power spectrum of the Lyman α absorption fluctuations probing small structures of the Universe (Rorai et al., 2017; Walther et al., 2019). The metal lines absorption present in

the quasar spectra can also be studied to put constraints on EoR (Becker et al., 2011). The UV photons from the galaxies enrich the IGM by metals (Madau et al., 1999) . Absorption profiles of such ions can reveal a lot of information about EoR. Different ions will trace different regions of the IGM. The relative ratios of the abundance of the metal lines also give constrain on the ionization state of the IGM (Peeples et al., 2014). For instance, OI is tightly coupled with HI, hence it traces neutral regions. Whereas CIV requires highly ionized gas and hence it traces ionized regions (Songaila, 1998).We can do similar evolution for the metal absorbers as the neutral hydrogen and do metal absorber statistics, to see the patchiness of the reionization and the bubble size distribution (Rahmati et al., 2016).

Using High-z Galaxies:

Lyman α emitters (LAEs) are star-forming galaxies with a very strong Lyman α emission discovered at redshifts up to $z \sim 7 - 10$ (Oesch et al., 2016; Zitrin et al., 2015; Roberts-Borsani et al., 2016). This provides a complementary probe having different statistics compared to 21 cm (Dijkstra, 2014). However, the absorption of Lyman-α photons by the IGM makes the visibility of the LAEs very low during EoR (Santos, 2004).

Clustering of LAEs is an indicator of the patchiness of reionization; if reionization has not yet completed, we will see enhanced clustering at intermediate scales (Ouchi et al., 2010; Konno et al., 2018). The abundance of LAEs also drops due to the high neutral fraction of the IGM, and hence various studies use a 1-point function, such as the luminosity function, to put constraints on the neutral fraction. The power spectrum of LAEs is used to understand the reionization topology, and effects like reionization-induced bias reinforce the probe's efficiency in constraining the Epoch of Reionization.

Higher order statistics like bispectrum are used to trace the morphology during the Epoch of Reionization, and can be used as a potential tracer to capture non-Gaussian features and non-linear effects that the power spectrum misses out (Majumdar et al., 2018). Although observationally, the bispectrum is not employed due to the lesser data volume of the surveys, studies are now being done based on synthetic mock data generation of the LAEs (Maitra et al., 2025) . Despite this, constraints from Lyman-α emitters are subject to various factors like uncertainties arising from the galaxy's intrinsic properties, CGM attenuation, and radiative transfer effects. As a result, most of the interpretations of the LAEs are model-dependent and have halo biases (Mason et al., 2018).

Now having discussed some of the probes of EoR, we will examine in detail how the Lyman Alpha Forest can be used as a probe of the IGM.

CHAPTER 3

Modeling Quasar Absorption Spectra

Absorption Profile

The absorption caused by the neutral hydrogen lying between the quasars and the observer along a line of sight is contributed to two factors: Uncertainty between the atom transition levels (here $1s - 2p$), that will be caused due to the finite lifetime of the excited state and Doppler broadening due to thermal motion of the particles in the absorbing regions. These two contributing factors are true not only for neutral hydrogen, but also metal absorbing systems. There may be peculiar velocity also associated with the cloud/ scatterer , this does not change the shape of the profile but shifts the line center frequency by amount $\nu(1-\frac{v_{LOS}}{c})$ where $v_{LOS} = v_{hubble} + v_{pec}$. We get a convolution of Lorentzian profile (due to the uncertainty) and a Gaussian profile (due to thermal broadening). The resultant convolution profile is called a Voigt profile (Voigt, 1912). This profile is given as:

$$H(a,x) = \frac{a}{\pi}\int_{-\infty}^{+\infty}\frac{e^{-y^2}}{(x-y)^2+a^2}dy \tag{3.1}$$

Here $x = (\lambda - \lambda_\alpha)/\Delta\lambda_D$ and $y = v/b$. Here $\Delta\lambda_D = \frac{b}{c}\lambda_\alpha$, where the E_i represents the energy of the transition , and λ_α represents Lyman α wavelength, b is the thermal parameter and v is the peculiar velocity.

This profile will decide how the spectrum will be affected by the temperature, recessional velocity, peculiar velocity (if present), neutral density, etc. The Lyman-α optical depth consists of this Voigt function and the general optical depth along a line of sight for a single absorber can be formulated as:

$$\tau(z_0) = \frac{cI_\alpha}{\sqrt{\pi}}\int dx \frac{n_{HI}[x,z(x)]}{b[x,z(x)][1+z(x)]} \times V[a, \frac{c[z(x)-z_0]}{b[x,z(x)](1+z_0)} + \frac{v[x,z(x)]}{b[x,z(x)]}] \tag{3.2}$$

Here $I_\alpha = 4.45 \times 10^{-18} cm^2$ is the effective cross-section (in terms of fundamental constants: $I_\alpha = \frac{\pi e^2}{m_e c} f_\alpha$, where f_α denotes the oscillator strength). For Lyman α, cross-section is: $\sigma_\alpha(\nu) = I_\alpha f_\alpha \phi(\nu)$, $\phi(\nu)$ represents the line profile of the transition. $n_{HI}[x, z(x)]$ represents the neutral hydrogen density and V(A,B) is the Voigt function, where the first term $a = \frac{\lambda_\alpha^2 \Gamma_i}{4\pi c \Delta\lambda_D}$ represents the damping factor (where Γ_i gives the natural decay rate) and the second term gives the contribution from the Doppler shift and second term v gives the peculiar velocity contribution . $z(x)$ represents the redshift grid from where we take the contributions to calculate the optical depth at z_0 redshift, and dx representing the

comoving distances over which we take the integrals. The thermal parameter is given by $b = \sqrt{\frac{2k_BT}{m}}$ and it is measured using FWHM of the absorption profile (Choudhury et al., 2001). Figure 3.1 represents the variation of the absorption profile with respect to the neutral hydrogen density. From Figure 3.1, even if we have density and temperature value at a single pixel, all the vicinity will be affected by it due to Voigt nature of the absorption profile. It is also important to note that the line of sight optical depth is cumulative in nature. The damping wings can extend upto 500-1000 km/s depending on the strength of the absorption line.

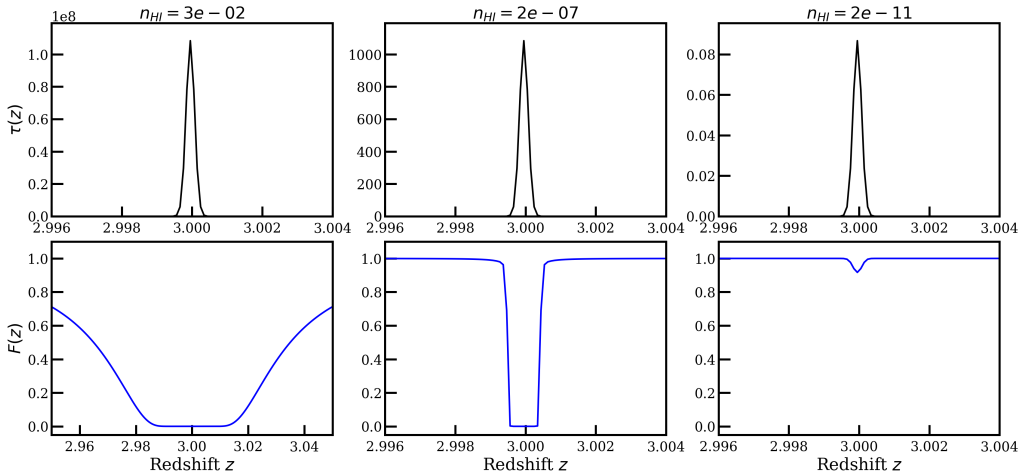


Figure 3.1: Shape of absorption profile for a single pixel absorber. Right to Left: As the neutral hydrogen density increases, the flux absorption profile becomes more broader and we get a clear strong damping wing for $n_H = 2.5 \times 10^{-2}$. The Voigt profile is dominated by Gaussian profile for low nH values, whereas in high nH values,damping wings are visible i.e., Lorentzian profile dominates.

Equivalent width is an useful construct when we do not have sufficient spectral resolution and it is used to measure strength of an absorption line. We define it as:

$$W = \frac{1}{f_0} \int (f_0 - f(\lambda)) d\lambda$$

In Figure 3.2, there are three different regimes of N_{HI}: for $N_{HI} \leq 10^{13-14}$, W increases linearly with the column density, so it is also called unsaturated absorption (flat portion of the curve of growth) . $10^{14} \leq N_{HI} \leq 10^{17}$ is true for low density IGM in Lyman α forest. For this regime , the strength of the absorption slows down and there is not much change of W with respect to the column density, thus known as saturated regions. For $N_{HI} \geq 10^{17}$, $W \propto \sqrt{N_{HI}}$ and this is the damping wing regime where the physics is dominated by the Lorentzian factor in the Voigt Function. The Lorentzian broadening arises from the finite lifetime of the energy levels from the Heisenberg uncertainty principle, producing extended damping wings. Strength of absorption increases with the Doppler parameter only in saturated regions, otherwise W remains independent of the thermal parameter.

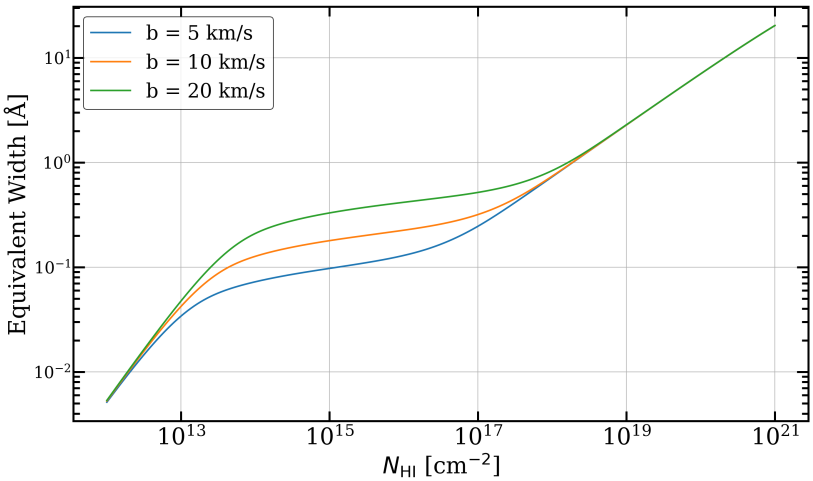


Figure 3.2: Growth Curve: Equivalent Width plotted as a function of column density N_H with changing thermal parameter. The strength of absorptioin line depends on oscillator stength,column density and thermal parameter,b. There are different regimes of the column density where the strength of absorption behaves differently. It is affected by the thermal parameter only in saturated regions.

3.1 Ionization State:

The absorption in spectra of quasars is produced due to resonance scattering or photoelectric absorption, and these two processes are dependent on the ionization state of the IGM. Young galaxies depending on the masses and luminosity functions will contribute differently to the reionization and photons having energies greater than the ionization potential of bounded electrons will be able to ionize the respective elements.We use quasars as background sources. If we are specific to hydrogen, photons having more than 13.6 eV energy will be able to ionize neutral hydrogen. Photoionization Rate is the rate of the number of photons available that can ionize the gas (a photon having frequency greater than threshold frequency (Lyman Limit)).

$$\Gamma_{HI} = \int_{\nu_T}^{\infty} d\nu \frac{4\pi J_\nu}{h_p \nu} \alpha_\nu^{HI}$$

Here, $\alpha_{HI}(\nu)$ denotes the photoionization cross-section and ν_T is the ionization frequency, $\frac{4\pi J_\nu}{c}$ represents the specific energy density, and J_ν is the specific intensity averaged over solid angle.
There can be two cases of photoionization: Case A type recombination, where it is assumed that photons can escape freely without recombination and every recombination is treated as true recombination and discrete recombination. Since most of the gas is optically thin, the photons emitted during this process can escape without getting absorbed again. Here, the probability of photons escaping is much more. Whereas in optically thick regions, we have Case B type recombination, where the emitted photons are reabsorbed immediately due to large density in the vicinity of absorption center. Case A type recombination, $\alpha_A(T)$ represents total rate coefficient for the radiative capture i.e $\alpha_A(T) = \Sigma_{n=1}^{\infty} \alpha_n(T)$, this includes all levels starting from n=1. Whereas in Case B recombination, we consider that recombination directly to the ground state is excluded

since the LyC photons will reionize the nearby atoms in lesser time. So we get $\alpha_B(T) = \sum_{n=2}^{\infty} \alpha_n(T)$, under these conditions, the photons are locally absorbed again. We consider case A since we are interested in diffused IGM where the medium is optically thin and the recombination photons are not able to reionize nearby atoms immediately.

We express the ionization rate equation in terms of neutral fraction and this formulation has been taken from (Meiksin, 2009):

$$\frac{dx_{\rm HI}}{dt} = -x_{\rm HI}\Gamma_{\rm HI} + x_{\rm HII} n_e \alpha_A(T), \tag{3.3}$$

$$\frac{dx_{\rm HII}}{dt} = -\frac{dx_{\rm HI}}{dt}. \tag{3.4}$$

We consider the IGM to be in a photoionization equilibrium i.e. $\frac{dx_{\rm HI}}{dt} = 0$, giving us:

$$\Gamma_{HI} x_{HI} = \alpha x_{HII} n_e$$

where $x_{HII} = 1 - x_{HI}$. The photoionization equilibrium equation can also be written in terms of the electron correction term $n_e = 1 + \frac{Y}{4(1-Y)}$ where Y represents the He mass fraction~ 0.24 that also accounts for the HeII ionization and $x_{\rm HI}$ alone:

$$x_{HI} = \frac{\alpha_A(T) n_e n_H}{\Gamma_{\rm HI}} \tag{3.5}$$

Here $\Gamma_{\rm HI}$ represents the photoionization rate and $\alpha_A(T)$ (cm^{3}s^{-1}) is the Case-A recombination rate which has a weak dependence on the temperature as: $\alpha_A(T) = 4.36 \times 10^{-10} T^{-0.7573}$ for $T \geq 5000$ K. We are considering simple case of reionization, where no collisional processes and clumping factors have been accounted. Figure 3.3 represents the evolution of $x_{\rm HI}$ throughout the redshift and giving us a picture of how reionization looked like. The ioninzation fraction in this case can be given as :

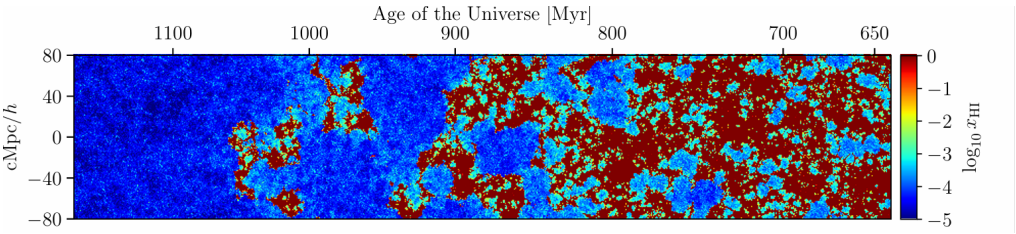


Figure 3.3: Neutral fraction evolution with lookback time, illustrating the reionization process; taken from Kulkarni et al. (2019)

3.2 Generating Lyman α Forest Code:

Equation 3.2 talks about a single absorber affecting its vicinity regions due to the Voigt nature of the absorption. A Real Lyman-α forest arises from continuous field or distribution of all such pixels, giving rise to a cumulative sum of the optical depth of all the absorbers and how they affect the whole IGM. We use following hydrogen density evolution with the redshift to find the neutral hydrogen fraction via Equation 3.5 for optical depth generation:

$$n_H(z) = \frac{(1-Y)\Omega_b}{m_p}\frac{3H_0^2}{8\pi G}(1+z)^3 \tag{3.6}$$

where $(1-Y)$ represents mass hydrogen fraction, $\Omega_c = \frac{3H_0^2}{8\pi G}$ represents the mean cosmic density at $z = 0$ and $H_0 = 100\,h\,\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$ represents the Hubble constant at present time, and Ω_b representing the baryon density. Equation 3.6 is used to calculate n_{HI} term in Equation 3.2. Fields like baryon density Ω_b, temperature T and peculiar velocities v are taken from from Sherwood line-of-sight simulations (Bolton et al., 2017) to generate the optical depth. We use periodic boundary conditions for flux generation and use Tepper-García (2006) relation for computational efficiency. The Voigt function can be approximated using Tepper-García (2006) as:

$$H(a,x) \approx H_0(x) - \frac{a}{\sqrt{\pi}x^2}\left[H_0(x)^2(4x^4+7x^2+4+Q) - Q - 1\right], \tag{3.7}$$

where

$$H_0(x) = e^{-x^2}, \quad Q = \frac{3}{2x^2}. \tag{3.8}$$

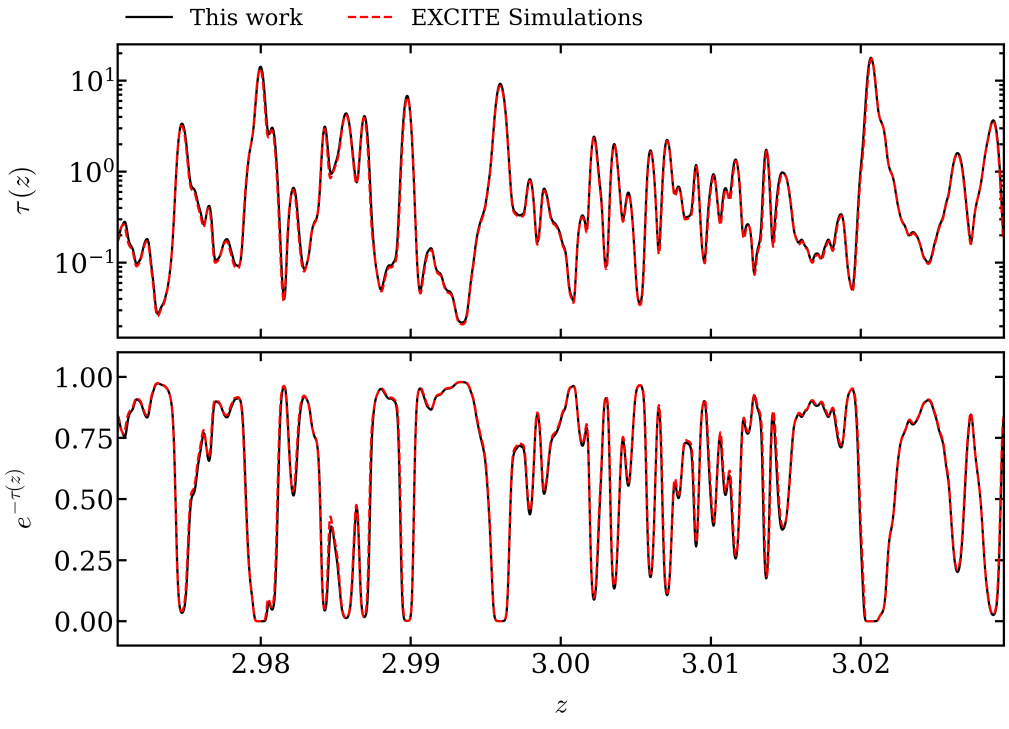


Figure 3.4: Synthetic Ly α forest generated using LOS optical depth 3.2. This is compared with ex-cite simulations (Gaikwad et al., 2023) and the match makes sure that code is working properly.

