Dissertation

submitted to the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of Heidelberg University, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Yash Mohan Sharma

born in New Delhi, India

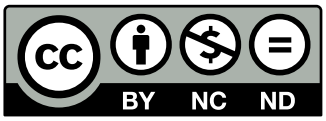
Oral examination: July 17th, 2025

Decoding the Reionization-Epoch Intergalactic Medium with the Most Distant Quasars

Referees:

Dr. Fabian Walter

Prof. Dr. Ralf Klessen



Abstract

At high-redshifts $z \geq 6$, the neutral hydrogen column densities in the IGM are so large that almost all of the Lyα photons around the resonance are absorbed. At those redshifts, the Lyα damping wing signals have proven to be instrumental in studying the epoch of reionization. With the recent advancements in the discovery and measurement of high-redshift sources (galaxies and quasars), it is becoming ever more crucial to explore what this new set of sources might reveal regarding the reionization history and its topology. Hence, in this thesis, we aim to comprehensively study signatures of astrophysical, IGM, and reionization source parameters on the reionization topology as seen through the spectra of an ensemble of damping wing profiles. For this, in chapter one, using 21cmFAST we generated the reionization models subjected to a large set of astrophysical parameters. We found that the neutral fraction, x_{HI}, quasar lifetime, t_q, quasar host halo mass, M_{qso}, and minimum halo mass that can support star formation, M_{min}, significantly impact both the median signal and the scatter of the ensemble of damping wings. But just the idea of an ensemble is not sufficient. We need to quantize it and confirm if such an ensemble is observationally possible. That's why, in the second chapter, we studied the constraining power of damping wings over the parameters filtered in chapter one. We showed that the constraints provided by only 64 quasars at redshift 7, $x_{HI} = 0.5^{+0.02}_{-0.02}$, $M_{min} = 8.78^{+0.53}_{-0.53}$, $\log t_q/\mathrm{yr} = 6.0^{+0.12}_{-0.12}$, and $\log M_{qso}/M_\odot = 11.52^{+0.32}_{-0.31}$ are comparable to the results from other observables like 21cm signal, and our methodology works for as low as 32 quasars. Even though this proved the strength of our damping wing analysis, it did not show that of our suite of models. Hence, using the similar reionization models developed in chapters one and two, in chapter three, instead of calculating Lyα optical depth, we looked at the Lyman continuum optical depth at $z = 6$. Subsequently, we studied the ionizing photons' mean free path (MFP) dependency on the filtered set of parameters from chapter one. We then plotted our models of transmission flux in the Lyman continuum regime against the stacked quasar spectra to estimate the range of these parameters that best fit the spectra. Our results implied that to explain the short MFP from the late reionization models, we would need $x_{HI} \geq 0.35$ and a very short t_q ($= 10^4$yrs) or $x_{HI} \approx 0.5$ for $t_q = 10^6$yrs. Thus, indicating that the late reionization models are insufficient to explain the short MFP of ionizing photons, we need some additional Lyman limit systems within the quasar proximity zone to explain the short MFP of the ionizing photons at $z = 6$.

Zusammenfassung

Bei hohen Rotverschiebungen $z \geq 6$, wo die Säulendichten des neutralen Wasserstoffs so groß sind, dass fast alle Lyα-Photonen um die Resonanz herum absorbiert werden, haben sich die Dämpfungsflügel-Signaturen als entscheidend für die Untersuchung der Reionisierungsära erwiesen. Mit den jüngsten Fortschritten bei der Entdeckung und Messung von Quellen mit hoher Rotverschiebung (Galaxien und Quasare) wird es immer wichtiger zu untersuchen, was diese neuen Quellen über die Geschichte der Reionisierung und ihre Topologie verraten könnten. Daher ist es das Ziel dieser Arbeit, Signaturen von astrophysikalischen, IGM- und Reionisierungsquellenparametern auf die Reionisierungstopologie, wie sie durch die Spektren eines Ensembles von Damping-Wing-Profilen sichtbar wird, umfassend zu untersuchen. Zu diesem Zweck haben wir in Kapitel 1 mit 21cmFAST Reionisierungsmodelle unter Verwendung einer großen Anzahl astrophysikalischer Parameter erstellt. Wir haben festgestellt, dass der Neutralanteil x_{HI}, die Lebensdauer von Quasaren t_q, die Masse des Quasar-Wirtshalos M_{qso} und die minimale Halomasse, die Sternentstehung unterstützen kann, M_{min}, sowohl das mittlere Signal als auch die Streuung des Ensembles von Damping Wings signifikant beeinflussen. Aber die Idee eines Ensembles allein reicht nicht aus. Wir müssen sie quantisieren und bestätigen, ob ein solches Ensemble beobachtbar ist. Aus diesem Grund haben wir im zweiten Kapitel die Einschränkungsfähigkeit von Dämpfungsflügeln auf die im ersten Kapitel gefilterten Parameter untersucht. Wir haben gezeigt, dass die Einschränkungen, die nur durch 64 Quasare bei Rotverschiebung 7, $x_{HI} = 0,5^{+0,02}_{-0,02}$, $M_{min} = 8,78^{+0,53}_{-0,53}$, $\log t_q/\mathrm{yr} = 6,0^{+0,12}_{-0,12}$ und $\log M_{qso}/M_\odot = 11,52^{+0,32}_{-0,31}$ sind vergleichbar mit den Ergebnissen anderer Beobachtungsgrößen wie dem 21-cm-Signal, und unsere Methodik funktioniert bereits für 32 Quasare. Auch wenn dies die Stärke unserer Dämpfungsflügelanalyse bewiesen hat, hat es nicht die Stärke unserer Modellreihe gezeigt. Daher haben wir in Kapitel 3 anstelle der Berechnung der optischen Tiefe Lyα die optische Tiefe des Lyman-Kontinuums bei $z = 6$ untersucht, wobei wir die in den Kapiteln 1 und 2 entwickelten Reionisationsmodelle verwendet haben. Anschließend haben wir die Abhängigkeit der mittleren freien Weglänge (MFW) der ionisierenden Photonen von den gefilterten Parametern aus Kapitel 1 untersucht. Anschließend haben wir unsere Modelle des Transmissionsflusses im Lyman-Kontinuum-Bereich gegen die gestapelten Quasarspektren aufgetragen, um den Bereich dieser Parameter zu schätzen, der am besten zu den Spektren passt. Unsere Ergebnisse deuten darauf hin, dass zur Erklärung der kurzen MFW aus den späten Reionisierungsmodellen $x_{HI} \geq 0,35$ und eine sehr kurze t_q ($= 10^4$Jahre) oder $x_{HI} \approx 0,5$ für $t_q = 10^6$Jahre erforderlich wären. Dies deutet darauf hin, dass die späten Reionisierungsmodelle nicht ausreichen, um die kurze MFW ionisierender Photonen zu erklären. Wir benötigen daher einige zusätzliche