Figure 3.4 illustrates the synthetic Lyman α forest using above methods. All the grids of the fields contain 512 pixels and we take the redshift window centered around $z \sim 3$. We see that our calculated Lyman α generated code matches

with EX-CITE simulations (Gaikwad et al., 2023) (refer to Chapter 4). This validates that our code is working fine and matching pixel-to-pixel with the synthetic Lyman α forest. Although the generation of the flux sightlines have a proper structured method, modelling the Lyman α forest comes with its own challenges. The large opacity fluctuations that are observed have to be captured by the simulations for studying the observables further. These fluctuations cannot be explained by density fluctuations alone, so various studies try to forward model the Lyman α forest and capture these fluctuations. Various reasons are given to explain these large opacity fluctuations including quasar clustering on large scales (Meiksin, 2020), patchy reionization, (Kulkarni et al., 2019), evolving short mean free path (Davies and Furlanetto, 2016; Gaikwad et al., 2023) etc (refer to the given papers for more details). We consider variation in the mean free path or evolving mean free path with the redshift to capture the opacity fluctuations (Gaikwad et al., 2023). Having discussed about the Lyman α forest generation, we can now see how the ionization state of the IGM can be inferred.

Using Dark Pixel Fraction: The dark pixel fraction are defined as fraction of pixels having no detectable transmission. These dark pixels are treated neutral below the detection level. This method is nearly model independent since it takes into the volume filling fraction of the IGM and compares the relative fraction of the dark pixels to put constraints on the neutral fraction. We call this nearly model independent because it does not account for the mean flux, just the fraction of the dark pixels. More fraction mean more absorption and we expect higher neutral fraction. It is simple and robust method since very less modeling is required and it also works in high saturation regions. Although dependent on the noise sensitivity, previous studies have used this method (McGreer et al., 2011; Davies et al., 2026) and placed upper bound constraints on the neutral fraction. The upper bounds are placed since even in a fully ionized universe there can be residual neutral hydrogen, density fluctuations which can produce optical depth $>> 1$ rendering the flux zero and hence giving maximum neutral fraction that must be present at that redshift to give same dark pixel fraction.

Using Dark Gap Statistics: All the above methods do provide the constraints on the neutral fraction, but they lose out the spatial information about the content of the neutral hydrogen in the IGM. Dark Gap statistics measure average neutral fraction and is more constraining since it provides more information about mean photoionization rate, reionization histories. One can also explore a large parameter space and different models using dark gaps found in the synthetic spectra using simulations. Different physics govern different statistics depicting a strong degeneracy and interplay between neutral fraction, mean free path and photoionization rate. We shall discuss more about Dark Gap Statistics in Chapter 5.

CHAPTER 4

Observations & Simulations

4.1 Observation Data-Set

The observational dataset for this analysis is taken from Ly-α observations from Zhu et al. (2021) based on 38 quasars taken from the E-XQR 30 sample (refer D'Odorico et al. (2023)). All the quasars are located at $z > 5.5$ with a high signal-to-noise ratio (SNR $> 10 - 20$ per 10 kms^{-1}) and the spectra is obtained from the X-Shooter instrument at Very Large Telescope at Keck (Vernet et al., 2011). The X-Shooter instrument is an intermediate-resolution echelle spectrograph that operates at a resolution of $\sim 7000 - 14000$, with three arms covering the ultraviolet, visible, and near-infrared. The quasars span $5.8 \leq z_{em} \leq 6.6$, and these quasars are selected from magnitude criteria ($J_{AB} \leq 20.0$ for $6.0 \leq z \leq 6.6$).

Although the Ly-α forest observations from (Zhu et al., 2021) use ESI quasars with XQR-30 spectra, we only consider the XQR-30 spectra for comparison due to high SNR. The observations in (Zhu et al., 2021) use optimized sky subtraction (Kelson, 2003), and the telluric corrections are done by Cerro Paranal Advanced Sky Model (Noll et al., 2012) (Jones et al., 2013). They have done a power law fitting of the zero point flux over the Ly-α forest to reduce foreground galaxy contamination.

The continuum reconstruction in the observations has been performed using log-PCA (Principal Component Analysis) fitting, as described in (Davies et al., 2018). There are 15 and 10 components on the red and blue sides, respectively, and the mapping from the red principal components to the blue components is performed via a projection matrix. Quasars with strong broad absorption lines have been left out, and the mild absorption lines are masked while fitting the spectra. They have also left out Ly-α emission peak and proximity zones, and the Ly-α continuum fitting uncertainty is less than 10 %.

As we have mentioned previously, we are using the Lyman-α forest observations from Zhu et al. (2021). For the observations, the spectra is rebinned to 1 h^{-1}Mpc and flux is normalized with respect to the continuum flux. If this normalized flux falls below 0.05 for a continuous region (this criteria is chosen similar as Zhu et al. (2021) as the non-detection of 2 σ of the lowest available SNR quasar sightlines), and the minimum length for this region is 1 h^{-1}Mpc, then it is categorized as a dark gap. For avoiding the quasar proximity region, they have identified the dark gaps 7 pMpc (approximately size of the biggest

proximity region found (Eilers et al., 2018)) from the quasar emission line. For lower bounds of searching, they have used the rest wavelength of Ly-β($\sim$ 1041 Å). Masking $\pm 75\,\mathrm{kms}^{-1}$ of the spectra interval is done for pixels having flux threshold above $3\sigma_{\mathrm{f}}$ to mask the sky-limited residuals (SLR) . No correction is applied for contamination from damped Lyman α systems and metal lines are identified in the observation spectra. We will follow similar procedures to identify the dark gaps in our simulation mock spectra for comparison.

4.2 Simulations & Model Suites

The code for the EXtended reionization using Code for Ionization and Temperature Evolution (Ex-CITE) simulation has been used to generate fluctuations in the photoionization rate and the mean free path (Gaikwad et al., 2023). EX-CITE runs on the hydrodynamical Sherwood Simulation Suite (Bolton et al., 2017) and is performed with the P-GAGDET-3 code (Springel, 2005). The Sherwood Simulations are then RT (radiative transfer) post-processed using the coupled equations for the photoionization rate and mean free path, which have been solved using the Octree method.

We are using the simulations with a box size of $80\,h^{-1}$Mpc, and 2×2048^3 particles are employed in the box, denoted as L80N2048. We have used outputs from a box size of $80\,h^{-1}$Mpc, as this is sufficient to capture the dark gap statistics. The resolution of the simulations is very high: $0.31\,\mathrm{h}^{-1}$cMpc (512^3 grids), which is more than twice of previous simulations (Davies et al., 2017; D'Aloisio et al., 2018). The values used for the ionizing emissivity parameter, mass halo, and mean free path power index are taken from the assumptions of Miralda-Escudé et al. (2000); Davies and Furlanetto (2016); Barkana and Loeb (2001). For consistency, the simulations use different skewers: L40N512, L40N2048, L80N1024, L80N2048, L160N512, and L160N1024. The Sherwood simulations are converged, with all skewers accounting for the box size, halo function, resolution, etc.

The simulation outputs are stored at 40 Myr starting from the initial redshift of 40. Since we are interested in the redshift range of $(5.0, 6.0)$, this corresponds to sampling of the outputs at redshift $z = 5.11, 5.26, 5.41, 5.58, 5.95$. This sampling is chosen because the simulations also compare with ATON simulations (Kulkarni et al., 2019) (ATON simulations using the same intervals).

The primary ionizing sources for the HI reionization come from the escaped fraction of ionizing photons from the young galaxies that ionize the IGM inhomogeneously. The simulation reproduces these patchy simulations by introducing fluctuations in the background by varying parameters like mean free path and photoionization rates. Thus, it is important that we introduce the theoretical framework related to these parameters. The MHR model (Miralda-Escudé et al., 2000) is used for sub-grid clumping factor approximations. The simulations also assume this conservative approach is more aligned with the Davies 2016 (Davies and Furlanetto, 2016) $\lambda_{\mathrm{mfp}} - \Gamma_{12}$ parameter variant models.

The emissivity of dark matter halos in the grid is determined by the halo mass function, without modeling details of galaxy formation and feedback. The HI ionizing emissivity weights are modeled by the dark matter halos as $\epsilon_{\mathrm{halo,i}} =$

$\epsilon_0 \times M^{\beta}_{\text{halo,i}}$ if $M_{\text{halo}} > 10^9$ else 0. $\beta = 1.0$ representing emissivity power law index and ϵ_0 is the normalization factor. This approximation denotes the weight to the halo mass and assumes a linear relationship between the mass and HI reionization. This simple model is shown to capture the fluctuations in the background (Gaikwad et al., 2023). The value of ϵ_0 affects the absolute value of the mean photoionization rate and is fixed through the total photon budget. The source variations have marginal effects (the effects of varying M_{cutoff} and β on measured quantities are small).

We also note the emissivity weights by normalizing the emissivity with respect to the total halo mass: $\omega_{\text{halo,i}} = \frac{M^{\beta}_{\text{halo,i}}}{\sum_{i=1}^{N_{halo}} M^{\beta}_{\text{halo,i}}}$ if halo mass satisfies the cutoff mass. The simulations vary the mean photoionization rate with ϵ_0 as a free parameter, which helps in exploring a large parameter space of mean free path and photoionization rate ($\lambda_{\text{mfp,HI}} - \langle\Gamma_{\text{HI}}\rangle$), independent of the redshifts, making it easier to compare with the observations.

The Lyman-α opacity fluctuations are captured by the fluctuations in photoionization rate ($\delta\Gamma_{HI}$) and mean free path variation. This variation in both parameters is dependent on each other, i.e., they form a coupled system of equations that have to be solved simultaneously. This set of equations are:

$$\frac{\Gamma_{HI}(x)}{\langle\Gamma_{HI}\rangle} = f_0 \sum_{i=1, j\neq i}^{N_{sources}} \frac{w_{halo}}{4\pi R_{ij}^2} e^{-\int_{R_i}^{R_j} dx/\lambda(x)} \tag{4.1}$$

where we sum over all the i indices up to the number of sources(N_{source}), f_0 acts as a normalization factor (ϵ_0 is absorbed in f_0) in the photoionization fluctuation maps, and R^2 variation accounts for the flux inverse square law. The attenuation by the absorbers in the IGM is accounted for by the mean free path parameter in the exponential factor.

The power-law variation of the mean free path with photoionization fluctuations is taken to be $p = 2/3$ for the reference (Davies and Furlanetto, 2016; D'Aloisio et al., 2018). The mean free path variation is given by:

$$\lambda(x) = \lambda_0 \Delta^{-1} (\delta\Gamma_{HI})^p \left(\frac{E_{bin}}{E_{ion}}\right)^{0.9} \tag{4.2}$$

Here, λ_0 acts as a spatially averaged mean free path parameter, which will serve as a parameter to generate the fluctuations. E_{bin} and E_{ion} denote the photon average energy in the frequency bin and the HI ionization potential, respectively. Now we can discuss briefly about how the numerical implementation is done in EXCITE:

We have initial set of parameters: λ_0 , 3-D density field ($\Delta(x)$), halo catalog, and maximum refinement level i.e. $R_{max} = 9$ for 512^3. Initially, R=6 is the initial coarse refinement level, and we assume a uniform mean free path in the simulation cell, making $\lambda_0 = \lambda_{mfp}$ and $\delta\Gamma_{HI} = 1$ for every cell. A continuous density grid is sampled by using SPH particle locations, smoothed with an SPH kernel on a grid (2^{3R}). We distribute emissivity weights onto this 2^{3R} grid using nearest point approximation and then recursively divide the simulation box into sub-cubes for octree construction. The total emissivity weight, sub-cube center, and size are saved at each tree level, and this center is used to calculate R_{ij}, i.e.,

distance between source and the sink. We calculate the mean free path variation 4.2 at each spatial location with the photoinization fluctuations at each sink. Bresenham rasterization (Bresenham, 1965) is used to compute the contribution of each source to find the shortest path between two points. While traversing from top to bottom (from sources to sink), this R_{ij} distance is calculated, and the criterion of the ratio of size and Rij is taken. If the ratio is less than the threshold of 0.7 ($s/R_{ij} < \theta$), then sources in that sub-cube are treated as one source, and the emissivity of all such sources is added to calculate the IGM attenuation along the sightlines. However, if this ratio exceeds 0.7, we traverse the tree downward until this criterion is satisfied. When this criterion is never satisfied i.e. when we are at the lowest resolution, we set the IGM attenuation to zero. In this way, we can capture fluctuations recursively efficiently (refer Gaikwad et al. (2023), for more details)

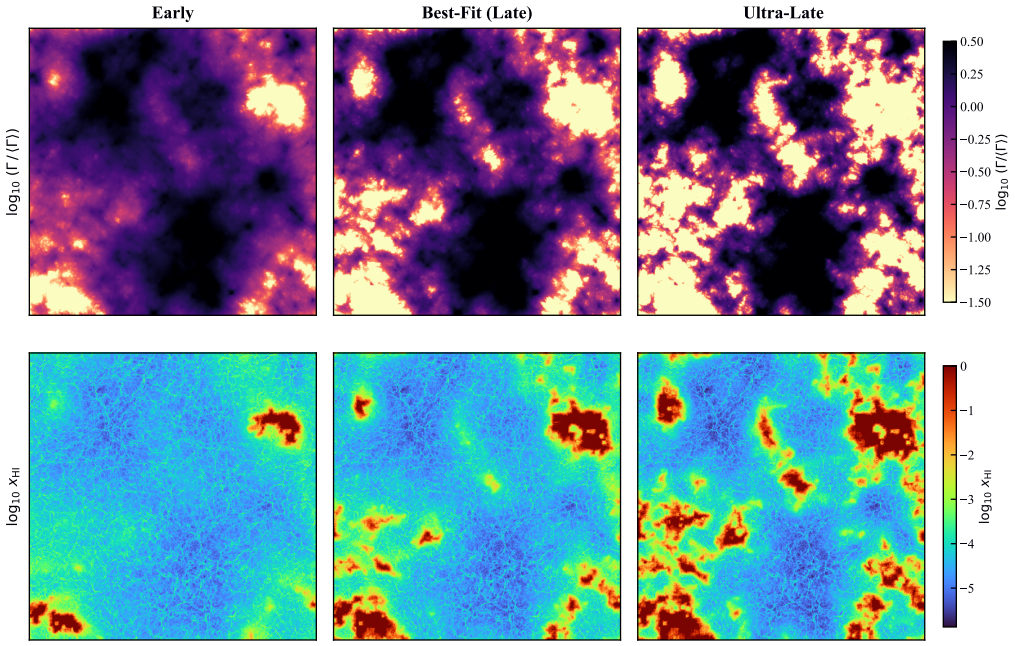


Figure 4.1: The slices of the simulation models plotted for photoionization fluctuations and neutral fraction at $z = 5.95$. Left to Right: Early Model (1σ lower limit of τ_{eff}), Best Fit (Fiducial) Model, Ultra-Late Model (1σ upper limit of τ_{eff}).

We need to see how models are generated and how we can map from the $\langle \Gamma_{HI} \rangle - \lambda_0$ parameter space to $\langle \Gamma_{HI} \rangle - \lambda_{mfp}$ space for seeing the variation in the mean free path. The photionization fluctuations produced at these 6 redshifts, which vary with λ_0, are used to generate the models, and a total of 1.36 million CPU hours are used to run all the models Gaikwad et al. (2023). The fluctuations are generated on a 512^3 grid, and we are able to successflully capture the fluctuations using Equation 4.1 and Equation 4.2. The photoionization fluctuations are also varied with the λ_0 parameter, and then normalized to 1. Now, we calculate the continuum optical depth along 2048 skewers in the simulation box:

$$\tau_{LyC} = \int n_{HI} \sigma_{HI} dx$$

where $\sigma_{HI} = 6.34 \times 10^{-18}\text{cm}^2$ represents the HI ionization cross-section, n_{HI} represents the number density of the neutral hydrogen and dx is in comoving distance(h^{-1}cMpc). The optical depth calculated along each sightline can then be turned into flux by $F_{LyC} = e^{-\tau_{\text{LyC}}}$. Then we average out these 2048 profiles, similar to stacking the spectra at 912 Å in observations. We then fit the continuum transmission profile using two free parameters (F_0, λ_{mfp}) as:

$$\langle F_{LyC} \rangle (x) = F_0 \text{exp}\left(\frac{-x}{\lambda_{mfp}}\right)$$

where F_0 is the normalization of the profile, λ_{mfp} is the true mean free path and x is in h^{-1}cMpc. This true mean free path is then matched with the observations and the corresponding photoionization rate is used to scale the fluctuations to produce the Ly-α forest. In this way, we make sure that we can generate models of varying mean free path.

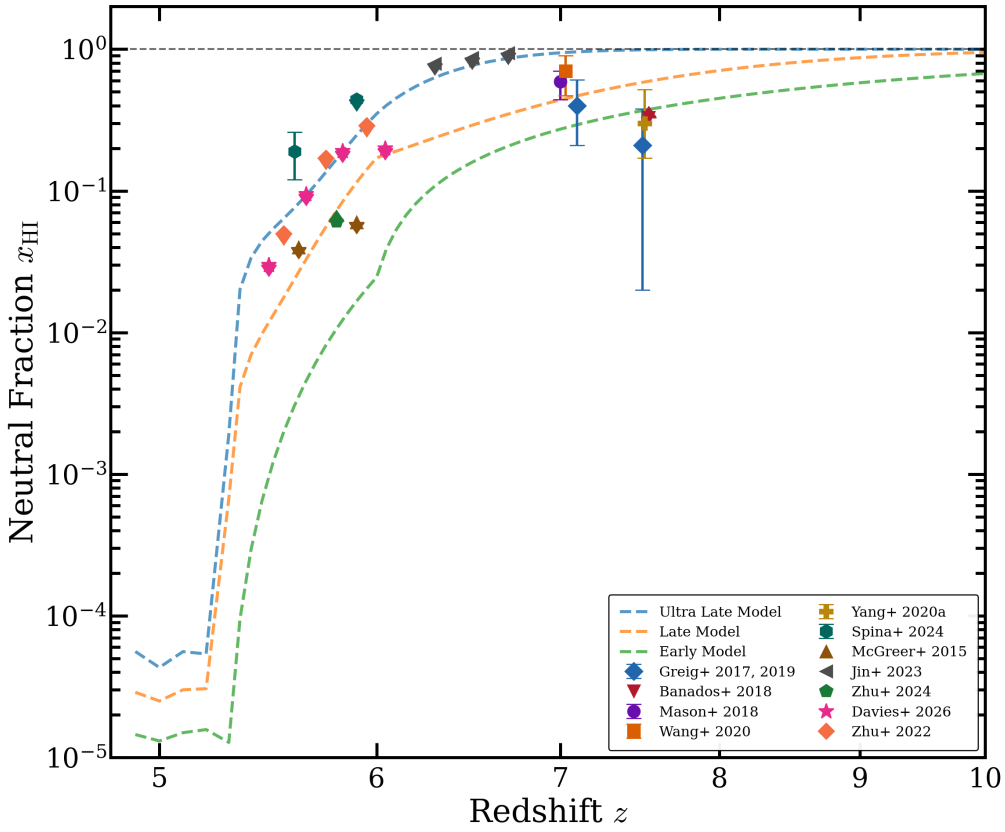


Figure 4.2: Evolution of neutral fraction with redshift across different models. The red dotted curve shows the Ultra-Late Model, orange one shows the Late Model (fiducial), and green shows Early Model. The observations include from damping wing analysis (Greig et al., 2017; Bañados et al., 2018; Greig et al., 2019; Wang et al., 2020; Yang et al., 2020), Lyman-α forest analysis (Spina et al., 2024; Zhu et al., 2024), and dark pixel fraction (McGreer et al., 2011; Mason et al., 2018; Jin et al., 2023; Davies et al., 2026).

The models are generated using the photoionization and are described below:

- **Best-Fit (Late) Model**: This reionization model evolves from $\lambda_0 = 8.318$ at $z = 6.0$ to $\lambda_0 = 46.452$ at $z = 5.1$. The redshift at which the reioniza-

tion completes $50\% = 7.20$. This model produces best-fit τ_{eff} scatter i.e it is a fiducial model.

- **Ultra-Late Model** : This reionization model evolves from $\lambda_0 = 4.266$ at $z = 6.0$ to $\lambda_0 = 27.543$ at $z = 5.1$. The redshift at which the reionization completes $50\% = 6.15$. This model represents 1σ (68 %) upper limit of τ_{eff} scatter.

- **Early Model**: This reionization model evolves from $\lambda_0 = 15.849$ at $z = 6.0$ to $\lambda_0 = 84.54$ at $z = 5.1$. The redshift at which the reionization completes $50\% = 8.35$. This model represents 1σ (68 %) lower limit of τ_{eff} scatter

The photoionization fluctuation maps and neutral fraction in different models is plotted in Figure 4.1; depicting physically sound reionization scenarios for each model. Figure 4.2 illustrates the evolution of neutral fraction across different reionization histories. There are different observations done by previous studies using damping wing analysis, dark pixel fraction, and Lyman-α forest analysis and is plotted with respect to the neutral fraction measurements done by these models. The model measurements are well in agreement with the observations, affirming that the bounds placed by the measurements in $\frac{\Gamma_{\rm HI}}{\langle\Gamma_{\rm HI}\rangle}$ and $\lambda_{\rm mfp}$ are accurate. We introduced this scaling parameter f_Γ to explore a large space of neutral fraction values without changing any temperature, density, or velocity fields. This fact will be exploited further in χ^2 minimization analysis (see Chapter 6) of the dark gap statistics.

Although, these simulations optimizes computing efficiency and parameter exploration, there are some important caveats that should be addressed. The sources use a simplified model, not considering stochasticity and feedback from the galaxies. No galaxy formation physics is included, all the parameters in the above formalisms are assumed to take care of the astrophysics. There are smaller uncertainties in the sub-grid approximations, and thermal modeling is also simplified (two-zone model).

Throughout the following chapters, we will be including the convention of scaling the mean photoionization rate i.e. $\Gamma_{\rm HI}(\rm x) = f_\Gamma \langle\Gamma_{\rm HI}\rangle$. We may use the mean photoionization rate and the scaling parameter interchangeably, since they are related by a simple multiplicative factor.

CHAPTER 5

Methods

Although the forward modeling of Lyman-α forest allows us to constrain the photoionization rate, mean free path via τ_{eff} scatter, but it does not give detailed information about the ionization structure of the IGM. This is due to the fact that this statistic covers average transmission along the sightlines i.e it is an integrated effect. There are a lot of degeneracies present in the τ_{eff} scatter distribution i.e different combinations of $\lambda_{\rm mfp}$ and $\langle\Gamma_{\rm HI}\rangle$ can produce similar distributions. In order to break this degeneracy, a different statistics to analyze the ionization state of the IGM should be considered : Dark Gap Statistics.

We define dark gaps as the continuous regions lying in the quasar sightline having a continuum-normalized flux below a certain threshold. The Ly α flux is suppressed in these regions, and these regions may arise from highly saturated regions or a fluctuating UVB. The lengths of the dark gaps reveal information about fluctuations and neutral content of the IGM. Dark gap statistics will reveal the distribution of dark gaps, i.e., extended regions of extremely high optical depth.

The transmitted flux is normalized with respect to the continuum of the quasars, and the spectra are rebinned to a comoving resolution of $1h^{-1}Mpc$. We had defined the flux threshold to be 0.05, taken from the quasars with the minimum signal-to-noise ratio (SNR) (see Section 4.1). The minimum length of dark gaps is taken to be $1h^{-1}Mpc$. Although the EXCITE simulations allow for a much higher resolution (see Section 4.2), we use a lower resolution to match with the observations.

We mask the contaminated regions by avoiding proximity regions, which are regions closer than seven proper Mpc distance from the quasar. We restrict our analysis to Ly-α redward of the Ly-β, i.e., around 1041 Å, and absorption takes place on the red side of this limit.The corrections for the telluric absorption and sky limited residuals have already been accounted for in the observations. Figure 5.1 gives the full pipeline for automated identification of the dark gaps using different stages: first, we convolve the generated Lyman-α sightline with the resolution of the XQR-30 (D'Odorico et al., 2023) to match the instruments. After that we do rebinning to 1 h^{-1}Mpc as stated above. In this stage only, we add the Gaussian random noise based on which quasar realization we want. Finally, we identify the dark gaps using the criteria $F_{\rm norm} < 0.05$ as stated above. We record the dark gaps by considering whether the transmission pixels are falling below this threshold and this is considered for the whole continuous region. The lengths and their starting and ending indices in the redshift grid is saved for using

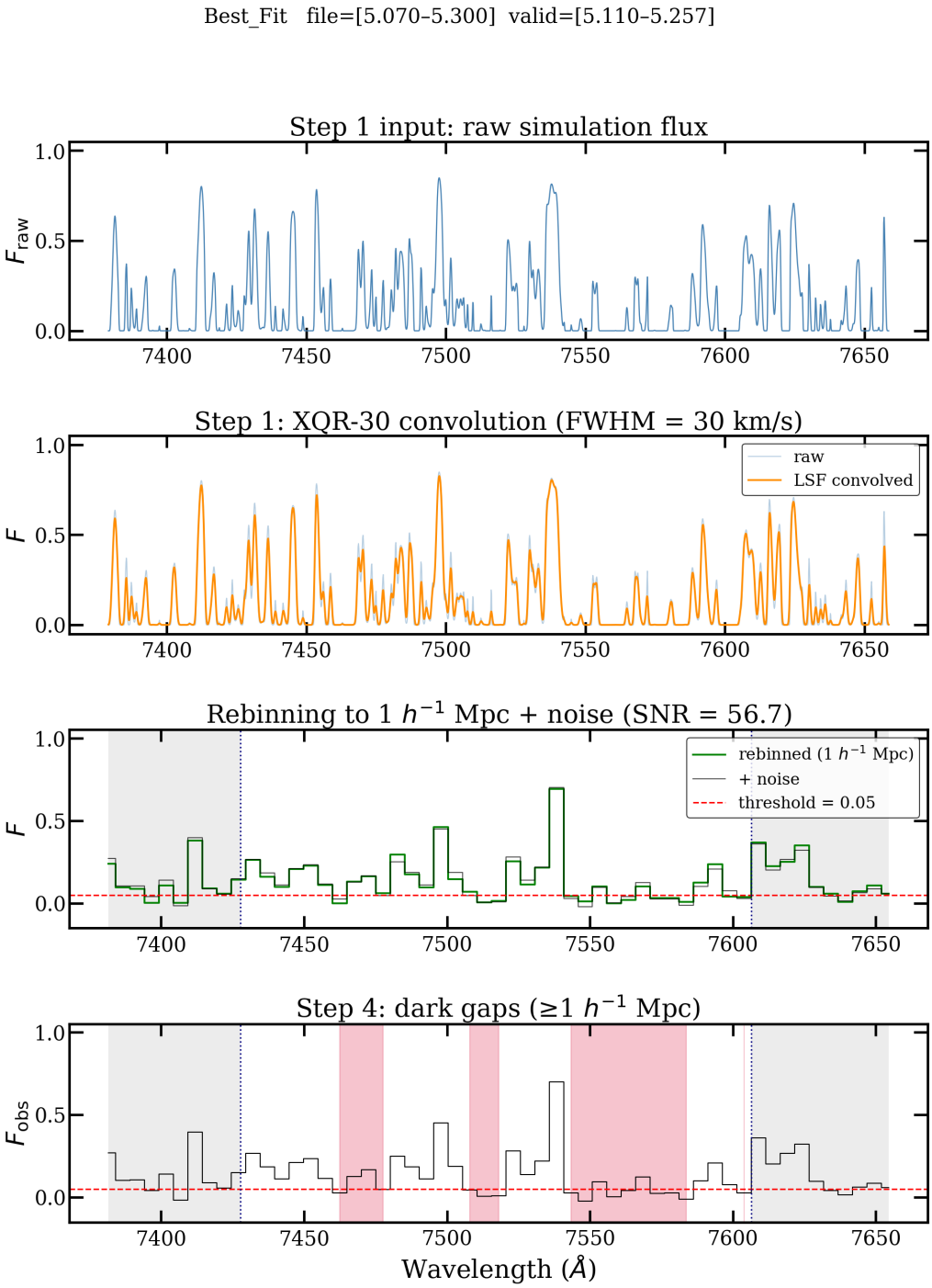


Figure 5.1: Dark gap detection steps illustrated in an example sightline,, at $z = 5.11 \rightarrow 5.26$. The steps involve matching the resolution with the instrument and then the orange curve represents the continuum normalized flux obtained from the synthetic spectra, and it is smoothed to $1\,h^{-1}$Mpc. Shaded regions represent the dark gaps obtained in the sightline, and the red dashed line represents the threshold $F = 0.05$.

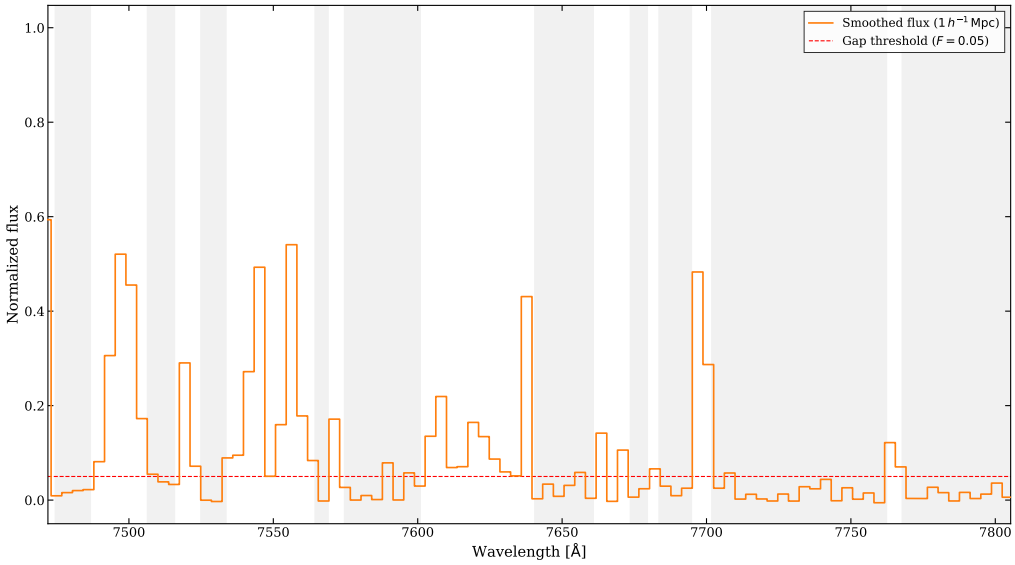


Figure 5.2: Dark gap detection is illustrated in an example sightline, PSO J183-12, at $z = 5.97$ with a signal-to-noise ratio of 61.8. The orange curve represents the continuum normalized flux obtained from the synthetic spectra, and it is smoothed to $1\,h^{-1}$Mpc. Shaded regions represent the dark gaps obtained in the sightline, and the red dashed line represents the threshold $F = 0.05$.

them in dark gap statistics.

We identify dark gaps in approximately 2000 sightlines, corresponding to 45 independent realizations of 38 quasars sightlines, by utilizing an automated dark gap detection pipeline. We perform our analysis for three extended reionization scenarios: Ultra Late, Late, and Early Reionization (see Section 4.2). Figure 5.2 shows an example sightline in our simulation corresponding to sightline PSO J183−12 located at $z = 5.97$ with a signal-to-noise ratio of 61.8. We have rebinned the spectrum to $1\,h^{-1}$Mpc and the shaded regions represent the dark gaps. Otherwise, longer dark gaps, the EXCITE simulations have a higher native resolution, leading to separation of these long dark gaps.

Saturated regions involving neutral hydrogen cannot be directly inferred from the observations since Lyman-α is a resonant transition. Each point in the Figure 5.3 represents a dark gap identified for a single realization across all the sightlines for Ultra Late Reionization. The x-axis in the plot indicates the location of the dark gap in this realization. The y-axis represents the dark gap length detected at each redshift. Although there is a significant scatter, we can still discern the key physical behavior: Dark gaps have shorter lengths at lower redshift ($z \sim 5$), representing a Universe that is predominantly ionized whereas at higher redshifts ($z > 5.5$), we see a broadening of the distribution and longer dark ($L > 10\,\mathrm{h}^{-1}$Mpc) gaps become more common. There is a large scatter for all the redshifts, indicating line-of-sight variance. This scatter is also evidence for the patchy reionization scenario. By changing the reionization history, we are unable to detect changes in the distribution due to the large scatter. So it is justifiable to use dark gap statistics rather than using individual dark gaps as indicators of saturation.

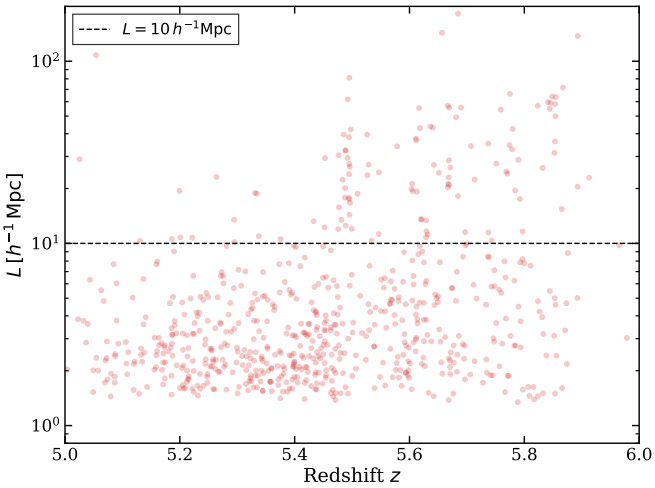


Figure 5.3: Dark gap lengths as a function of redshift z for the Ultra-Late reionization scenario. Each point represents an individual dark gap identified across all sightlines. The black dashed line marks the threshold $L = 10\,h^{-1}$ Mpc used to define long dark gaps. Longer dark gaps become increasingly prominent toward higher redshift.

Dark Gap Statistics

From the Figure 5.3, it is clear that the dark gaps occupy a wide range of comoving scale lengths, with some even reaching $100\,h^{-1}$Mpc. The majority of the population lies within the range $< 10\,h^{-1}$Mpc, indicating a rapid decline in the population at the tails of the distribution. Density fluctuations dominate gaps on smaller scales and are marginally sensitive to the neutral fraction. The mean photoionization rate controls these fluctuations. The dark gaps greater than $10\,h^{-1}Mpc$ hold a lot of information as they are more likely to be affected by the UVB fluctuations, reionization history, and residual neutral regions. We will see how this length scale will be used in defining statistics of long dark gaps.

Cumulative Distribution Function (CDF):

The cumulative distribution (CDF) of the dark gaps can be used to understand effect of reionization history, mean free path, and photoionization rates, on the saturated regions. The CDF quantifies the probability of a dark gap having length greater than some length L_0, i.e., $P(L \leq L0)$. All the models employ spatial fluctuations (refer Section 4.2), giving a wider distribution of the tails in the CDF. The distribution shape is also affected by the ionization state of the IGM. If IGM is more neutral, it will result in much longer dark gaps, hence skewing the CDF shape. The dark gaps are binned such that each bin ranging from $5.0 - 5.3$, $5.3 - 5.6$, $5.6 - 6.0$ contains equal number of dark gaps to avoid small count statistics.

The CDF is primarily used for model comparison, and, due to the large number of samples, we expect a smaller p-value, compelling us to use distributions of p-values/D-values across the different realizations of the sightlines. The CDF also varies with the mean free path: at a given length, the CDF shifts to larger values for shorter mean free path suggesting that there are more number of dark

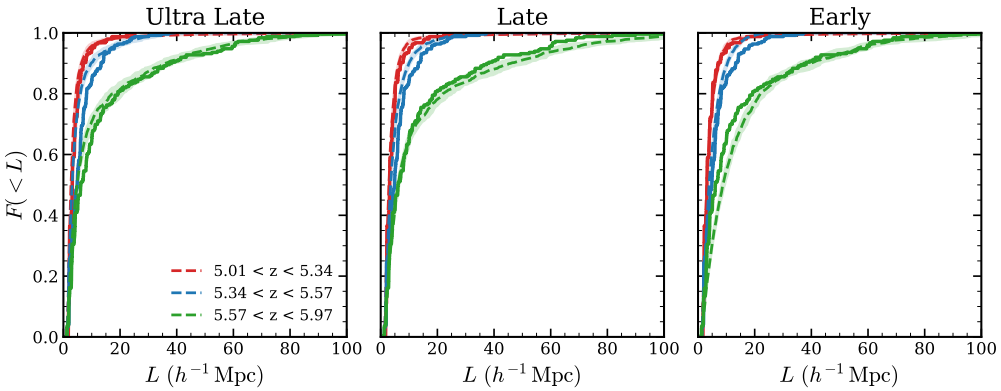


Figure 5.4: Cumulative Distribution Function of the dark gaps plotted as a function of their lengths. The dotted lines represents the data points for the fiducial model and solid lines are the observations (Zhu et al., 2021). This plot serves an χ by eye-test to check whether the models are able to produce the observations.

gaps present with shorter length. Although the shape of the CDF is less dependent on the choice of reionization history, the photoionization rate and neutral fraction content affect the CDF equally. The uncertainties are addressed by considering different realizations of the same sightlines, and the evolution of the observations relative to the models across redshift bins can be inferred from the D-distributions of their CDFs.