Lyman-Limit-Systeme in der Quasar-Nähezone, um die kurze MFW der ionisierenden Photonen bei $z = 6$ zu erklären.

Acknowledgements

The pathway towards completing any Ph.D. is never smooth. Neither was this, but it was a joyful experience because of the support from many people who helped me push at various stages.

At first, I would like to thank Fred Davies and MPIA for offering me the opportunity to work on such an exciting topic and for opening the doors of reionization physics for me. I am grateful to Fred, who, despite my diverse background, trusted me and guided me through my Ph.D. journey. I am also thankful to our reionization group, which, other than Fred, includes Fahad Nasir, Prakash Gaikwad, Barun Maity, Sarah Bosman, and her group, for all our discussions and group meetings. Of course, getting to MPIA wouldn't be possible if it wasn't for Girish Kulkarni from the Tata Institute of Fundamental Research (TIFR), Mumbai. I cannot express my gratitude for Girish's support and guidance with words alone. Without him, I wouldn't be studying Cosmology. Besides scientific guidance, I also appreciate his advice for my personal growth. Similarly, I would not have met Girish if it wasn't for Savinder Kaur and Sushil Kumar from Sri Guru Teg Bahadur (SGTB) Khalsa College, Delhi University, who guided and shaped my interest in Physics during my undergrad. I thank Fabian Walter and Ralf Klessen for guiding my thesis progress, providing fruitful comments, and agreeing to review my thesis. I would also like to thank Christian Fendt, who was always present during the tough times of my Ph.D.

This Ph.D. wouldn't be as joyful as it was without the support of all my friends from MPIA and outside of MPIA, mainly thanks to Tanja Heise, Akash Sampathkumar, Prateek Alakh, and Ashish Maithani for all their support during difficult times. Thanks to all my friends, I could appreciate every moment I spent in Heidelberg and never felt distant from home. Nonetheless, the support from my home and family can never be understated, no matter how far away I am. I cannot thank my parents and siblings enough, who ensured I could follow my dreams. I am deeply grateful to my cousin Devashish Binjola, who was my inspiration and closest friend during my childhood. I wouldn't enjoy science as much as I do if it weren't for his shared knowledge. I am grateful to all my cousins and family members for their support and trust in me. Lastly, I thank Bhavesh Rajpoot, who helped me immensely during my thesis writing phase. If a thank you could suffice my appreciation towards all these people, I am really thankful to them.

Contents

List of Figures

List of Tables

Acronyms

Acronym	Description
AGN	Active Galactic Nuclei
BAO	Baryon Acoustic Oscillations
BBN	Big Bang Nucleosynthesis
ΛCDM	Λ-Cold Dark Matter
CMB	Cosmic Microwave Background
DM	Dark Matter
EoR	Epoch of Reionization
ES	Excursion-Set
FIM	Fisher Information Matrix
FLRW	Fiedmann-Lemaitre-Robertson-Walker
HI	neutral hydrogen
HMF	Halo Mass Function
IGM	Intergalactic Medium
IMF	Initial Mass Function
ISM	Interstellar Medium
LF	Luminosity Function
LLS	Lyman Limit Systems
LPT	Linear Perturbation Theory
LyC	Lyman Continuum
LyS	Lyman Series
MFP	Mean Free Path
MPIA	Max Planck Institute for Astronomy
PS	Press-Schechter
S/N	Signal-to-Noise Ratio
SED	Spectral Energy Distribution
SFR	star formation rate
SMBH	Super Massive Black Holes
ST	Sheth-Tormen