Probability Distribution Function (PDF):

Although the CDF provides a good idea of the evolution of the neutral fraction, it does not provide error bars for model comparison. The independence of the redshift bins is less problematic in the PDF than in the CDF, an assumption we use throughout this study. The neutral fraction in different simulations is usually inferred from the cumulative distribution of the effective optical depth (Gaikwad et al., 2023; Kulkarni et al., 2019; Davies and Furlanetto, 2016). But the probability distribution of dark gaps gives the hidden information that is averaged out in the effective optical depth analysis. In $log(\Delta \mathrm{z})$, the probability distribution is defined as the probability of a randomly selected dark gap lying in a certain interval with constant equi-spaced logarithmic bins. The empirical choices do affect the shape of the PDF, but we have taken consistent, constant bins and bin width to retain all the important information needed for the dark gap analysis. We consider bin parameters (to retain the shape of the PDF) in log-space as: $log(\Delta z)$ is from $-3.0 \rightarrow 0.05$ with a bin-width of 3.82×10^{-4}. These bin parameters are kept constant throughout the calculation of the global PDF.

The probability distribution function is computed by uniform binning in logarithmic redshift space for each mock realization, then averaged over all realizations. This method is also used to calculate the variance across all the mocks and to estimate the bin-wise uncertainties. For such a global PDF defined at the intervals of $[5.0, 6.0]$, we mask the observational PDF where its value is less than 0.5 to avoid noisy measurement in the tails of the distribution.We chose this hard cut-off to avoid the bimodal nature of the χ^2 (although this leads to loss of the information, refer to Appendix A for more details). The error bars on the PDF are

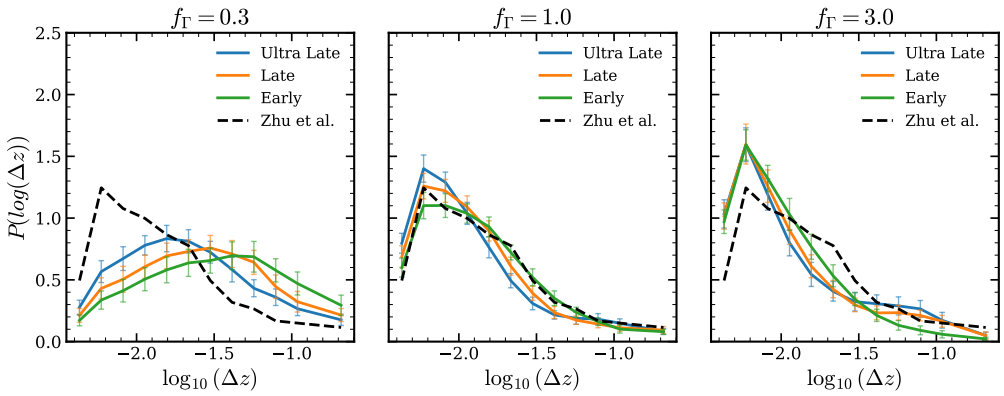


Figure 5.5: The global PDF of different reionization histories across redshift $5.0 \leq z \leq 6.0$ as a function of photoionization rate.Left to Right: The scaling values change from $f_\Gamma = 0.3 \rightarrow 3.0$. The mean photoionization rate for this redshift range is taken to be $\Gamma_{12} = 0.35$. The dashed lines represent the observations (Zhu et al., 2021) and the solid lines indicate different models based on z_{50} (Ultra Late: Blue, Late(Best-Fit) Model: Orange, Early Model: Green Solid lines respectively) with error bars representing the 68% uncertainties on the PDF.

not only coming from the Poisson counting uncertainty but also from the scatter among the different mock realizations for each model and f_Γ value. This associated uncertainty is accounted from the sample standard deviation across the mocks. The following equations effectively summarizes the method to generate error bars on the PDF:

$$\bar{p}_k = \frac{1}{N_{\text{mock}}} \sum_{m=1}^{N_{\text{mock}}} p_{m,k} \tag{5.1}$$

Here, $\bar{p}_k$ represents the mean PDF taken across all the mocks for each bin index k. $N_{\text{mocks}} = 45$ represents the total number of mocks and $p_{m,k}$ represents the PDF of each mock realization for each bin index k.

$$\sigma_k = \sqrt{\frac{1}{N_{\text{mock}} - 1} \sum_{m=1}^{N_{\text{mock}}} (p_{m,k} - \bar{p}_k)^2} \tag{5.2}$$

Here, σ_k represents the standard deviation for each bin index k. These error bars effectively model uncertainties from the finite sampling.

We also compute this statistic for different redshift bins. To achieve a smoother evolution, we use Kernel Density Estimation (Figure 5.6) to plot the PDFs of the observations and simulations (approximately 200 dark gaps per bin). We have the bin-width for the KDE as : bw $= 3.5n^{-1/3}$, where n is the grid length. This smooths out most of the small-feature artifacts introduced by the limited number of sightlines in the observations. The only difference between the binned PDF and the global PDF is that the binned PDF is now dependent not only on the mean photoionization rate but also on the redshift evolution. The redshift evolution of the neutral fraction will be determined by varying the photoionization rates, and the solution that best matches the observations will yield an optimized mean photoionization rate. This optimized photoionized rate will provide the

neutral fraction associated with it. The redshift bins will determine the error on the $log(\Delta z)$, and the neutral fraction will depend on the type of reionization history present. The error calculation in the KDE is identical to that of the global PDF: taking the mean PDF across all the bin indices and then the sample standard deviation to compute the uncertainty.

Figure 5.5 illustrates that the PDF is not a direct indicator of neutral regions. The PDF is a composite statistic with numerous processes affecting its shape, width, tail, amplitude, etc. The reionization history primarily determines the shape and skewness of the tail distribution. However, variation in the mean photoionization rate is degenerate with the reionization history. Considering the mean photoionization rates in different regimes, we observe different behavior in the reionization topology. At lower f_Γ, the mean free path and UVB fluctuations are major factors in deciding the abundance of the dark gaps. The occurrence of the short dark gaps in the Late Model is more compared to that of the Early Model due to shorter mean free path.

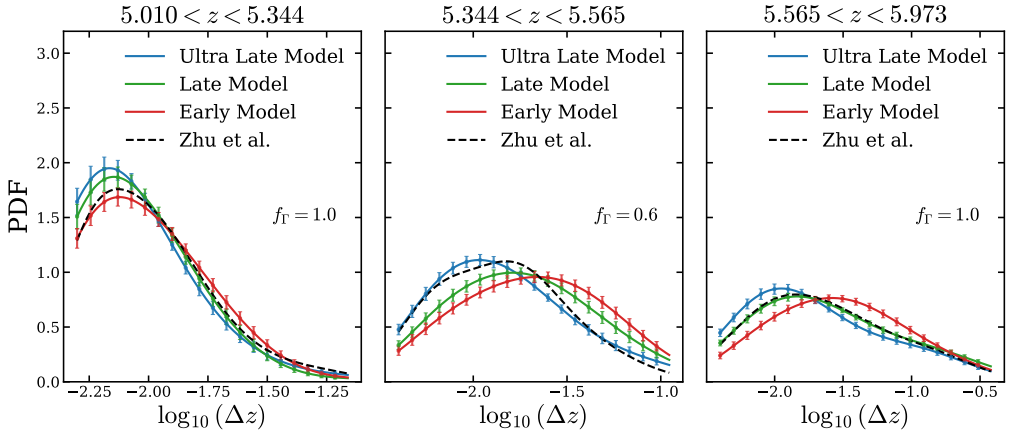


Figure 5.6: The KDE plotted for the distribution of the dark gaps of different models across different redshift models. The dotted line represent the observations (Zhu et al., 2021) whereas the solid lines represent the simulation KDE curves. We get suppressed f_Γ in second redshift bin due to averaging of the photoionization rate due to high opacity fluctuations via rapid evolution of neutral fraction in this redshift bin.

The PDF shape and peaks are more sensitive to the photoionization rate than to the reionization history. At low f_Γ, the PDF is broad, and the peak shifts towards larger dark gap length scales. This feature is due to a weak ionizing background, hence making the dark gaps longer. As f_Γ is increased, the PDF becomes narrower and the peak shifts to the left side due to more transmission by a stronger ionizing background. For a given value of f_Γ, the reionization history cannot alter the shape of the PDF; i.e., they overlap. In this regime, the IGM is optically thin; thus, the bubble size or reionization topology does not affect the scales of the dark gaps. The neutral regions are scarce; hence, the probability of the tails of the PDF distribution approaches zero. The mean photoionization rate matches the observation approximately around $f_\Gamma = 1.0$. The PDF is normalized per mock allowing us to compare the mocks with different total number of gaps.

F10 Statistics:

To study the evolution of long dark gaps, which provide much more stringent constraints on the neutral fraction due to the sensitivity of the longer saturated regions being more sensitive to the neutral fraction. We consider statistics of long dark gaps to probe into long saturated regions: F10 Statistics. F10 statistics is the cumulative distribution of all the dark gaps greater than $10\,\mathrm{h}^{-1}$Mpc, probing the tail end of the PDF distribution. Previous works (Zhu et al., 2021; Maity et al., 2026) use F30 curves for model comparison and consistency.

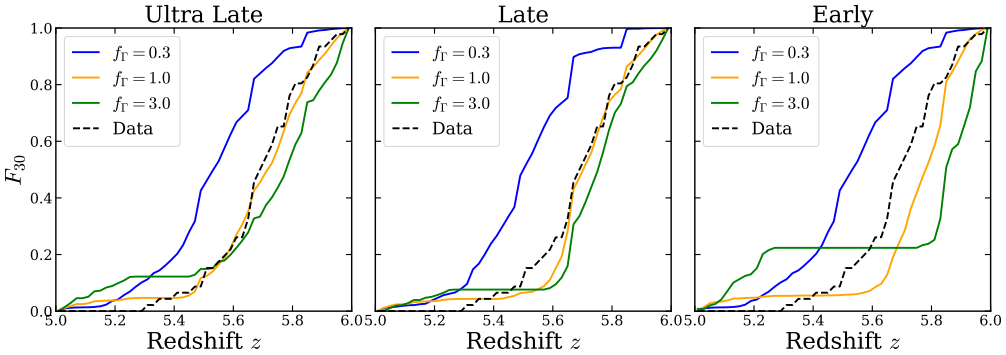


Figure 5.7: Cumulative F30 Statistics Distribution for different reionization models and mean photoionization rate (Left to Right: Ultra Late Model, Late Model, Early Model). The long dark gaps are sensitive to the mean photoionization rate and the observation is well within the extreme values of the mean photoionization rates. All the reionization histories show similar distribution, strengthening the argument that longer dark gaps are not affected much by the reionization histories.

While F30 statistics (see Figure 5.7) traces more extended saturated regions, the abundance of the dark gaps reduces drastically. Hence, we can use F10 statistics ($L > 10\,h^{-1}$Mpc), which represent a trade-off between sensitivity to neutral regions and statistical power. F10 includes a significantly larger fraction of dark gaps, which helps reduce small-count statistics. It is essential to note that changing the threshold only probes a different part of the same distribution; hence, it is expected that the shape of the curves remains the similar in both F10 and F30 Statistics. We will consider F10 statistics, as the same reionization physics drives the constraints. We can see that the Figure 5.8 depicting F10 statistics is smoother than the F30 statistics (Figure 5.7). This feature also motivates us to use F10 statistics for consistency.

Figure 5.8 illustrates the F10 distribution of the dark gaps for different reionization histories: Ultra Late, Late, and Early Models ($z_{50} = 6.15, 7.2, 8.35$, respectively). Figure 5.8 illustrates that regardless of the reionization history, the F10 statistics is more dependent on the the mean photionization rate. The shape of the distribution is retained across all models, and it is monotonic, implying that different reionization histories produce similar dark gap statistics by adjusting the mean photoionization rate. We can also infer that the abundance of long, dark gaps is primarily controlled by the ionizing background, rather than the reionization history. The observations are in good agreement with the simulations, i.e., they lie within the extreme photoionization rates. The lower photoionization rate

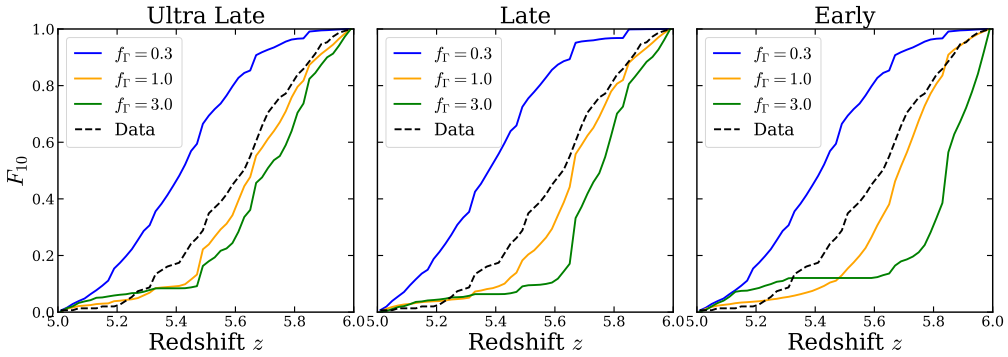


Figure 5.8: F10 Statistics evolving with the redshift for different reionization histories and different mean photoionization rates(Left to Right: Ultra Late Model, Late Model, Early Model). This acts as an eye-test we see that all the models produce similar shape as the observations (Zhu et al., 2021).

values result in an excess of long, dark gaps, whereas higher values suppress the abundance of long gaps.

F10 shifts leftwards on decreasing f_Γ, showing that longer dark gaps are present at earlier redshifts. This result is expected because lower photoionization rates correspond to higher neutral fractions. Consequently, smaller photoionization values have steeper curves than larger ones, implying that the PDF of these curves shifts to larger length scales. As we have seen, the mean photoionization rate is treated as a free parameter and serves as a parameter to vary the neutral fraction. Visually, the consistency between the F10 statistics and the observations is evident around $f_\Gamma = 1.0$ regardless of the reionization history.

The slope of the F10 defines the rate at which dark gaps appear as we move to higher redshifts. It is unlikely that dark gaps greater than $10\,h^{-1}$Mpc are produced by ionized regions since the probability of UVB fluctuations to produce such long dark gaps is very less. Hence, we can assert that the slope traces the timing of neutral island disappearance. F10 is also insensitive to the small-scale topology of the IGM, as it is computed by integrating many sightlines, making it a robust statistic for determining the neutral fraction. Even though there is a shift in reionization histories, their similar behavior asserts the argument that the reionization timing is degenerate with UVB fluctuation amplitude.

The dark gaps are regions of high optical depth, and the optical depth is controlled by the geometry of neutral regions and the amplitude of the neutral fraction. The former is decided by the patchiness of the reionization and the latter by the mean photoionization rate (f_Γ). Dark Gap Statistics infer the integrated effect; it is unable to distinguish the source of photoionization leading to ionized regions. A different combination of mean photoionization rate and reionization history can produce the same effect: indistinguishable Ly-α absorption leading to degeneracy.

Reionization history describes the timing of the disappearance of neutral regions. Earlier models predict more isolated neutral regions within ionized regions, whereas later models predict longer dark gaps. The mean photoionization rate decides the residual neutral fraction. So, we can have late reionization models with higher photoionization rates or early reionization models with lower

photoionization rates. We cannot tell whether a dark gap is caused by neutral hydrogen (longer dark gaps constrained by z_{50}) or by a typical neutral fraction spread over a considerable length with a strong ionization background. The parameter space (z_{50}, f_Γ) will be crucial in understanding what kind of neutral fraction can be obtained evolved over the redshift range $5.0 < z < 6.0$.

We can use dark gap statistics in the form of PDF and cumulative F10 statistics. The PDF addresses smaller and intermediate scales and is sensitive to the mean photoionization rate (f_Γ). The PDF has a weak dependence on the reionization history, changing only the peak at extreme scales. Cumulative F10 statistics deal with the evolution of neutral regions with the redshift, i.e., timing of reionization. The disappearance of the neutral islands can trace the reionization timeline. F10 statistics is weakly dependent on the photoionization rates, making it constrain reionization history (z_{50}), given a photoionization rate. We will obtain an optimized mean photoionization rate for each reionization history model by matching observations using a χ^2 analysis (since we have only one free parameter, f_Γ).

Neutral Fraction Constraints via χ^2 minimization:

The optimization of the free parameters is done by χ^2 analysis. The simulation curves in the PDF match the observations. We obtain different mean photoionization rates, i.e., f_Γ values, using χ^2 analysis. Since we have finite scaling values with neutral fractions, we interpolate between the simulation values to obtain the final result. The PDF and F10 plots exhibit monotonic, smooth behavior, allowing us to proceed with this interpolation. The residuals are calculated at each redshift bin center, and the variance is estimated from sample variance across all simulation mocks. The simulation mocks provide independent realizations of the same sightlines to capture cosmic scatter. This is used in calculating χ^2 for each redshift center : $\chi^2_{model} = \Sigma_i \frac{(S_i - M_i)^2}{\sigma_i^2}$, where we calculate χ^2_i for each redshift center. Low-variance regions in the tails of the distribution are discarded by our earlier method of masking regions with the condition $P_{obs} < 0.5$. We compute χ^2 for different mean photoionization rates and plot the χ^2 vs interpolated f_Γ. The minimum point on this curve corresponds to the optimized mean photoionization rate. For different reionization histories, this rate varies, and we obtain an optimized parameter space for these two parameters that can reproduce the observations, indicating a degeneracy between the reionization history and the mean photoionization rate.

The confidence interval on the f_Γ parameter can be found out by considering $\Delta\chi^2$. Since the degrees of freedom dof $= 1$; we consider the following values of $\Delta\chi^2$:

$$\Delta\chi^2 = \chi^2 - \chi^2_{\text{min}} = 1.0 \text{ for } 68\% \, (1\sigma)$$

$$\Delta\chi^2 = \chi^2 - \chi^2_{\text{min}} = 4.0 \text{ for } 95\% \, (2\sigma)$$

The primary objective of this work is to determine whether a specific reionization history can accurately reproduce the observations for some physically allowed photoionization rate. All the statistics have been calculated as an ensemble over many skewers with Gaussian scatter over the redshift bins. The

mean photoionization rate f_Γ is not a reionization parameter; rather, it encapsulates all unresolved uncertainties in the ionizing emissivity and related factors. This is the only effective parameter, and it is constrained over three discrete reionization histories, i.e., z_{50}, so χ^2 is sufficient for the parameter estimation.

EX-CITE simulations (refer to Section 4.2) encapsulate all calculations of fluctuations in the mean photoionization rates and mean free path. The optimized scaling values obtained from dark gap statistics can be converted into a neutral fraction by interpolation. We use piecewise linear interpolation since the behavior of the statistics is smooth and monotonic. Having discussed the methods, we now apply the above framework to present the constraints on neutral hydrogen fraction, i.e, x_{HI}, using dark gap statistics and χ^2 analysis.

CHAPTER 6

Results

In this section, we will now discuss all the primary results from dark gap statistics: CDF, PDF and F10 Statistics. We will examine the χ^2 minimization curves and then find constraints on the upper bounds of the neutral fraction.

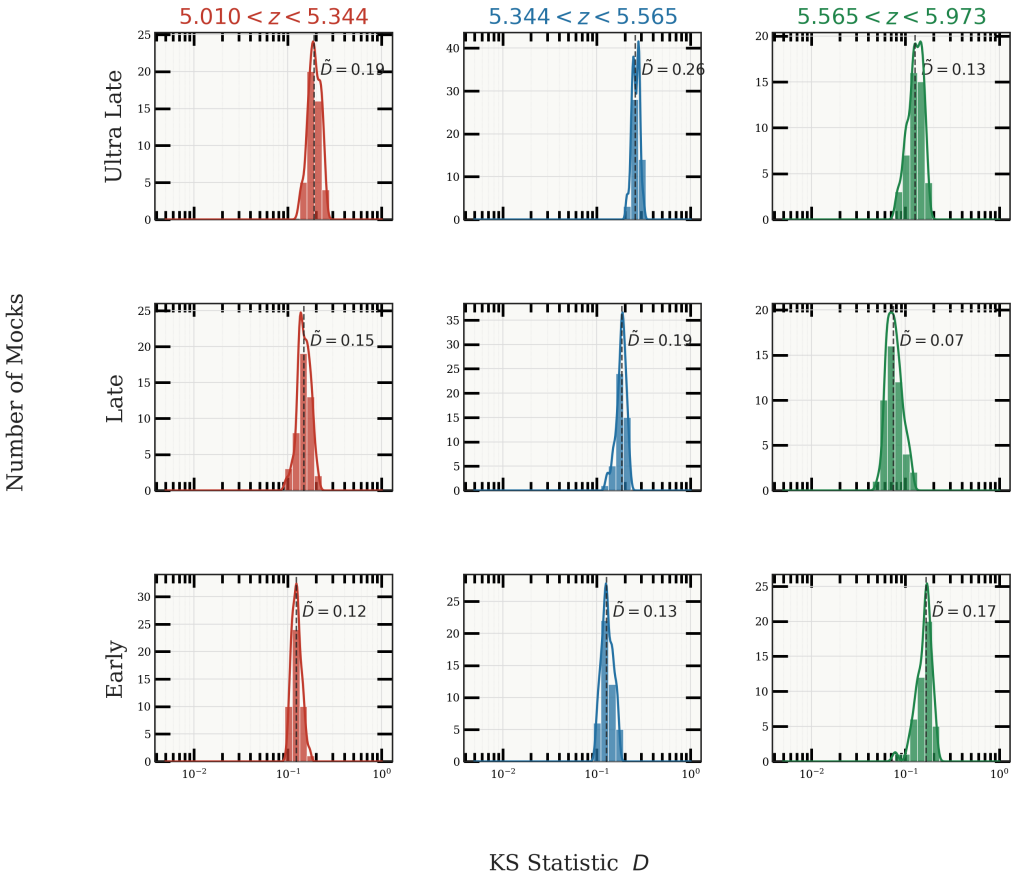


Figure 6.1: The D-Value Distribution plotted across different mocks for different redshift bins and models. The tightness and low value of KS statistic D are indicators of the closeness of the models to the observations.

We consider D-value distribution across different mocks and then seeing the lowest median D-values to see which model is preferred at what redshifts (KS-2 test fails due to large sample size). The D-value is the supremum of the difference between the observation and the simulation CDF curves. Smaller D value results in better agreement than the larger D case. Figure 6.1 illustrate the D-

value distribution across different mocks for different redshift bins. In Figure 6.1, for $5.0 \leq z \leq 5.3$, D-values are smaller and tightly clustered compared to other redshift bins giving a better fit for the models. For $5.3 \leq z \leq 5.6$, there is a shift to higher D-values and the tension grows for all the models. For $5.6 \leq z \leq 6.0$, we get the poorest fit out of all the redshift bins. The models are systematically worse at higher redshift bins. For confirming the quantitative values of the D-distribution, we consider the median of the D-values across all the mocks for each model and each redshift.

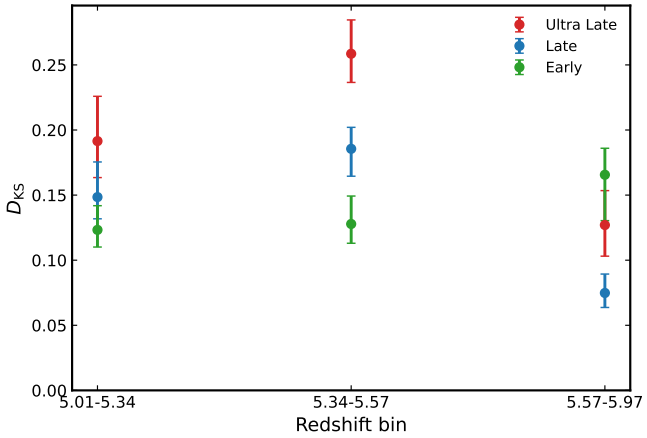


Figure 6.2: Median value of the D distribution across different redshift bins. The preference of the models can be easily inferred: For $5.0 \leq z \leq 5.3$: Early Model; For $5.3 \leq z \leq 5.6$: Early Model; For $5.6 \leq z \leq 6.0$: Late (Best-Fit) Model

Figure 6.2 depicts the D median values for all the models across different redshift bins. These values are taken from the D-value distribution and the preference to the models for redshift bins $(5.0, 5.3)$, $(5.3, 5.6)$, $(5.6, 6.0)$ is Late, Early and Early Model respectively. It is important to note that we see the CDF as a diagnostic tool to compare different reionization history and how different z bins change the preference of these models.

For a given value of the reionization history, we constrain the mean photoionization rate using a χ^2 analysis; hence, we perform marginalization of the reionization history, z_{50}. We calculate χ^2 for the logarithmic redshift center bins, and the total number of these bins will give us the degrees of freedom for the PDF. Figure 6.3 illustrates the minimum χ^2 curve for Early, Late(Best-Fit) and Ultra-Late to obtain an optimized f_Γ value for global redshift range of $5.0 \leq z \leq 6.0$.

The minimum χ^2 over all the combinations of (f_Γ, z_{50}) will be taken as the reference for the parameter contour plots. We use only one parameter: mean photoionization rate, i.e., f_Γ, which encapsulates the effects of all the reionization physics in the IGM. We do not consider the covariance matrix, as calculating covariance from sightline-limited observations would be noisy and not meaningful. The bin correlations in the PDF are nonlinear and model-dependent. Thus, using a single covariance matrix for all the parameter space will be inconsistent. Mock realizations have the LOS variance that is already captured by the diagonal variances. This also means that the dominant uncertainty is due to sample

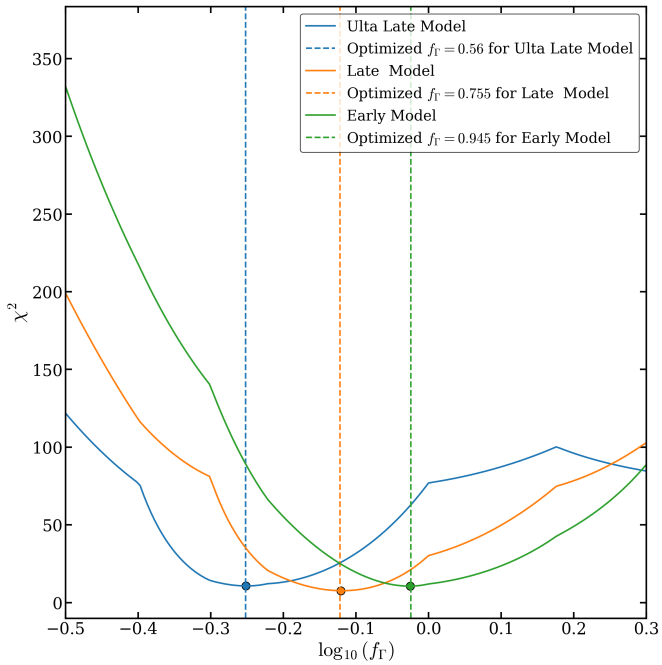


Figure 6.3: Minimum χ^2 curve plotted for different models over $5.0 \leq z \leq 6.0$. The optimized values for f_Γ are as expected for different models. All three different reionization histories are able to reproduce the PDF for a certain value of f_Γ indicating the degeneracy in the dark gap statistics.

variance in all the realizations. We treat all the redshift bins as independent and consider only the diagonal elements of the covariance matrix.

In Figure 5.5, the photoioinization rate (f_Γ) sets the mean free path of the ionizing photons and the shape is more sensitive to f_Γ than reionization history (z_{50}). At low f_Γ, the neutral regions are larger, meaning long opaque stretches surve and the simulation PDF develops a broad tail towards larger Δz. As f_Γ is increased, the PDF gets sharpened and shifts to the left side. At very high f_Γ, the IGM is more ionized resulting in shorter dark gaps (faster ionization of long dark gaps), hence narrows the PDF and peaking at smaller Δz. In this regime ($f_\Gamma = 3.0$), all reionization histories produce similar PDF since topology becomes irrelevant and all neutral regions get ionized at the same rate irrespective of their length.

In 5.5, for a given photoionization rate, the reionization history (z_{50}) controls the topology of the neutral gas. In case of Ultra-Late reionization, the reionization is not complete and the neutral content of the IGM is very high. The PDF is flattened and broad since the all length scales of the dark gaps are found in the PDF. This model has short mean free path, meaning the reionization is more patchy compared to other models. This also produces more number of shorter dark gaps than the Early Model. In contrast to the Ultra-Late Model, Early Model makes the PDF more concentrated at smaller intermmediate length scales and we see a steeper drop off at large gap sizes. Since the IGM is more diffused in Early Reionization (high mean free path), the short dark gaps are less in abundance compared to the Ultra-Late Model.

The minimum χ^2 will give the best fit f_Γ value for a particular model. A good fit f_Γ does not guarantee the preference of a specific reionization model.

However, this yields three different scaling values for all three models, thereby forming a degenerate pair of (z_{50}, f_Γ).

There is an enhanced population of longer dark gaps in the Early Model due to extended regions of less ionized regions of the IGM. At intermediate photoionization rates, the early model of reionization suppresses the tail of the PDF distribution and shifts the peak to shorter dark gaps. However, this does not mean that the early model has more abundance in shorter dark gaps. Since the ionization fronts have more time to coalesce in the early model, they are more diffused, and the correlation scales are therefore larger. Subsequently, the PDF trends in the early model are consistent with those of a larger true mean free path. Early models dominate the population of gaps at intermediate scales. Similarly, ultra-late/late models have a broader tail distribution, outlining longer dark gaps due to later reionization in the Universe. The leftward peak of the ultra-late/late model is consistent with a more fragmented ionization, characterized by shorter, darker gaps dominating the distribution, even though long, dark gaps are present in its tails. In a regime of a very high photoionization rate, the dark gaps are dominated by neutral islands, leading to overlap between reionization models at the tails of the distribution.

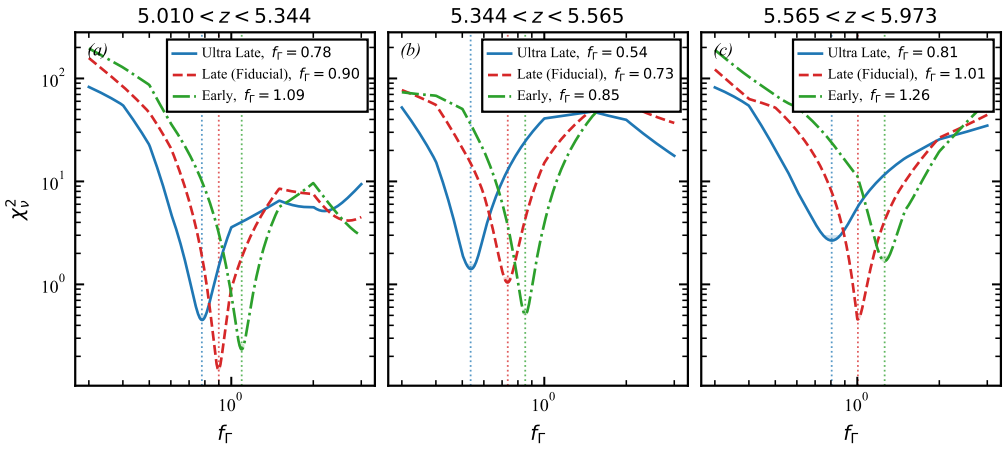


Figure 6.4: Minimum χ^2 curves plotted for different models across different redshifts. The optimized f_Γ values follow similar trend as in global minimization curve.

Figure 5.8 illustrates the F10 statistics for different models and scaling parameter, f_Γ. In the Early reionization model, reionization is more complete at higher redshifts, meaning large neutral islands are ionized earlier. The long dark gaps become rarer at high redshift in this case. F10 rises steeply at $z \sim 5.5-5.8$, and the curves shift to higher redshifts compared to other models. The long gaps that do appear in the Early model are spread over a wider redshift range. Hence, there is no sharp transition within the redshift range $5.0 \leq z \leq 6.0$. That is why we see a flattened curve in the Early Model for this redshift range. In contrast to this model, Ultra Late reionization allows for larger neutral islands to exist at lower redshifts. At redshift range $5.2 \leq z \leq 5.5$, suddenly many long dark gaps appear, producing a steep concentrated rise in F10.

For any given reionization history, the scaling parameter in Figure 5.8 governs the slope and amplitude of the F10 statistics. Decreasing f_Γ results in

steeper F10 curves that rise more sharply over a narrow reshift range. This is due to the concentration of neutral regions at lower photoionization rates. Higher f_Γ values smooth the transition since IGM is more ionized at all epochs. At very high f_Γ, all three reionization histories converge to have similar shapes of F10. When the mean photoionization rate is very high, the mean free path of the ionizing photons becomes very large, and any neutral islands are rapidly photoionized.In this regime, the IGM opacity is no longer sensitive to the reionization topology. F10 Statistics loses its ability to discriminate between different reionization histories. F10 acts as a consistency check; the observation F10 (Zhu et al., 2021) can be reproduced for a suitable f_Γ for any reionization model depicting the degeneracy one more time.

Figure 5.6 discusses the evolution of the KDE PDFs across different redshift bins. The dramatic change of shape depicts the evolution of reionization. Consider redshift bin $5.0 \leq z \leq 5.3$: There is a narrow, sharp peak around $log\Delta z = -2.05$ and all models cluster tightly together around intermediate length scales. At this redshift bin, reionization is mostly complete, and the IGM is ionized, leading to a larger occurrence at small scales. Short dark gaps thus dominate the gap-size distribution. All the reionization histories have converged to a similar ionization state.

For bin $5.3 \leq z \leq 5.6$, we obtain a broadened, flattened PDF. This redshift bin is important for reionization physics because it marks the beginning of the transition epochs in most reionization models. At $z \sim 5.45$, the IGM is in an overlap phase: we get different ionized and neutral regions with different lines of sight. The scatter in sightlines is large, so the effective photoionization rate averages out to a lower value than expected. This is because there is a very rapid evolution of the neutral fraction in this redshift bin (from $10^{-2} \rightarrow 10^{-5}$, see 6.6). This rapid evolution produces large fluctuations in opacity, and with a changing sightline, we get different opacities, lowering the average photoionization rate, as seen in Figure 5.6 ($f_\Gamma = 0.6$). The shadowing effect of the neutral islands is present across all bins, which also lowers the photoionization rate.

For the redshift bin $5.6 \leq z \leq 6.0$, the PDF is more extended to large dark gaps, up to $log\Delta z = -0.5$. In this redshift bin, large neutral islands are abundant and span large length scales. The models also differ, and they peak at higher $log\Delta z$, indicating a more neutral IGM; reionization is not complete in most models. In Figure 5.6, for all the redshift bins, the Ultra-Late Model elevates the PDF at large dark gap lengths. Compared to other models, the Early Model has a more complete reionization. The PDFs are suppressed at large Δz due to low neutral fraction in the IGM. The value of f_Γ is not driving the shape change, but the evolving neutral fraction and neutral island topology.

To ensure the accuracy of dark gap statistics, we use similar methods as mentioned in Chapter 5 for ATON simulations (Kulkarni et al., 2019). This will serve as a consistency check to see whether our dark gap statistics also hold for other simulations, such as ATON. The PDF of ATON is illustrated in Figure 6.5 with respect to the EXCITE simulations (Gaikwad et al., 2023) and the observations (Zhu et al., 2021). The measured photoionization rate of ATON simulations is less than that of EXCITE simulations in the redshift range of $5.0 \leq z \leq 5.5$, hence producing fewer numbers of shorter dark gaps than EXCITE (Refer to Figure 10 in Gaikwad et al. (2023)). This results in ATON producing longer

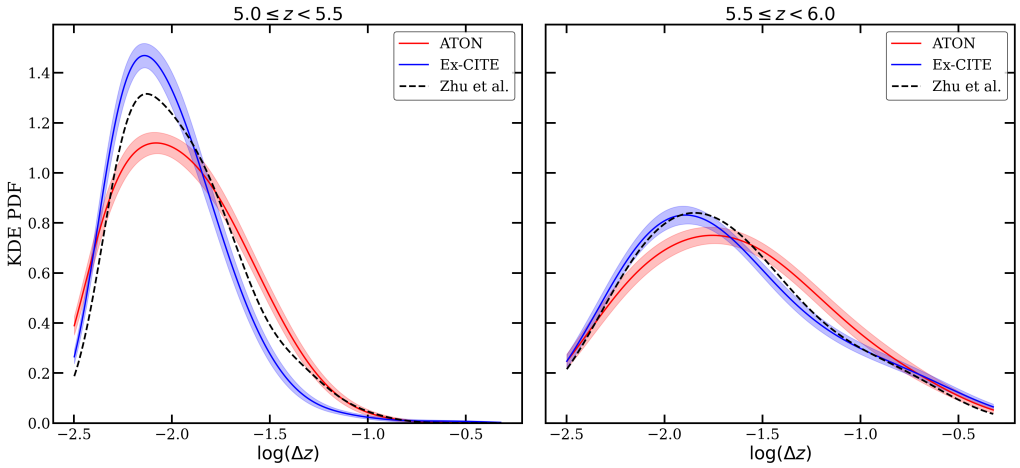


Figure 6.5: Using Dark Gap Statistics on ATON simulations (Kulkarni et al., 2019) to check the consistency of our methods against EXCITE simulations (Gaikwad et al., 2023). The blue curve represents EXCITE simulations and red curve represents the ATON simulations. The PDF behaves as expected due to low photoionization rates measured in ATON compared to EXCITE.

dark gaps compared to EXCITE. In the redshift range of $5.5 \leq z \leq 6.0$, ATON simulations produce photoionization rates similar to those of EXCITE, hence χ by eye gives a much more similar PDF compared to the previous redshift bin.

In Figure 6.3,the scaling values, f_Γ are 0.945, 0.755, 0.560 for Early, Late (fiducial) and Ultra-Late Model respectively. These scaling values are then matched with the neutral fraction presented in simulations (Section 4.2) to give us upper bounds on the neutral fraction over $5.0 \leq z \leq 6.0$: **0.0439,0.0958,0.2040** for Early, Late and Ultra-Late model of reionization respectively.

Table 6.1: Constraints on the neutral hydrogen fraction $\langle x_{\rm HI} \rangle$ and the optimised photoionisation-rate rescaling factor f_Γ for the **Early Reionization** model. The mean IGM photoionisation rate $\langle \Gamma_{12,\rm HI} \rangle$ (in $\rm s^{-1}$) is averaged over all simulation snapshots taken from Gaikwad et al. (2023) measurements. All upper bounds on $\langle x_{\rm HI} \rangle$ are quoted at 1σ. 1σ.