Acronym	**Description**
UVB	ultraviolet background radiation

1 Theoretical Framework

Contents

The bright stars twinkling in the night sky, guiding people in the earliest eras of human voyages, were not only followed by sails and ships but also by the human mind and curiosity. The ever-so-lasting fascination with the motion of stars, moon, and sun on the blanket of the sky grew more mysterious with every exploration, presenting herself

in all manners of work, from poetic tales of romance to the fundamentals of human culture, everyone everywhere has bathed in the ocean of her mysteries. The mysteries began with counting the heavenly bodies in the sky and eventually drove people to question the very nature of the Universe. Modern cosmology, the subject of studying the cosmos (Universe), sails on the same ocean, covering the nature and evolution of the Universe from its birth to the present era, hoping to see the shore someday. We, too, will sail the same boat with similar aspirations. In this thesis of our journey, we explore and strengthen our understanding of one of the most actively researched periods in the evolution of the Universe, called the Epoch of Reionization (EoR). We will first begin our journey by studying the evolution of the Universe and develop the necessary skills for the development of the later chapters. From chapter two onwards, we will study the work I carried out during my PhD days at Max Planck Institute for Astronomy (MPIA), and finally, in chapter five, we will conclude with a summary and future applications of my work.

1.1 The Standard Cosmological Model

Most of the discussions and derivations in this section are inspired by the descriptions in Peebles, 1993 and lecture notes from Prof. Luca Amendola.

Before studying the Universe, let's describe the Universe we want to study. The standard model of cosmology that we will use in this study follows Λ-Cold Dark Matter (ΛCDM). Where Λ refers to the dark energy and CDM represents cold Dark Matter (DM), which means the DM is nonrelativistic and only interacts via gravity. ΛCDM assumes the cosmological principle, that the Universe on a large scale is homogeneous and isotropic. The other parameters that define ΛCDM are $\Omega_m, \Omega_\Lambda, \Omega_r, \sigma_8, h$ are described in the section below.

1.1.1 Expanding Universe Dynamics

The expanding Universe, following the cosmological principle, can be described by the Fiedmann-Lemaitre-Robertson-Walker (FLRW) metric:

$$ds^2 = c^2dt^2 - a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right] \tag{1.1}$$

Here (r, θ, ϕ) are the comoving coordinates that do not change with the Universe's expansion, and c is the speed of light. Looking at the equations, one could see the

spherical symmetry on the spatial part of FLRW due to the cosmological principle. We define comoving coordinates from the proper coordinates by multiplying the proper coordinates by the time-dependent scale factor $a(t)$, which has been normalized so that at present $a(t_0) = 1$. The factor k on the dr part of the Eq. (1.1) represents the spatial curvature of the Universe. To formulate the expansion of the Universe, all we now would like to know is the scale factor $a(t)$. This can be derived from Einstein's field equations and the energy-momentum conservation law. Separating the spatial and temporal parts, we get the Friedmann equations.

$$\begin{aligned} \left(\frac{\dot{a}}{a}\right)^2 \equiv H^2(t) &= \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} \\ \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} \end{aligned} \tag{1.2}$$

Λ is the cosmological constant and represents the contribution from dark energy. $H(t) \equiv \dot{a}/a$ is known as the Hubble-Lemaître parameter, which gives a measurement of how fast the Universe is expanding. The present value of the Hubble-Lemaître parameter (H_0) is commonly parametrized using h or h^{-1}, where $h = H_0/(100 kms^{-1}Mpc^{-1})$. The term $\rho(t)$ is the total matter-energy density, and $p(t)$ is the total pressure in the Universe at time t.

For an object at rest in the comoving frame, the distance D and velocity V in the proper coordinates are given by,

$$\begin{aligned} D &= ar \\ \dot{D} &= \dot{a}r = HD \end{aligned} \tag{1.3}$$

This is called the Hubble-Lemaître law.

Using the energy conservation law, we can define the matter density and pressure relation:

$$\dot{p}(t) + 3\frac{\dot{(a)}(t)}{a(t)}(\rho + p) = 0 \tag{1.4}$$

We can solve the above equation by substituting ($p = w\rho$). This gives us the energy-density relation. The value of w depends on which kind of interaction dominates. Solving Eq. (1.4) for ρ, we get

$$\rho \propto a^{-3(1+w)}$$

Substituting this relation in the first Friedmann equations, Eq. (1.2), we get relation for scale factor,

$$a(t) \propto t^{\frac{2}{3(1+w)}}. \tag{1.5}$$

Another useful quantity related to the scale factor $a(t)$ is the redshift. As the Universe is expanding, the wavelength of the photon traveling through the Universe stretches. We can represent this stretch using a parameter called redshift z, which is nothing but the Doppler shift of the photon's wavelength.

$$z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} \tag{1.6}$$

The relation between z and a can be easily derived from Eq. (1.3). For the radiation emitted at t_{em} with wavelength λ_{em}, which is observed at t_o with wavelength λ_o.

$$\frac{d\lambda}{\lambda} = \frac{V}{c} = \frac{HD}{c} = -Hdt = -\frac{da}{a}$$

. The negative sign is because $\lambda_o > \lambda_{em}$ while $dt = t_{em} - t_o$ This gives us $\lambda_{em} = a\lambda_o$. Substituitng this in Eq. (1.6), we get,

$$1 + z = a^{-1} \tag{1.7}$$

Now let's try to break down the density relations for different components (matter, radiation, curvature, and dark energy) from 1.2, to see which component drives which part of the Universe.