Redshift bin	$z_{\rm mid}$ range	f_Γ [1σ interval]	$\langle x_{\rm HI} \rangle_{\rm upper}$	$\langle \Gamma_{12,\rm HI} \rangle$
Bin 1	5.010–5.344	1.092 [1.072, 1.107]	$< 4.0 \times 10^{-5}$	0.809
Bin 2	5.344–5.565	0.851 [0.836, 0.867]	$< 9.0 \times 10^{-5}$	0.576
Bin 3	5.565–5.972	1.260 [1.220, 1.301]	$< 3.55 \times 10^{-3}$	0.409

Figure 6.4 give the minimum χ^2 curves for redshift bins. We get reasonable χ^2 values for each reionization model where we used hard cut-off of 0.5. The χ^2 curves are asymmetric in nature i.e. they have steep nature on low f_Γ and shallow on high f_Γ. Reducing the scaling parameter below the optimized one leads to over-production of the long dark gaps, hence shifting the PDF to high $log\Delta z$ tails. Increasing the scaling parameter above the optimized one will ionize the IGM more leading to suppression of dark gaps. The PDF loses sensitivity at high mean photoionization rates since the IGM becomes optically thin. We can

Table 6.2: Constraints on the neutral hydrogen fraction $\langle x_{\rm HI}\rangle$ and the optimised photoionisation-rate rescaling factor f_Γ for the **Late (Best-Fit)** reionisation model. The mean IGM photoionisation rate $\langle\Gamma_{12,\rm HI}\rangle$ (in $\rm s^{-1}$) is averaged over all simulation snapshots taken from Gaikwad et al. (2023) measurements. All upper bounds on $\langle x_{\rm HI}\rangle$ are quoted at 1σ.

Redshift bin	$z_{\rm mid}$ range	f_Γ [1σ interval]	$\langle x_{\rm HI}\rangle_{\rm upper}$	$\langle\Gamma_{12,\rm HI}\rangle$
Bin 1	5.010–5.344	0.900 [0.887, 0.908]	$< 1.4 \times 10^{-4}$	0.493
Bin 2	5.344–5.565	0.734 [0.714, 0.751]	$< 2.76 \times 10^{-3}$	0.358
Bin 3	5.565–5.972	1.005 [0.996, 1.024]	$< 4.705 \times 10^{-2}$	0.218

Table 6.3: Constraints on the neutral hydrogen fraction $\langle x_{\rm HI}\rangle$ and the optimised photoionisation-rate rescaling factor f_Γ for the **Ultra-Late Reionization** model. The mean IGM photoionisation rate $\langle\Gamma_{12,\rm HI}\rangle$ (in $\rm s^{-1}$) is averaged over all simulation snapshots taken from Gaikwad et al. (2023) measurements. All upper bounds on $\langle x_{\rm HI}\rangle$ are quoted at 1σ.

Redshift bin	$z_{\rm mid}$ range	f_Γ [1σ interval]	$\langle x_{\rm HI}\rangle_{\rm upper}$	$\langle\Gamma_{12,\rm HI}\rangle$
Bin 1	5.010–5.344	0.780 [0.762, 0.802]	$< 7.9 \times 10^{-4}$	0.305
Bin 2	5.344–5.565	0.537 [0.520, 0.554]	$< 1.423 \times 10^{-2}$	0.230
Bin 3	5.565–5.972	0.805 [0.755, 0.859]	$< 8.398 \times 10^{-2}$	0.121

see the effects of changing the cut off in the Appendix A in much more detail. We obtain the optimized scaling values from the minimum of the χ^2 curves.The optimized scaling values , f_Γ, neutral fraction values ($x_{\rm HI}$) and the mean photoionization rate ($\langle\Gamma_{\rm HI}\rangle$) are listed in the summary table for each reionization model. Table 6.1, 6.2, 6.3 give the summary of this work, particularly the upper bound constraints on the neutral fraction for each reionization history.

All the results are consistent with the physical reionization histories and their features and most importantly, within the bounds of the observational uncertainties and Figure 6.6 depicts the constraints on the neutral fraction from this analysis.

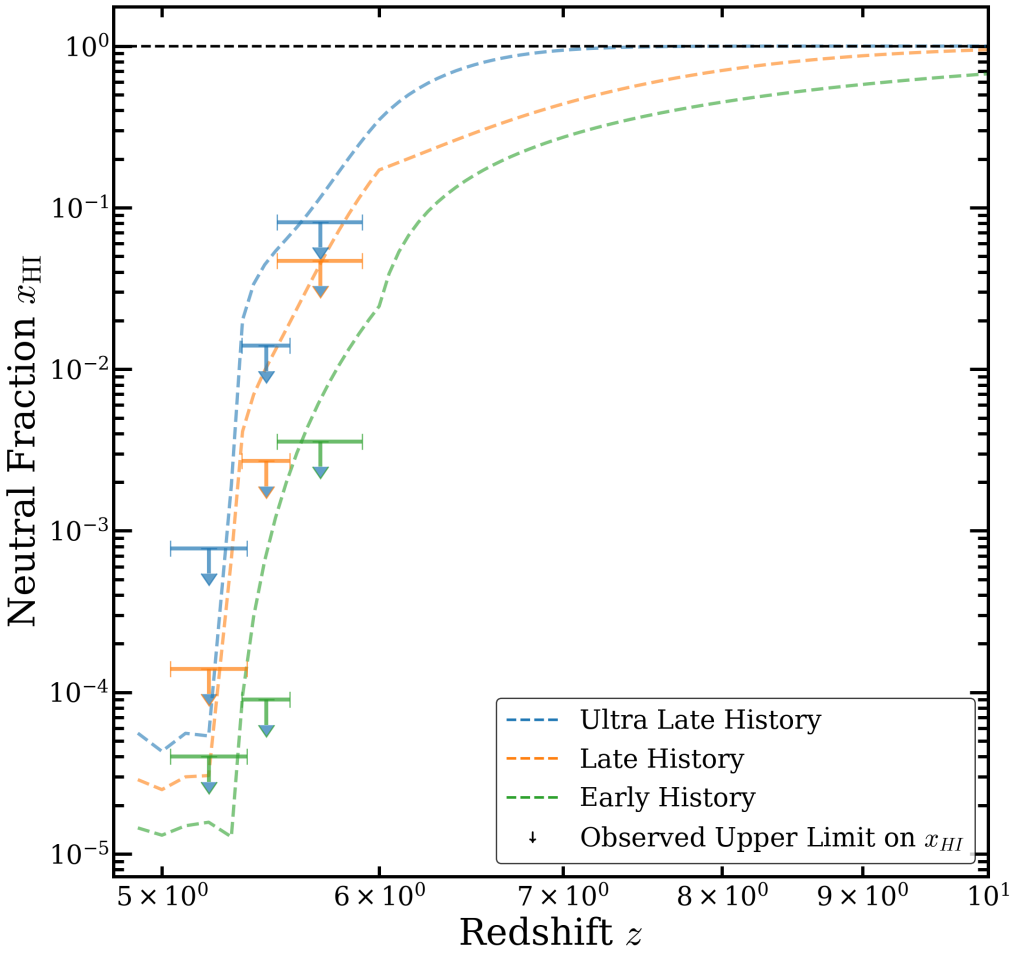


Figure 6.6: Upper Bound Constraints on the neutral fraction evolving with the redshift over different reionization histories.

CHAPTER 7

Conclusion and Discussion

- We have modeled the Lymanα forest by considering the optical depth absorption profiles and then generated Lyman α sightlines of the quasars. The main equation that governs the optical depth generation is Equation 3.2, and we have applied periodic boundary conditions and matched with the Sherwood sightlines for consistency.

- For the generated sightlines, we have implemented automated dark gap detection pipeline which can match the instrumental resolution of XQR-30, and add the gaussian random noise based on quasar SNR (D'Odorico et al., 2023) to generate the quasar sightlines. We have made around 2000 sightlines which consists of 45 mock realizations of the quasars. For all of these sightlines, we have re-binned the spectra to $1\,h^{-1}$cMpc and considered continuous regions having $F_{\rm norm} < 0.05$ with minimum length of $1h^{-1}$Mpc as the dark gap regions (Zhu et al., 2021). See Figure 5.1 for more details.

- We have used the cumulative distribution function (CDF) to get an idea of the preference of the models in different redshift bins. For all the models, the CDF is not able to repoduce the observations. Since the sample size of the sightlines is large, KS test no longer becomes unreliable, so we use D values to get the order of the preference of models: Early Reionization Model for $5.0 \leq z \leq 5.3$ and $5.3 \leq z \leq 5.6$ and Late (Fiducial) Model for $5.6 \leq z \leq 6.0$.

- We have used a global PDF (Figure 5.5) to estimate the mean neutral hydrogen fraction over the redshift bin $5.0 \leq z \leq 6.0$. The global PDF varies strongly with the scaling parameter and we the see the degeneracy in play for all the models.

- For the global PDF, we have used χ^2 minimum curves (Figure 6.3) to get constraints on f_Γ. In f_Γ and we get optimized values at 0.945, 0.755, and 0.560 for Early, Late (Fiducial) and Ultra-Late Model respectively. This gives us the mean global neutral hydrogen fraction ($\bar{x}_{\rm HI}$) constraints as : $0.0439,\ 0.00958,\ 0.2040$.

- We have then used KDE for estimating the upper bounds on the neutral fraction using binning of the redshift (Figure 5.6). The shape of the

KDE changes with the redshift bins revealing important information about reionization physics going on in different redshift bins.

- The KDE for $5.0 \leq z \leq 5.3$ tells us how the dark gaps are concentrated at the small scales depicting that most of the regions in the IGM are ionized. All reionization histories converge at small scales in this redshift bin. We get f_Γ constraints as 1.092, 0.900, 0.780 for Early, Late and Ultra-Late Models respectively.

- The KDE for $5.3 \leq z \leq 5.6$ shows a much flatter and broader PDF, revealing a huge range of dark gaps length present at this epoch. This redshift bin represents the transition epoch and there is a rapid evolution of the neutral fraction, the KDE gives really huge opacity fluctuations (i.e. with different sightlines, the opacity fluctuates greatly), reducing the optimized scaling rate on average. All the reionization models differ greatly from the observation in this redshift bin depicting that it is unable to capture such a rapid evolution of neutral fraction. We get f_Γ values as: 0.851, 0.734, 0.537 for Early, Late and Ultra-Late Models respectively.

- The KDE for $5.6 \leq z \leq 6.0$ gives an extended and smooth reionization picture where most of the dark gap lengths lie at longer scales representing an IGM affected more by the neutral regions rather than the fluctuations. We get f_Γ values as: 1.260, 1.005, 0.805 for Early, Late and Ultra-Late Models respectively.

- We also use F10 Statistics (i.e. giving the spatial distribution of the dark gaps $> 10\text{h}^{-1}\text{Mpc}$) for consistency and find that all the reionization histories are able to produce the observation F10 statistics.

- The upper bounds on neutral fraction are model-dependent: we get different upper bounds for different reionization histories illustrating the degeneracy in EoR. The neutral fraction constraints are given in the Summary Tables 6.1 6.2 6.3 for Early, Late and Ultra Late Models respectively.

- This work displays that there is indeed a strong degeneracy between photoioinization rate and reionization history, which can be removed further by having an anchor of pure observations.

CHAPTER 8

Future Prospects

We have constrained the upper bounds of the neutral fraction using dark gap statistics for the redshift $5.0 \leq z \leq 6.0$. But the constraints are still model dependent compelling us to look for other techniques to constrain the neutral fraction. Primarly, our plan is to explore the Lyman α forest by stacking the quasar spectra to recover damping wings around $z \sim 5.5$. This method have its own advantages: we get nearly model independent constraints (not upper bounds), now observationally confirmed by (Spina et al., 2024), we can look forward to recovering the damping signals by stacking in our EXCITE simulations, giving a much closer constraints on the ionization state of the IGM. We plan to explore other methods like using the quasar proximity regions to get constraints at redshifts like $z \sim 7$. We can also use dark pixel fraction where we quantify all the dark pixels are due to neutral regions giving us conservative nearly model independent constraints on neutral fraction. Using higher-order forests for dark pixel fraction (in Lyβ & Lyγ forest) to get constraints on otherwise unprobed regions due to contamination (like between $6.0 \leq z \leq 7.0$, as shown by (Jin et al., 2023)). We also are planning to constrain the ionization state using Lyman Alpha Emitters and Lyman Break Galaxies i.e by considering their number density evolution to see amount of saturated regions present during EoR and UV luminosity function of the Lyman Break galaxies to give photon budget independently constraining the ionization state of the IGM. We also consider methods like IGM tomography to get constraints on the ionization state where instead of using a single sightline, we use the background sources for reconstruction of 3D maps of absorption in the IGM. We also consider to extend the research to 21 cm forest, Line Intensity mapping of metal lines like OI, CIV etc to directly probe into the ionization state of the IGM. And, above all our priority will be cross-validate or use different probes together to break degeneracies and get coherent picture of reionization. For instance, we can consider the cross-validating probes like dark gap statistics and LAEs: the dark gaps are dominated by the neutral patches and fluctuations, whereas the visibility of Lyα is sensitive to ionized bubbles around galaxies. Both of the results can complement each other.

CHAPTER 9

Appendix A: Effect of the threshold on the binned PDFs

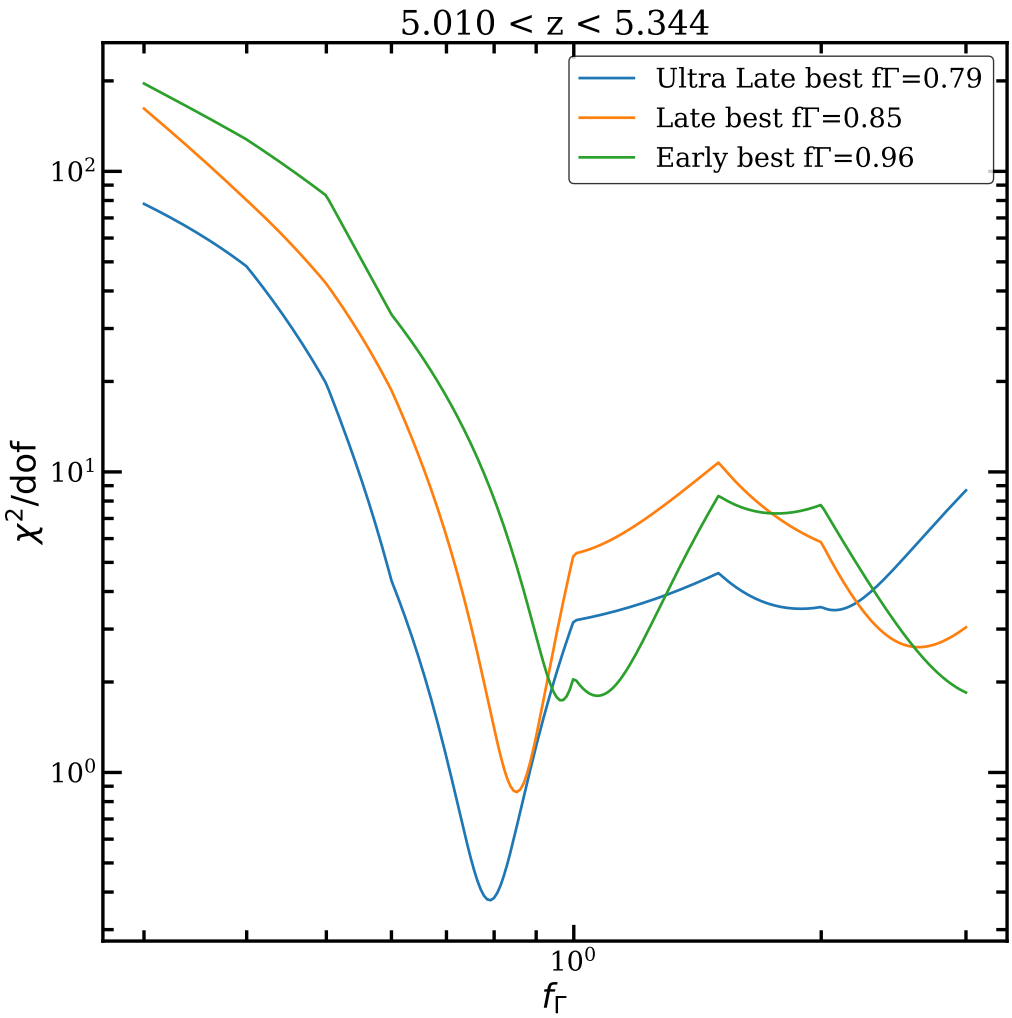


Figure 9.1: χ^2 double minimum (Green curve) for the binned PDF due to domination of Poisson Statistics. The masking of ≤ 0.1 gives such a double minimum , but we consider that changing the threshold still keeps the optimized scaling parameter f_Γ within the uncertainties.

We have put a threshold on the probability distribution function (PDF) as

mentioned in Chapter 5 to avoid the measurement noise on the tails of the distribution: where we have masked regions of $PDF < 0.1$ in global PDF and $PDF < 0.5$ in binned PDFs. The masking of $PDF < 0.5$ is too aggressive but there is a good reason for us to use this masking. When we consider $PDF < 0.1$ as the cutoff, we get double minimum in χ^2 minimum curve for $5.1 \leq z \leq 5.3$ as seen in Figure 9.1. It becomes unclear as to consider which minimum giving us an ambiguity in observation backed f_{Γ}. This prompted us to look for the threshold $PDF < 0.5$ which did remove the double minima but we lose information in doing such a process. This completely disregards of the longer dark gaps and give us a constrain on the intermediate scales. Nevertheless, this is a good deal of trade-off since we consider different minimums and best fit scaling value still stays consistent within the error.

CHAPTER 10

Appendix B: Finding True Mean Free Path

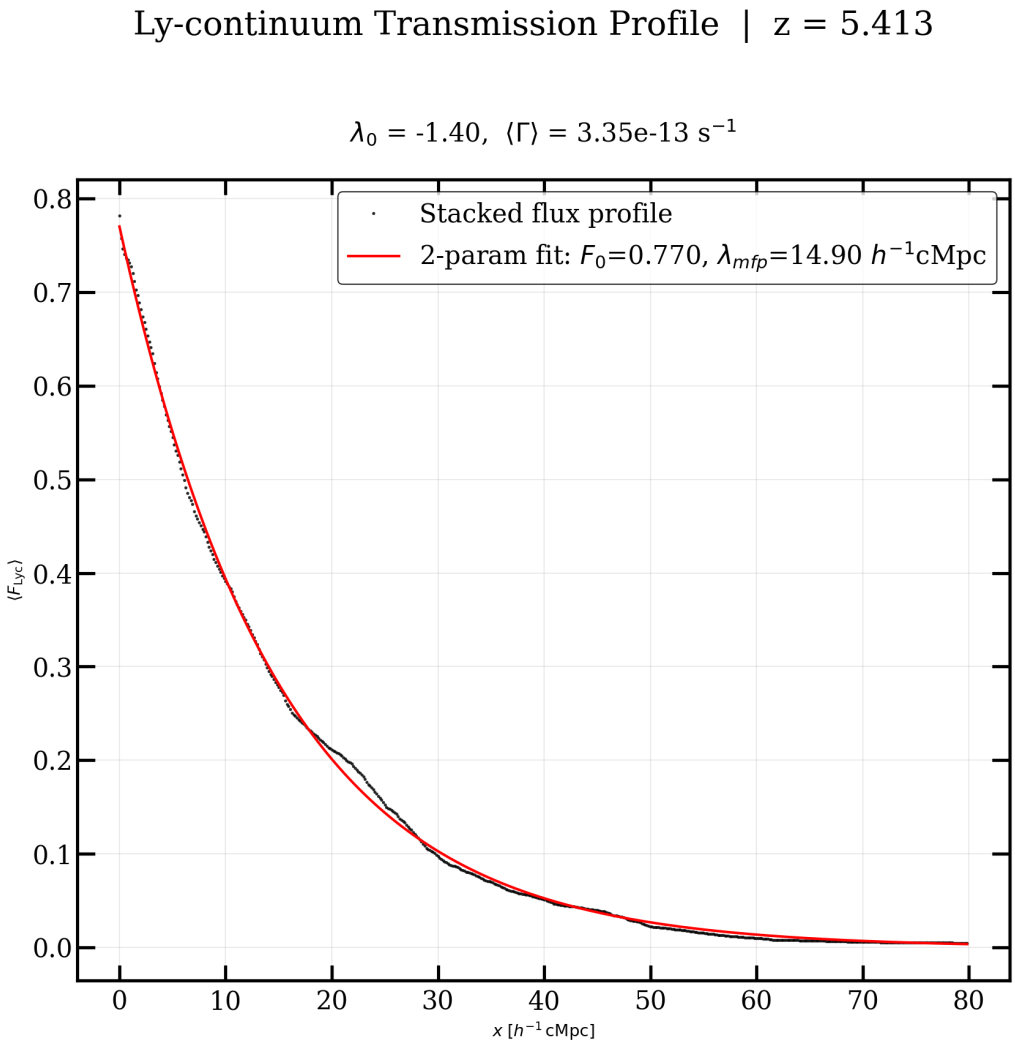


Figure 10.1: Fitting of the stacked flux profile for $\lambda_0 = 1.4$ and a $\langle \Gamma_{\rm HI} \rangle = 3.35 \times 10^{-13}\,{\rm s}^{-1}$.

In the following section, we will consider the method of finding true mean free path from λ_0 which is a convenient parameter used in our simulations. Following is the process that we use to find the true mean free path:

- 1)Shoot a large number of sightlines and find out the mean photoionization rate across all the sightlines.
- 2)Use the mean photoionization rate to find the Lyman continuum optical depth by following formula:

  $\tau_{LC} = \int n_{HI}\sigma_{HI}dx$; where dx is the physical length; $\sigma_{HI} = 6.34e - 18\,\mathrm{cm}^2$ and $n_{HI} = \alpha n_e n_H^2/(\Gamma_{mean})$.
- 3) Find the LyC flux as $\langle F\rangle = \langle e^{(-1.0*\tau_{LC})}\rangle$
- 4) Take the average of the LyC flux along all the sightlines. This will be similar as stacking the profiles at 912 Å.
- 5) Fit the stacked flux with an exponential to obtain the true mean free path as: (see Figure 10.1) $F = F_0 e^{-(dx/\lambda_{\mathrm{mfp}})}$;
- 6) But if flux is non-zero only for pixels< 5: consider using the first value of $F = F_0/e$ point as the true mfp.

Bibliography

Abel, T. and Wandelt, B. D. (2002). Adaptive ray tracing for radiative transfer around point sources. , 330(3):L53–L56.

Alpher, R. A., Bethe, H., and Gamow, G. (1948). The Origin of Chemical Elements. *Physical Review*, 73(7):803–804.

Alvarez, M. A. (2016). The Kinetic Sunyaev-Zel'dovich Effect from Reionization: Simulated Full-sky Maps at Arcminute Resolution. , 824(2):118.

Bañados, E., Venemans, B. P., Mazzucchelli, C., Farina, E. P., Walter, F., Wang, F., Decarli, R., Stern, D., Fan, X., Davies, F. B., Hennawi, J. F., Simcoe, R. A., Turner, M. L., Rix, H.-W., Yang, J., Kelson, D. D., Rudie, G. C., and Winters, J. M. (2018). An 800-million-solar-mass black hole in a significantly neutral Universe at a redshift of 7.5. , 553(7689):473–476.

Barkana, R. and Loeb, A. (2001). In the beginning: the first sources of light and the reionization of the universe. , 349(2):125–238.

Becker, G. D., Bolton, J. S., Haehnelt, M. G., and Sargent, W. L. W. (2011). Detection of extended He II reionization in the temperature evolution of the intergalactic medium. , 410(2):1096–1112.

Becker, G. D., Bolton, J. S., and Lidz, A. (2015). Reionisation and High-Redshift Galaxies: The View from Quasar Absorption Lines. , 32:e045.

Becker, G. D., D'Aloisio, A., Christenson, H. M., Zhu, Y., Worseck, G., and Bolton, J. S. (2021). The mean free path of ionizing photons at $5 < z < 6$: evidence for rapid evolution near reionization. , 508(2):1853–1869.

Bolton, J. S., Haehnelt, M. G., Warren, S. J., Hewett, P. C., Mortlock, D. J., Venemans, B. P., McMahon, R. G., and Simpson, C. (2011). How neutral is the intergalactic medium surrounding the redshift z = 7.085 quasar ULAS J1120+0641? , 416(1):L70–L74.

Bolton, J. S., Puchwein, E., Sijacki, D., Haehnelt, M. G., Kim, T.-S., Meiksin, A., Regan, J. A., and Viel, M. (2017). The Sherwood simulation suite: overview and data comparisons with the Lyman α forest at redshifts $2 \leq z \leq 5$. , 464(1):897–914.

Bresenham, J. E. (1965). Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 4(1):25–30.

Bromm, V. and Larson, R. B. (2004). The First Stars. , 42(1):79–118.

Bromm, V. and Yoshida, N. (2011). The First Galaxies. , 49(1):373–407.

Bull, P., Ferreira, P. G., Patel, P., and Santos, M. G. (2015). Late-time Cosmology with 21 cm Intensity Mapping Experiments. , 803(1):21.

Carilli, C. L., Gnedin, N. Y., and Owen, F. (2002). H I 21 Centimeter Absorption beyond the Epoch of Reionization. , 577(1):22–30.

Carniani, S., Hainline, K., D'Eugenio, F., Eisenstein, D. J., Jakobsen, P., Witstok, J., Johnson, B. D., Chevallard, J., Maiolino, R., Helton, J. M., Willott, C., Robertson, B., Alberts, S., Arribas, S., Baker, W. M., Bhatawdekar, R., Boyett, K., Bunker, A. J., Cameron, A. J., Cargile, P. A., Charlot, S., Curti, M., Curtis-Lake, E., Egami, E., Giardino, G., Isaak, K., Ji, Z., Jones, G. C., Kumari, N., Maseda, M. V., Parlanti, E., Pérez-González, P. G., Rawle, T., Rieke, G., Rieke, M., Del Pino, B. R., Saxena, A., Scholtz, J., Smit, R., Sun, F., Tacchella, S., Übler, H., Venturi, G., Williams, C. C., and Willmer, C. N. A. (2024). Spectroscopic confirmation of two luminous galaxies at a redshift of 14. , 633(8029):318–322.

Chakraborty, A. and Choudhury, T. R. (2026). Reionization and its sources. In *Encyclopedia of Astrophysics, Volume 4*, volume 4, pages 433–452.

Choudhury, T. R. and Ferrara, A. (2006). Physics of Cosmic Reionization. *arXiv e-prints*, pages astro–ph/0603149.

Choudhury, T. R., Srianand, R., and Padmanabhan, T. (2001). Semianalytic Approach to Understanding the Distribution of Neutral Hydrogen in the Universe: Comparison of Simulations with Observations. , 559(1):29–40.

Ciardi, B. and Ferrara, A. (2005). The First Cosmic Structures and Their Effects. , 116(3-4):625–705.

Cyburt, R. H., Fields, B. D., Olive, K. A., and Yeh, T.-H. (2016). Big bang nucleosynthesis: Present status. *Reviews of Modern Physics*, 88(1):015004.

D'Aloisio, A., McQuinn, M., Davies, F. B., and Furlanetto, S. R. (2018). Large fluctuations in the high-redshift metagalactic ionizing background. , 473(1):560–575.

Davies, F. B., Bosman, S. E. I., D'Odorico, V., Campo, S., Mesinger, A., Qin, Y., Becker, G. D., Bañados, E., Chen, H., Cristiani, S., Fan, X., Gallerani, S., Haehnelt, M. G., Keating, L. C., Lai, S., Ryan-Weber, E., Wang, F., Yang, J., and Zhu, Y. (2026). Updated dark pixel fraction constraints on reionization's end from the Lyman-series forests of XQR−30. , 545(2):staf1862.

Davies, F. B. and Furlanetto, S. R. (2016). Large fluctuations in the hydrogen-ionizing background and mean free path following the epoch of reionization. , 460(2):1328–1339.

Davies, F. B., Furlanetto, S. R., and Dixon, K. L. (2017). A self-consistent 3D model of fluctuations in the helium-ionizing background. , 465(3):2886–2894.

Davies, F. B., Hennawi, J. F., Eilers, A.-C., and Lukić, Z. (2018). A New Method to Measure the Post-reionization Ionizing Background from the Joint Distribution of Lyα and Lyβ Forest Transmission. , 855(2):106.

Dayal, P. and Ferrara, A. (2018). Early galaxy formation and its large-scale effects. , 780:1–64.

Dijkstra, M. (2014). Lyα Emitting Galaxies as a Probe of Reionisation. , 31:e040.

D'Odorico, V., Bañados, E., Becker, G. D., Bischetti, M., Bosman, S. E. I., Cupani, G., Davies, R., Farina, E. P., Ferrara, A., Feruglio, C., Mazzucchelli, C., Ryan-Weber, E., Schindler, J.-T., Sodini, A., Venemans, B. P., Walter, F., Chen, H., Lai, S., Zhu, Y., Bian, F., Campo, S., Carniani, S., Cristiani, S., Davies, F., Decarli, R., Drake, A., Eilers, A.-C., Fan, X., Gaikwad, P., Gallerani, S., Greig, B., Haehnelt, M. G., Hennawi, J., Keating, L., Kulkarni, G., Mesinger, A., Meyer, R. A., Neeleman, M., Onoue, M., Pallottini, A., Qin, Y., Rojas-Ruiz, S., Satyavolu, S., Sebastian, A., Tripodi, R., Wang, F., Wolfson, M., Yang, J., and Zanchettin, M. V. (2023). XQR-30: The ultimate XSHOOTER quasar sample at the reionization epoch. , 523(1):1399–1420.

Dolgov, A. D. (2002). Neutrinos in cosmology. , 370(4-5):333–535.

Eilers, A.-C., Davies, F. B., and Hennawi, J. F. (2018). The Opacity of the Intergalactic Medium Measured along Quasar Sightlines at $\tilde{z}$ 6. , 864(1):53.

Fan, X., Strauss, M. A., Becker, R. H., White, R. L., Gunn, J. E., Knapp, G. R., Richards, G. T., Schneider, D. P., Brinkmann, J., and Fukugita, M. (2006). Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with z$\sim$6 Quasars. II. A Sample of 19 Quasars. , 132(1):117–136.

Ferraro, S., Hill, J. C., Battaglia, N., Liu, J., and Spergel, D. N. (2016). Kinematic Sunyaev-Zel'dovich effect with projected fields. II. Prospects, challenges, and comparison with simulations. , 94(12):123526.

Field, G. B. (1959). The Spin Temperature of Intergalactic Neutral Hydrogen. , 129:536.

Fragos, T., Lehmer, B., Tremmel, M., Tzanavaris, P., Basu-Zych, A., Belczynski, K., Hornschemeier, A., Jenkins, L., Kalogera, V., Ptak, A., and Zezas, A. (2013). X-Ray Binary Evolution Across Cosmic Time. , 764(1):41.

Furlanetto, S. R., Oh, S. P., and Briggs, F. H. (2006). Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. , 433(4-6):181–301.

Gaikwad, P., Haehnelt, M. G., Davies, F. B., Bosman, S. E. I., Molaro, M., Kulkarni, G., D'Odorico, V., Becker, G. D., Davies, R. L., Nasir, F., Bolton, J. S., Keating, L. C., Iršič, V., Puchwein, E., Zhu, Y., Asthana, S., Yang, J., Lai, S., and Eilers, A.-C. (2023). Measuring the photoionization rate, neutral fraction, and mean free path of H I ionizing photons at $4.9 < z < 6.0$ from a large sample of XShooter and ESI spectra. , 525(3):4093–4120.

Gaikwad, P., Srianand, R., Khaire, V., and Choudhury, T. R. (2019). Effect of non-equilibrium ionization on derived physical conditions of the high-z intergalactic medium. , 490(2):1588–1604.

George, E. M., Reichardt, C. L., Aird, K. A., Benson, B. A., Bleem, L. E., Carlstrom, J. E., Chang, C. L., Cho, H.-M., Crawford, T. M., Crites, A. T., de Haan, T., Dobbs, M. A., Dudley, J., Halverson, N. W., Harrington, N. L., Holder, G. P., Holzapfel, W. L., Hou, Z., Hrubes, J. D., Keisler, R., Knox, L., Lee, A. T., Leitch, E. M., Lueker, M., Luong-Van, D., McMahon, J. J., Mehl, J., Meyer, S. S., Millea, M., Mocanu, L. M., Mohr, J. J., Montroy, T. E., Padin, S., Plagge, T., Pryke, C., Ruhl, J. E., Schaffer, K. K., Shaw, L., Shirokoff, E., Spieler, H. G., Staniszewski, Z., Stark, A. A., Story, K. T., van Engelen, A., Vanderlinde, K., Vieira, J. D., Williamson, R., and Zahn, O. (2015). A Measurement of Secondary Cosmic Microwave Background Anisotropies from the 2500 Square-degree SPT-SZ Survey. , 799(2):177.

Greig, B., Mesinger, A., and Bañados, E. (2019). Constraints on reionization from the z = 7.5 QSO ULASJ1342+0928. , 484(4):5094–5101.

Greig, B., Mesinger, A., Haiman, Z., and Simcoe, R. A. (2017). Are we witnessing the epoch of reionisation at z = 7.1 from the spectrum of J1120+0641? , 466(4):4239–4249.

Gunn, J. E. and Peterson, B. A. (1965). On the Density of Neutral Hydrogen in Intergalactic Space. , 142:1633–1636.

Guth, A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. , 23(2):347–356.

Haiman, Z., Rees, M. J., and Loeb, A. (1996). H 2 Cooling of Primordial Gas Triggered by UV Irradiation. , 467:522.

Hirata, C. M. (2006). Wouthuysen-Field coupling strength and application to high-redshift 21-cm radiation. , 367(1):259–274.

Hui, L. and Gnedin, N. Y. (1997). Equation of state of the photoionized intergalactic medium. , 292(1):27–42.

Jin, X., Yang, J., Fan, X., Wang, F., Bañados, E., Bian, F., Davies, F. B., Eilers, A.-C., Farina, E. P., Hennawi, J. F., Pacucci, F., Venemans, B., and Walter, F. (2023). (Nearly) Model-independent Constraints on the Neutral Hydrogen Fraction in the Intergalactic Medium at z 5-7 Using Dark Pixel Fractions in Lyα and Lyβ Forests. , 942(2):59.

Jones, A., Noll, S., Kausch, W., Szyszka, C., and Kimeswenger, S. (2013). An advanced scattered moonlight model for Cerro Paranal. , 560:A91.

Kelson, D. D. (2003). Optimal Techniques in Two-dimensional Spectroscopy: Background Subtraction for the 21st Century. , 115(808):688–699.

Kolb, E. W. and Turner, M. S. (1990). *The early universe*, volume 69.

Konno, A., Ouchi, M., Shibuya, T., Ono, Y., Shimasaku, K., Taniguchi, Y., Nagao, T., Kobayashi, M. A. R., Kajisawa, M., Kashikawa, N., Inoue, A. K., Oguri, M., Furusawa, H., Goto, T., Harikane, Y., Higuchi, R., Komiyama, Y., Kusakabe, H., Miyazaki, S., Nakajima, K., and Wang, S.-Y. (2018). SILVERRUSH. IV. Lyα luminosity functions at z = 5.7 and 6.6 studied with $\sim$1300 Lyα emitters on the 14-21 deg^2 sky. , 70:S16.

Koopmans, L., Pritchard, J., Mellema, G., Aguirre, J., Ahn, K., Barkana, R., van Bemmel, I., Bernardi, G., Bonaldi, A., Briggs, F., de Bruyn, A. G., Chang, T. C., Chapman, E., Chen, X., Ciardi, B., Dayal, P., Ferrara, A., Fialkov, A., Fiore, F., Ichiki, K., Illiev, I. T., Inoue, S., Jelic, V., Jones, M., Lazio, J., Maio, U., Majumdar, S., Mack, K. J., Mesinger, A., Morales, M. F., Parsons, A., Pen, U. L., Santos, M., Schneider, R., Semelin, B., de Souza, R. S., Subrahmanyan, R., Takeuchi, T., Vedantham, H., Wagg, J., Webster, R., Wyithe, S., Datta, K. K., and Trott, C. (2015). The Cosmic Dawn and Epoch of Reionisation with SKA. In *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 1.

Kulkarni, G., Keating, L. C., Haehnelt, M. G., Bosman, S. E. I., Puchwein, E., Chardin, J., and Aubert, D. (2019). Large Ly α opacity fluctuations and low CMB τ in models of late reionization with large islands of neutral hydrogen extending to z < 5.5. , 485(1):L24–L28.

Lidz, A., McQuinn, M., Zaldarriaga, M., Hernquist, L., and Dutta, S. (2007). Quasar Proximity Zones and Patchy Reionization. , 670(1):39–59.

Lidz, A., Zahn, O., Furlanetto, S. R., McQuinn, M., Hernquist, L., and Zaldarriaga, M. (2009). Probing Reionization with the 21 cm Galaxy Cross-Power Spectrum. , 690(1):252–266.

Loeb, A. and Barkana, R. (2001). The Reionization of the Universe by the First Stars and Quasars. , 39:19–66.

Loeb, A. and Furlanetto, S. R. (2013). *The First Galaxies in the Universe*.

Madau, P. and Dickinson, M. (2014). Cosmic Star-Formation History. , 52:415–486.

Madau, P. and Haardt, F. (2015). Cosmic Reionization after Planck: Could Quasars Do It All? , 813(1):L8.

Madau, P., Haardt, F., and Rees, M. J. (1999). Radiative Transfer in a Clumpy Universe. III. The Nature of Cosmological Ionizing Sources. , 514(2):648–659.