- Radiation: The radiation component (relativistic in nature) follows $p_r = \rho_r/3$, that is $w = 1/3$ and $\rho \propto a^{-4}$. Substituting w in Eq. (1.5), we get $a(t) \propto t^{1/2}$. The Universe must be radiation dominated at small values of a or at early times

- Matter: The nonrelativistic matter (DM + baryons) contains the bulk of the mass of the Universe. They interact primarily via gravity. The equation of state for a cold DM is $p = 0$. The total matter density is then $\rho_m = \rho_{DM} + \rho_b$, this implies $\rho_m \propto a^{-3}$, and from the Eq. (1.5), we get $a(t) = t^{2/3}$. Matter must have started to dominate after the radiation dominance ended. The Universe was matter-dominated until $z \sim 0.3$ or $t_{age} \sim 10$ billion years.

- Dark energy: The dark energy or the cosmological constant, which accelerates the expansion of the Universe, is assumed to follow the following equation of state

$p_\Lambda = -\rho_\Lambda$, where the pressure is negative. Thus, the density is constant throughout the evolution of the Universe. The scale factor changes exponentially with time for dark energy domination. The Universe is dominated by dark energy at present.

- Curvature: The final component is the curvature of the Universe, which can be seen from Eq. (1.2) to be evolving as $\propto a^{-2}$.

These components in the Friedmann equation are represented in the form of density parameters, defined as $\Omega_\alpha = \frac{\rho_\alpha(t)}{\rho_c(t)}$, with the critical density $\rho_c(t) = 3H^2(t)/8\pi G$. Thus, the first Friedmann equation can be rewritten as:

$$H^2(t) = H_0^2 \left[\frac{\Omega_{m,0}}{a^3} + \frac{\Omega_{r,0}}{a^4} + \Omega_\Lambda + \frac{\Omega_{k,0}}{a^2} \right] \tag{1.8}$$

Where the subscript '0' denotes the value of the density parameters at present. The values used in this work are provided by Aghanim et al., 2020, with cosmological parameters $h = 0.676$, $\Omega_m = 0.309$, $\Omega_b = 0.0489$, and $\sigma_8 = 0.810, h = 0.7$. Thus completing the description of our standard model.

1.2 Traveling Through Cosmic Time

The evolution of the Universe is divided into various epochs defined by specific physical conditions and dominant processes. This section will briefly discuss these epochs and the chronological transitions between them. Everything started with the Big Bang, *The Planck Epoch*, around 13.8 Gyrs ago, when the Universe was highly energetic and dense. The Universe then transitions into the phases of cooling and reheating throughout its evolutionary stages, giving rise to the Universe we see today. The Fig. 1.1 gives the pictorial view of the evolution of the Universe. After the Big Bang, around $t_{age} \sim 10^{-36}s$, the Universe started to expand exponentially, entering into the *Inflationary Epoch*. The Universe has been expanding ever since. During this period, the Universe underwent exponential expansion (of the order of 10^{26}) within a fraction of a second, from $t_{age} \sim 10^{-36}s$ to $t_{age} \sim 10^{-33}s$. This extreme expansion is hypothesized to cause the flat geometry (flatness problem) and stretch the quantum fluctuations to the astronomical scales, which later became the density perturbations (Guth, 1981). It ended in a transition phase called reheating, leading to the formation of standard model particles. The Universe continued to expand after this, though at a much slower rate, which also allowed the Universe to cool down and enter into various particle-forming

epochs (e.g., Quark epoch, Hadron epoch, Lepton epoch, etc). Around $t_{age} \sim 200-1200s$ the Universe entered the Big Bang Nucleosynthesis (BBN) epoch. During this period, the Universe had cooled enough to allow protons and neutrons to form primordial hydrogen and other light elements like helium, lithium, etc. (Alpher et al., 1948).