Madau, P., Meiksin, A., and Rees, M. J. (1997). 21 Centimeter Tomography of the Intergalactic Medium at High Redshift. , 475(2):429–444.

Maitra, S., Kulkarni, G., Asthana, S., Bolton, J. S., Haehnelt, M. G., and Keating, L. (2025). The Lyman α emitter bispectrum as a probe of reionization morphology. , 542(1):486–507.

Maity, B., Davies, F. B., Spina, B., and Bosman, S. E. I. (2026). Studying dark gaps in Ly- forest transmission with large reionization simulation. *arXiv e-prints*, page arXiv:2602.00301.

Majumdar, S., Pritchard, J. R., Mondal, R., Watkinson, C. A., Bharadwaj, S., and Mellema, G. (2018). Quantifying the non-Gaussianity in the EoR 21-cm signal through bispectrum. , 476(3):4007–4024.

Malloy, M. and Lidz, A. (2015). How to Search for Islands of Neutral Hydrogen in the z $\sim$5.5 IGM. , 799(2):179.

Mason, C. A., Treu, T., Dijkstra, M., Mesinger, A., Trenti, M., Pentericci, L., de Barros, S., and Vanzella, E. (2018). The Universe Is Reionizing at z $\sim$ 7: Bayesian Inference of the IGM Neutral Fraction Using Lyα Emission from Galaxies. , 856(1):2.

McGreer, I. D., Mesinger, A., and Fan, X. (2011). The first (nearly) model-independent constraint on the neutral hydrogen fraction at z $\sim$ 6. , 415(4):3237–3246.

McQuinn, M. (2016). The Evolution of the Intergalactic Medium. , 54:313–362.

McQuinn, M., Furlanetto, S. R., Hernquist, L., Zahn, O., and Zaldarriaga, M. (2005). The Kinetic Sunyaev-Zel'dovich Effect from Reionization. , 630(2):643–656.

Meiksin, A. (2020). The influence of metagalactic ultraviolet background fluctuations on the high-redshift Lyα forest. , 491(4):4884–4893.

Meiksin, A. A. (2009). The physics of the intergalactic medium. *Reviews of Modern Physics*, 81(4):1405–1469.

Mesinger, A., Furlanetto, S., and Cen, R. (2011). 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal. , 411(2):955–972.

Miralda-Escudé, J. (1998). Reionization of the Intergalactic Medium and the Damping Wing of the Gunn-Peterson Trough. , 501(1):15–22.

Miralda-Escudé, J., Haehnelt, M., and Rees, M. J. (2000). Reionization of the Inhomogeneous Universe. , 530(1):1–16.

Mo, H., van den Bosch, F. C., and White, S. (2010). *Galaxy Formation and Evolution*.

Mueller, E.-M., de Bernardis, F., Bean, R., and Niemack, M. D. (2015). Constraints on Gravity and Dark Energy from the Pairwise Kinematic Sunyaev-Zel'dovich Effect. , 808(1):47.

Mukhanov, V. F. and Chibisov, G. V. (1981). Quantum fluctuations and a non-singular universe. *Soviet Journal of Experimental and Theoretical Physics Letters*, 33:532.

Noll, S., Kausch, W., Barden, M., Jones, A. M., Szyszka, C., Kimeswenger, S., and Vinther, J. (2012). An atmospheric radiation model for Cerro Paranal. I. The optical spectral range. , 543:A92.

Oesch, P. A., Brammer, G., van Dokkum, P. G., Illingworth, G. D., Bouwens, R. J., Labbé, I., Franx, M., Momcheva, I., Ashby, M. L. N., Fazio, G. G., Gonzalez, V., Holden, B., Magee, D., Skelton, R. E., Smit, R., Spitler, L. R., Trenti, M., and Willner, S. P. (2016). A Remarkably Luminous Galaxy at z=11.1 Measured with Hubble Space Telescope Grism Spectroscopy. , 819(2):129.

Ouchi, M., Shimasaku, K., Furusawa, H., Saito, T., Yoshida, M., Akiyama, M., Ono, Y., Yamada, T., Ota, K., Kashikawa, N., Iye, M., Kodama, T., Okamura, S., Simpson, C., and Yoshida, M. (2010). Statistics of 207 Lyα Emitters at a Redshift Near 7: Constraints on Reionization and Galaxy Formation Models. , 723(1):869–894.

Peebles, P. J. E. (1993). *Principles of Physical Cosmology*.

Peeples, M. S., Werk, J. K., Tumlinson, J., Oppenheimer, B. D., Prochaska, J. X., Katz, N., and Weinberg, D. H. (2014). A Budget and Accounting of Metals at z $\sim$0: Results from the COS-Halos Survey. , 786(1):54.

Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N., Basak, S., Battye, R., Benabed, K., Bernard, J.-P., Bersanelli, M., Bielewicz, P., Bock, J. J., Bond, J. R., Borrill, J., Bouchet, F. R., Boulanger, F., Bucher, M., Burigana, C., Butler, R. C., Calabrese, E., Cardoso, J.-F., Carron, J., Challinor, A., Chiang, H. C., Chluba, J., Colombo, L. P. L., Combet, C., Contreras, D., Crill, B. P., Cuttaia, F., de Bernardis, P., de Zotti, G., Delabrouille, J., Delouis, J.-M., Di Valentino, E., Diego, J. M., Doré, O., Douspis, M., Ducout, A., Dupac, X., Dusini, S., Efstathiou, G., Elsner, F., Enßlin, T. A., Eriksen, H. K., Fantaye, Y., Farhang, M., Fergusson, J., Fernandez-Cobos, R., Finelli, F., Forastieri, F., Frailis, M., Fraisse, A. A., Franceschi, E., Frolov, A., Galeotta, S., Galli, S., Ganga, K., Génova-Santos, R. T., Gerbino, M., Ghosh, T., González-Nuevo, J., Górski, K. M., Gratton, S., Gruppuso, A., Gudmundsson, J. E., Hamann, J., Handley, W., Hansen, F. K., Herranz, D., Hildebrandt, S. R., Hivon, E., Huang, Z., Jaffe, A. H., Jones, W. C., Karakci, A., Keihänen, E., Keskitalo, R., Kiiveri, K., Kim, J., Kisner, T. S., Knox, L., Krachmalnicoff, N., Kunz, M., Kurki-Suonio, H., Lagache, G., Lamarre, J.-M., Lasenby, A., Lattanzi, M., Lawrence, C. R., Le Jeune, M., Lemos, P., Lesgourgues, J., Levrier, F., Lewis, A., Liguori, M., Lilje, P. B., Lilley, M., Lindholm, V., López-Caniego, M., Lubin, P. M., Ma, Y.-Z., Macías-Pérez, J. F., Maggio, G., Maino, D., Mandolesi, N., Mangilli, A., Marcos-Caballero, A., Maris, M., Martin, P. G., Martinelli, M., Martínez-González, E., Matarrese, S., Mauri, N., McEwen, J. D., Meinhold, P. R., Melchiorri, A., Mennella,

A., Migliaccio, M., Millea, M., Mitra, S., Miville-Deschênes, M.-A., Molinari, D., Montier, L., Morgante, G., Moss, A., Natoli, P., Nørgaard-Nielsen, H. U., Pagano, L., Paoletti, D., Partridge, B., Patanchon, G., Peiris, H. V., Perrotta, F., Pettorino, V., Piacentini, F., Polastri, L., Polenta, G., Puget, J.-L., Rachen, J. P., Reinecke, M., Remazeilles, M., Renzi, A., Rocha, G., Rosset, C., Roudier, G., Rubiño-Martín, J. A., Ruiz-Granados, B., Salvati, L., Sandri, M., Savelainen, M., Scott, D., Shellard, E. P. S., Sirignano, C., Sirri, G., Spencer, L. D., Sunyaev, R., Suur-Uski, A.-S., Tauber, J. A., Tavagnacco, D., Tenti, M., Toffolatti, L., Tomasi, M., Trombetti, T., Valenziano, L., Valiviita, J., Van Tent, B., Vibert, L., Vielva, P., Villa, F., Vittorio, N., Wandelt, B. D., Wehus, I. K., White, M., White, S. D. M., Zacchei, A., and Zonca, A. (2020). Planck 2018 results. VI. Cosmological parameters. , 641:A6.

Press, W. H. and Schechter, P. (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. , 187:425–438.

Pritchard, J. R. and Loeb, A. (2012). 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901.

Raghunathan, S., Ade, P. A. R., Anderson, A. J., Ansarinejad, B., Archipley, M., Austermann, J. E., Balkenhol, L., Beall, J. A., Benabed, K., Bender, A. N., Benson, B. A., Bianchini, F., Bleem, L. E., Bock, J., Bouchet, F. R., Bryant, L., Camphuis, E., Carlstrom, J. E., Cecil, T. W., Chang, C. L., Chaubal, P., Chiang, H. C., Chichura, P. M., Chou, T.-L., Citron, R., Coerver, A., Crawford, T. M., Crites, A. T., Cukierman, A., Daley, C., Dibert, K. R., Dobbs, M. A., Doussot, A., Dutcher, D., Everett, W., Feng, C., Ferguson, K. R., Fichman, K., Foster, A., Galli, S., Gallicchio, J., Gambrel, A. E., Gardner, R. W., Ge, F., George, E. M., Goeckner-Wald, N., Gualtieri, R., Guidi, F., Guns, S., Gupta, N., de Haan, T., Halverson, N. W., Hivon, E., Holder, G. P., Holzapfel, W. L., Hood, J. C., Hrubes, J. D., Hryciuk, A., Huang, N., Hubmayr, J., Irwin, K. D., Kéruzoré, F., Khalife, A. R., Knox, L., Korman, M., Kornoelje, K., Kuo, C.-L., Lee, A. T., Levy, K., Li, D., Lowitz, A. E., Lu, C., Maniyar, A., Martsen, E. S., McMahon, J. J., Menanteau, F., Millea, M., Montgomery, J., Corbett Moran, C., Nakato, Y., Natoli, T., Nibarger, J. P., Noble, G. I., Novosad, V., Omori, Y., Padin, S., Pan, Z., Paschos, P., Patil, S., Phadke, K. A., Prabhu, K., Pryke, C., Quan, W., Rahimi, M., Rahlin, A., Reichardt, C. L., Rouble, M., Ruhl, J. E., Saliwanchik, B. R., Schaffer, K. K., Schiappucci, E., Sievers, C., Smecher, G., Sobrin, J. A., Stark, A. A., Stephen, J., Suzuki, A., Tandoi, C., Thompson, K. L., Thorne, B., Trendafilova, C., Tucker, C., Umilta, C., Veach, T., Vieira, J. D., Viero, M. P., Wan, Y., Wang, G., Whitehorn, N., Wu, W. L. K., Yefremenko, V., Young, M. R., Zebrowski, J. A., Zemcov, M., SPT-3G, and SPTpol Collaboration (2024). First Constraints on the Epoch of Reionization Using the Non-Gaussianity of the Kinematic Sunyaev-Zel'dovich Effect from the South Pole Telescope and Herschel-SPIRE Observations. , 133(12):121004.

Rahmati, A., Schaye, J., Crain, R. A., Oppenheimer, B. D., Schaller, M., and Theuns, T. (2016). Cosmic distribution of highly ionized metals and their physical conditions in the EAGLE simulations. , 459(1):310–332.

Roberts-Borsani, G. W., Bouwens, R. J., Oesch, P. A., Labbe, I., Smit, R., Illingworth, G. D., van Dokkum, P., Holden, B., Gonzalez, V., Stefanon, M., Holwerda, B., and Wilkins, S. (2016). z $\gtrsim$ 7 Galaxies with Red Spitzer/IRAC [3.6]-[4.5] Colors in the Full CANDELS Data Set: The Brightest-Known Galaxies at z $\sim$7-9 and a Probable Spectroscopic Confirmation at z = 7.48. , 823(2):143.

Robertson, B. E. (2022). Galaxy Formation and Reionization: Key Unknowns and Expected Breakthroughs by the James Webb Space Telescope. , 60:121–158.

Robertson, B. E., Ellis, R. S., Furlanetto, S. R., and Dunlop, J. S. (2015). Cosmic Reionization and Early Star-forming Galaxies: A Joint Analysis of New Constraints from Planck and the Hubble Space Telescope. , 802(2):L19.

Rorai, A., Hennawi, J. F., Oñorbe, J., White, M., Prochaska, J. X., Kulkarni, G., Walther, M., Lukić, Z., and Lee, K.-G. (2017). Measurement of the small-scale structure of the intergalactic medium using close quasar pairs. *Science*, 356(6336):418–422.

Santos, M. R. (2004). Probing reionization with Lyman α emission lines. , 349(3):1137–1152.

Schaye, J., Theuns, T., Leonard, A., and Efstathiou, G. (1999). Measuring the equation of state of the intergalactic medium. , 310(1):57–70.

Seager, S., Sasselov, D. D., and Scott, D. (1999). A New Calculation of the Recombination Epoch. , 523(1):L1–L5.

Seager, S., Sasselov, D. D., and Scott, D. (2000). How Exactly Did the Universe Become Neutral? , 128(2):407–430.

Shull, J. M., Harness, A., Trenti, M., and Smith, B. D. (2012). Critical Star Formation Rates for Reionization: Full Reionization Occurs at Redshift z $\approx$ 7. , 747(2):100.

Silva, M. B., Santos, M. G., Gong, Y., Cooray, A., and Bock, J. (2013). Intensity Mapping of Lyα Emission during the Epoch of Reionization. , 763(2):132.

Songaila, A. (1998). The Redshift Evolution of the Metagalactic Ionizing Flux Inferred from Metal Line Ratios in the Lyman Forest. , 115(6):2184–2205.

Spina, B., Bosman, S. E. I., Davies, F. B., Gaikwad, P., and Zhu, Y. (2024). Damping wings in the Lyman-α forest: A model-independent measurement of the neutral fraction at 5.4 < z < 6.1. , 688:L26.

Springel, V. (2005). The cosmological simulation code GADGET-2. , 364(4):1105–1134.

Sunyaev, R. A. and Zeldovich, Y. B. (1980). The velocity of clusters of galaxies relative to the microwave background - The possibility of its measurement. , 190:413–420.

Tepper-García, T. (2006). Voigt profile fitting to quasar absorption lines: an analytic approximation to the Voigt-Hjerting function. , 369(4):2025–2035.

Vernet, J., Dekker, H., D'Odorico, S., Kaper, L., Kjaergaard, P., Hammer, F., Randich, S., Zerbi, F., Groot, P. J., Hjorth, J., Guinouard, I., Navarro, R., Adolfse, T., Albers, P. W., Amans, J.-P., Andersen, J. J., Andersen, M. I., Binetruy, P., Bristow, P., Castillo, R., Chemla, F., Christensen, L., Conconi, P., Conzelmann, R., Dam, J., de Caprio, V., de Ugarte Postigo, A., Delabre, B., di Marcantonio, P., Downing, M., Elswijk, E., Finger, G., Fischer, G., Flores, H., François, P., Goldoni, P., Guglielmi, L., Haigron, R., Hanenburg, H., Hendriks, I., Horrobin, M., Horville, D., Jessen, N. C., Kerber, F., Kern, L., Kiekebusch, M., Kleszcz, P., Klougart, J., Kragt, J., Larsen, H. H., Lizon, J.-L., Lucuix, C., Mainieri, V., Manuputy, R., Martayan, C., Mason, E., Mazzoleni, R., Michaelsen, N., Modigliani, A., Moehler, S., Møller, P., Norup Sørensen, A., Nørregaard, P., Péroux, C., Patat, F., Pena, E., Pragt, J., Reinero, C., Rigal, F., Riva, M., Roelfsema, R., Royer, F., Sacco, G., Santin, P., Schoenmaker, T., Spano, P., Sweers, E., Ter Horst, R., Tintori, M., Tromp, N., van Dael, P., van der Vliet, H., Venema, L., Vidali, M., Vinther, J., Vola, P., Winters, R., Wistisen, D., Wulterkens, G., and Zacchei, A. (2011). X-shooter, the new wide band intermediate resolution spectrograph at the ESO Very Large Telescope. , 536:A105.

Voigt, W. (1912). *Über das Gesetz der Intensitätsverteilung innerhalb der Linien eines Gasspektrums*, volume 1912,25 of *Sitzungsberichte*. München.

Walther, M., Oñorbe, J., Hennawi, J. F., and Lukić, Z. (2019). New Constraints on IGM Thermal Evolution from the Lyα Forest Power Spectrum. , 872(1):13.

Wang, F., Davies, F. B., Yang, J., Hennawi, J. F., Fan, X., Barth, A. J., Jiang, L., Wu, X.-B., Mudd, D. M., Bañados, E., Bian, F., Decarli, R., Eilers, A.-C., Farina, E. P., Venemans, B., Walter, F., and Yue, M. (2020). A Significantly Neutral Intergalactic Medium Around the Luminous z = 7 Quasar J0252-0503. , 896(1):23.

Weinberg, S. (2008). *Cosmology*.

Yang, J., Wang, F., Fan, X., Hennawi, J. F., Davies, F. B., Yue, M., Banados, E., Wu, X.-B., Venemans, B., Barth, A. J., Bian, F., Boutsia, K., Decarli, R., Farina, E. P., Green, R., Jiang, L., Li, J.-T., Mazzucchelli, C., and Walter, F. (2020). Pōniuā'ena: A Luminous z = 7.5 Quasar Hosting a 1.5 Billion Solar Mass Black Hole. , 897(1):L14.

Zel'Dovich, Y. B. (1970). Reprint of 1970A&A.....5...84Z. Gravitational instability: an approximate theory for large density perturbations. , 500:13–18.

Zhu, Y., Becker, G. D., Bosman, S. E. I., Cain, C., Keating, L. C., Nasir, F., D'Odorico, V., Bañados, E., Bian, F., Bischetti, M., Bolton, J. S., Chen, H., D'Aloisio, A., Davies, F. B., Davies, R. L., Eilers, A.-C., Fan, X., Gaikwad, P., Greig, B., Haehnelt, M. G., Kulkarni, G., Lai, S., Puchwein, E., Qin, Y., Ryan-Weber, E. V., Satyavolu, S., Spina, B., Walter, F., Wang, F., Wolfson,

M., and Yang, J. (2024). Damping wing-like features in the stacked Ly α forest: Potential neutral hydrogen islands at z < 6. , 533(1):L49–L56.

Zhu, Y., Becker, G. D., Bosman, S. E. I., Keating, L. C., Christenson, H. M., Bañados, E., Bian, F., Davies, F. B., D'Odorico, V., Eilers, A.-C., Fan, X., Haehnelt, M. G., Kulkarni, G., Pallottini, A., Qin, Y., Wang, F., and Yang, J. (2021). Chasing the Tail of Cosmic Reionization with Dark Gap Statistics in the Lyα Forest over 5 < z < 6. , 923(2):223.

Zitrin, A., Labbé, I., Belli, S., Bouwens, R., Ellis, R. S., Roberts-Borsani, G., Stark, D. P., Oesch, P. A., and Smit, R. (2015). Lymanα Emission from a Luminous z = 8.68 Galaxy: Implications for Galaxies as Tracers of Cosmic Reionization. , 810(1):L12.