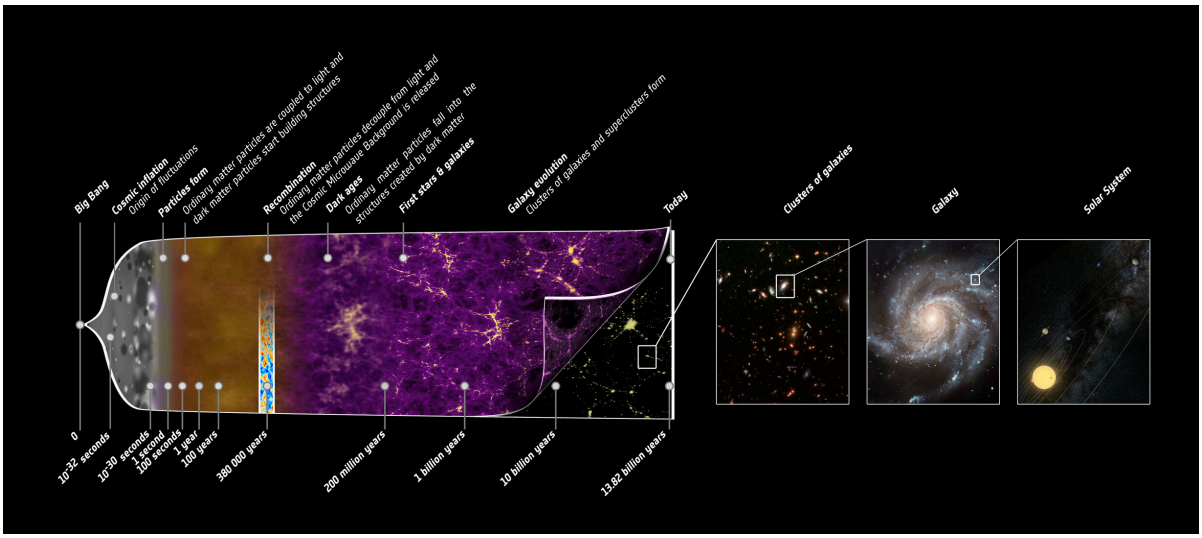


Figure 1.1. This image depicts the evolution of the Universe, marked with major epochs/eras along with their time stamps. The mini panel zooms into the galaxy cluster, the Milky Way galaxy, and the Solar System from left to right. Credits: ESA – C. Carreau

.

After the end of BBN ($t_{age} \sim 1200s$ till $t_{age} \sim 380,000$ yrs), the Universe was in the photon epoch. Initially in this epoch, the Universe was in a radiation-dominated phase, where the energy density of radiation (photons and neutrinos) was greater than that of matter (DM and baryons). The photon pressure was the dominant factor in determining the Universe's expansion rate, see Eq. (1.2) for reference. Since the energy density of radiation decreases much faster than the energy density of matter due to expansion, it eventually lead to the matter-radiation equality, which happened around $z \sim 3500$ or t_{age} 43,000 yrs. After this, the energy density of the Universe was dominated by matter. During the whole photon epoch, photons were constantly scattered by free electrons via Thomson scattering, coupling them into a photon-baryon fluid, leading to a completely opaque Universe. The coupling of photons to the baryonic perturbations caused them to oscillate as acoustic waves known as Baryon Acoustic Oscillations (BAO). The imprints of BAO can be seen in the Cosmic Microwave Background (CMB).

The Universe remained opaque to radiation until z 1100 or t_{age} 380,000, yrs, causing the average photon energy to drop low enough that they cannot ionize the neutral hydrogen (HI) atoms. The electrons can now combine with protons to form stable HI.

This is known as recombination, and the era is hence called the recombination epoch. As the free electrons combined with protons, the number density of electrons reduced, which allowed photons to travel freely without being scattered. This event is called photon decoupling. The Universe started to become transparent to photons. The photons, after decoupling, were observed today as CMB (Penzias and Wilson, 1979). After their last scattering from free electrons during the recombination epoch, the photons could represent the size of the Universe where it transitions from being opaque to being transparent. This is called the surface of last scattering. The CMB photons around the recombination had a blackbody spectrum corresponding to $T \sim 3000\ K$, which, after being redshifted to us, corresponds to the blackbody radiation of $T \sim 2.75\ K$. Due to the primordial density perturbations in the early Universe, small-scale inhomogeneities and anisotropies can also be seen in the CMB. These fluctuations in the CMB can be used to constrain the properties of the Universe (Aghanim et al., 2020). This is the furthest we can see from our telescopes. However, current efforts aim to identify smaller-scale perturbations, probing the Universe as early as $\sim 10^{-14}$s (Biscoveanu et al., 2020; Collaboration et al., 2015; Amaro-Seoane et al., 2023). This era of recombination continued till the rate of recombination was greater than the rate of expansion of the Universe. After this point, the amount of HI in the Universe froze.

After the recombination epoch ended, the Universe entered the Dark Ages until the first bright objects formed. The Dark Age continued from $z \sim 1100$ or $t_{age} \sim 380,000$ yrs, till $z \sim 30$ or $t_{age} \sim 100$ Myrs. There were no galaxies or stars, the Universe was filled with HI and helium. Due to the initial perturbations in the matter density fields, the DM, unaffected by the CMB radiation, started to collapse gravitationally and form the bound structures, called DM halos. The baryonic matter density closely followed the DM gravitational potential wells. The collapse of baryonic gas was opposed by its gas pressure, eventually reaching a quasi-equilibrium state. The gas, primarily HI, could cool down due to thermal collisions followed by de-excitation, hence collapsing further within the DM halos. This eventually led to the formation of the first generation of stars, referred to as Population III or Pop III. The Pop III stars were formed around $z \sim 30$ or $t_{age} \sim 100$ Myrs, leading to the era of Cosmic Dawn. These extremely massive stars were one of the earliest sources of UV radiation ($\geq 100 - 200 M_{\odot}$) (Bromm et al., 1999; Abel et al., 2000; Klessen and Glover, 2023). Due to their large mass, these stars were short-lived (of the order $\sim$ 10 Myr) and ended with supernovae. This led to the metal enrichment of their environment. The metal enrichment allowed for more

effective cooling. Due to this, the gas could further fragment into smaller clumps and form smaller but much longer-lived Pop II stars. These stars built up denser stellar clusters and complicated structures, which eventually led to the assembly of the first galaxies. These galaxies were also a strong UV source. The Pop III stars, the source of UV radiation, were able to ionize the surrounding gas and the Universe. This was also the beginning of the reionization of the Universe. However, due to the short lifespan of Pop III stars, they could not sustain it and could only ionize the Universe locally. Galaxies, on the other hand, were very stable and carried large-scale reionization of the Intergalactic Medium (IGM) via bubble-like structures. These bubbles, which were isolated initially, started to grow and eventually overlap as more and more galaxies were formed. This marked the end of the cosmic dawn era and the beginning of the EoR. The cosmic dawn ended approximately at $z \sim 15$ or $t_{age} \sim 250$ Myrs, when the first galaxies were formed. The epoch of reionization can be debated to have begun around $z \sim 30-15$ from the formation of the first stars to the first galaxies, but it ended approximately at $z \sim 5.5$ or $t_{age} \sim 1$ Gyr (Bosman et al., 2022).

Galaxies of various masses were formed at this time, some were big enough to host Active Galactic Nuclei (AGN) with black holes of mass $M_{BH} \sim 10^6 M_\odot$ (Larson et al., 2023). These AGNs were the progenitors of the earliest Super Massive Black Holes (SMBH) or quasars, around $z \sim 7.5$ or $t_{age} \sim 650$ Myrs. The Universe was almost completely ionized when the reionization ended; the IGM was reheated to $t \sim 10^4$ K. This heating potentially suppressed the formation of very small galaxies. The new Universe has now transitioned to the era where galaxy structure formation dominates. Smaller galaxies and DM halos could merge to form larger and more complex structures. These structures also grew by accreting more and more gas. Baryonic gas was heated because of reionization, and heated the IGM. The gas within the halos was cooler and got cooler due to much more efficient cooling mechanisms (e.g., molecular cooling, etc). This increased the star formation rate (SFR) within the halos. The SFR peaked around $z \sim 2$ or $t_{age} \sim 3.2$ Gyrs (Madau and Dickinson, 2014). Around $z \sim 0.3 - 0.7$ or $t_{age} \sim 7-10$ Gyrs, the energy density from dark energy started to dominate, accelerating the expansion of the Universe, entering into the accelerated expansion or dark energy dominated era, which defines the present era ($z = 0$). We expect this accelerated expansion to continue into the future.

1.3 Theory of Structure Formation

Most of the discussions and derivations in this section are inspired from the works in Peacock, 1998; Padmanabhan, 2000; Barkana and Loeb, 2001; Narlikar, 2002; Mo et al., 2010; Peebles, 1993.

Starting with the standard cosmological principle, where the Universe is homogeneous and isotropic at large scales. The structures present today, such as galaxies, galaxy clusters, and the overall cosmic web, could only be formed if we assume some small-scale fluctuations in the initial density field. The fluctuations in the CMB indicate that the Universe is largely inhomogeneous at small scales. These initial density perturbations will eventually grow over time, giving rise to gravitationally bound stable objects. Eventually, the perturbations would grow so large that the baryonic matter would start to collapse within these stable objects, giving rise to the first stars and galaxies. These objects, being the first sources of light, will eventually mark the beginning of the era of reionization. In this section, we will address how the large-scale and small-scale structures are formed from the initial density perturbations, how they evolve, and how we calculate them in our simulations.

1.3.1 Linear Perturbation Thoery

As the Universe continued to evolve and began its journey to take its modern shape, the initial density perturbations also grew. The scale of these perturbations can be further divided into two categories, first when the amplitude of the perturbation was small and the size of the structure formed was much smaller than the horizon size ($c/H(t)$), such that the density contrast relative to the background was much smaller than unity. The second is when the density contrast is much higher than unity. In this section, we will study the former, the linear perturbation theory.

1.3.1.1 Ideal Fluid

We begin the treatment of the effects of perturbations by using the Newtonian theory for the evolution of the density ρ, under pressure P and velocity $\vec{v}$ of a non-relativistic fluid influenced by the gravitational potential ϕ. This approximation works well for both the pressureless particles, like dark matter[1], and baryons under local thermal equilibrium due to frequent collisions. The equations of motion of such a fluid are then given by

[1] Given the velocity dispersion is small enough to be neglected.

the following equations:

$$\text{Euler}: \frac{D\vec{u}(\vec{r},t)}{Dt} = -\frac{\vec{\nabla}_r p}{\rho(\vec{r},t)} - \vec{\nabla}_r \Phi(\vec{r},t) \tag{1.9}$$

$$\text{Continuity}: \frac{D\rho(\vec{r},t)}{Dt} + \rho\vec{\nabla}_r.[\vec{v}(\vec{r},t)] = 0 \tag{1.10}$$

$$\text{Poisson}: \nabla_r^2 \Phi(\vec{r},t) = 4\pi G\rho(\vec{r},t) \tag{1.11}$$

Here $\vec{r}$[2] is in proper coordinates and $D/Dt = \partial/\partial t + \vec{u}.\nabla$. The above equations have six unknown variables ($\rho, u_x, u_y, u_z, P,$ and ϕ), but only five equations. Thus, these need to be supplemented with the equation of state for the fluid pressure P. Before we write down the equation of state, let's first describe our perturbations in the density $\delta(\vec{r},t)$ as,

$$\delta(\vec{r},t) = \frac{\rho(\vec{r},t) - \langle\rho(t)\rangle}{\langle\rho(t)\rangle} \tag{1.12}$$

where $\rho(\vec{x},t)$ is the density (matter or energy) of the Universe at a position $\vec{x}$ and time t, where $\langle\rho(t)\rangle$ is the mean density over the whole space at that time. It is useful to switch to comoving units when discussing the time evolution of the FLRW Universe by employing the following transformation:

$$\begin{aligned} &\vec{r} = a\vec{x} \\ &\vec{u} = \dot{a}\vec{x} + \vec{v}, \vec{v} \equiv a\dot{\vec{x}} \\ &\nabla_{\vec{r}} \;\rightarrow\; \frac{1}{a}\nabla_x; \frac{\partial}{\partial t} \;\rightarrow\; \frac{\partial}{\partial t} - \frac{\dot{a}}{a}\vec{x}\cdot\nabla_x \end{aligned} \tag{1.13}$$

Where, $\vec{v}$ and $\vec{x}$ are comoving peculiar velocities and distances, a is the scale factor, and $\dot{a}/a = H(t)$ is the Hubble-Lemaître constant. The $\partial/\partial t$ is now with fixed $\vec{x}$. For the sake of convenience, we will drop the $\vec{x}$ subscripts here onwards. Furthermore, unless stated otherwise, all the variables will implicitly depend on $\vec{x}$ and t. Now rewriting, Eq. (1.9), Eq. (1.10) and Eq. (1.11) and using the relation that $\langle\rho\rangle \propto a^{-3}$, we get the following equations in comoving space,

$$\text{Euler}: \frac{\partial\vec{v}}{\partial t} + \frac{\dot{a}}{a}\vec{v} + \frac{1}{a}(\vec{v}\cdot\nabla)\vec{v} = -\frac{\nabla\Phi}{a} - \frac{\nabla P}{a\langle\rho\rangle(1+\delta)} \tag{1.14}$$

[2]From here onward, we will follow the notation; $\vec{r},\vec{u}$ are the distance and velocity in proper coordinates, and $\vec{x},\vec{v}$ are the distance and velocity in comoving coordinates.

$$\text{Continuity}: \frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot [(1+\delta)\vec{v}] = 0 \tag{1.15}$$

$$\text{Poisson}: \nabla^2 \phi = 4\pi G \langle \rho \rangle a^2 \delta, \quad \phi \equiv \phi + a\ddot{a}x^2/2. \tag{1.16}$$

Now to derive the equation of state, we need to consider the thermodynamic processes acting on the fluid. For a general case the Pressure P of the fluid can depend upon both the density ρ and the specific entropy S of the fluid, i.e.,

$$P = P(\rho, S) \tag{1.17}$$

Using the second law of thermodynamics, $dS = dQ/T$, where T is the temperature of the background. The dQ is the infinitesimal addition of heat to the unit mass fluid, and using the difference between heating ($\mathcal{H}$) and cooling ($\mathcal{C}$) rates per unit volume V to describe the change dQ over time, i.e, ($dQ = V(\mathcal{H} - \mathcal{C})dt$), we get

$$T\frac{dS}{dt} = \frac{\mathcal{H} - \mathcal{C}}{\rho}, \tag{1.18}$$

In §1.4, we show that for a gas in IGM, $\mathcal{H}$ can be determined by physical processes such as radiation from stars, and quasars. While $\mathcal{C}$ could be determined by collisional excitation or absorption processes. Now applying the first law of thermodynamics for the monoatomic non-relativistic ideal gas of unit mass ($dQ = dU + W$), we get

$$T\,dS = d\left(\frac{3}{2}\frac{P}{\rho}\right) + P\,d\left(\frac{1}{\rho}\right), \tag{1.19}$$

Using $P = \left(\frac{\rho}{\mu m_p}\right) k_B T$, where μ is the mean molecular weight in terms of m_p, we get the following equation of state,

$$P \propto \rho^{5/3} \exp\left(\frac{2}{3}\frac{\mu m_p}{k_B} S\right). \tag{1.20}$$

Using the Eq. (1.20) to derive the $\nabla P/\bar{P}$ term in the Euler Eq. (1.14),

$$\frac{\nabla P}{\bar{P}} = \frac{1}{\langle \rho \rangle}\left[\left(\frac{\partial P}{\partial \rho}\right)_S \nabla \rho + \left(\frac{\partial P}{\partial S}\right)_\rho \nabla S\right], \tag{1.21}$$

using the definition of adiabatic sound,

$$c_s^2 \equiv \left(\frac{\partial P}{\partial \rho}\right)_S \tag{1.22}$$

we get,

$$\begin{aligned}\frac{\nabla P}{\bar{P}} &= \frac{1}{\langle\rho\rangle}\left[c_s^2\nabla\rho + \frac{2}{3}\rho\, T\,\nabla S\right] \\ &= c_s^2\nabla\delta + \frac{2}{3}(1+\delta)\, T\,\nabla S\end{aligned} \tag{1.23}$$

Substituting the above relation in the Euler Eq. (1.14), we get,

$$\frac{\partial\vec{v}}{\partial t} + \frac{\dot{a}}{a}\vec{v} + \frac{1}{a}(\vec{v}\cdot\nabla)\vec{v} = -\frac{\nabla\Phi}{a} - \frac{c_s^2}{a(1+\delta)}\nabla\delta - \frac{2T}{3a}\nabla S \tag{1.24}$$

We can argue that for the linear regime, the perturbations $\delta \ll 1$ and the velocity $\vec{v} \ll 1$, we can ignore all the non-linear terms. Hence we can rewrite continuity Eq. (1.15), and Euler equation from Eq. (1.24) into a linearized form as,

$$\text{Continuity}: \frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla\cdot\vec{v} = 0 \tag{1.25}$$

$$\text{Euler}: \frac{\partial\vec{v}}{\partial t} + \frac{\dot{a}}{a}\vec{v} = -\frac{\nabla\Phi}{a} - \frac{c_s^2}{a}\nabla\delta - \frac{2T}{3a}\nabla S \tag{1.26}$$

Now if we take the partial derivative of Eq. (1.25) with respect to t, we get the following sets of relations:

$$\begin{aligned}\nabla\cdot\vec{v} &= -a\cdot\frac{\partial\delta}{\partial t} \\ \nabla\cdot\frac{\delta\vec{v}}{\delta t} &= -a\cdot\frac{\partial^2\delta}{\partial t^2}\end{aligned} \tag{1.27}$$

Taking the gradient of the Eq. (1.26), then substituting for the results from above and the Eq. (1.16), we get

$$\boxed{\frac{\partial^2\delta}{\partial t^2} + 2H\frac{\partial\delta}{\partial t} = 4\pi G\langle\rho\rangle\delta + \frac{c_s^2}{a^2}\nabla^2\delta + \frac{2T}{3a^2}\nabla^2 S,} \tag{1.28}$$

Where, $H = \dot{a}/a$ is the Hubble-Lemaître constant. Let's compare the above equation with a damped harmonic oscillator. We can see that the second term on the left-hand side of the above equation is like the damping coefficient; it represents the Hubble drag, which does surpass the perturbation growth via the expansion of the Universe. The terms on the right hand side represent the external force, this external force could either aid in the growth or dampen the system depending on their relative phase with the $\partial\delta/\partial t$.

The first term on the right hand side, the gravitational term causes the perturbation to grow because of gravitational instability. The second term, $\nabla^2\delta$ represents the spatial variations on density while $\nabla^2\S$ represents the spatial variations on entropy.

For the linear regime approximation, where the size of the structures are much smaller than $(c/H(t))$, we can neglect the curvature of the Universe. In that case we can represent the perturbations as a Fourier transform of plane waves, i.e.,

$$\delta(\vec{x},t) = \sum_{\vec{k}} \delta_{\vec{k}}(t)\exp(i\vec{k}\cdot\vec{x}), \quad \delta_{\vec{k}}(t) = \frac{1}{V_{box}}\int \delta(\vec{x},t)\exp(-i\vec{k}\cdot\vec{x})\,d^3\vec{x} \tag{1.29}$$

Where V_{box} is the volume of our simulation box with periodic boundary conditions. The wave vector $\vec{k}$ is also written in comoving space. Now taking the Fourier transform of Eq. (1.28), which is done simply by replacing $\delta(\vec{x}) \rightarrow \delta(\vec{k})$, $S(\vec{x}) \rightarrow S(\vec{k})$, $\nabla \rightarrow i\vec{k}$ and $\nabla^2 \rightarrow -k^2$, we get the equations of motion for the perturbations in the Fourier space,

$$\frac{d^2\delta_{\vec{k}}}{dt^2} + 2\frac{\dot{a}}{a}\frac{d\delta_{\vec{k}}}{dt} = \left[4\pi G\langle\rho\rangle - \frac{k^2c_s^2}{a^2}\right]\delta_{\vec{k}} - \frac{2T}{3a^2}k^2S_{\vec{k}} \tag{1.30}$$

The Fourier transform of the continuity Eq. (1.25) gives the relation for the velocity field,

$$\vec{v}_{\vec{k}} = \frac{iak}{k^2}\frac{d\delta_{\vec{k}}}{dt}. \tag{1.31}$$

As the above calculations are much faster in Fourier space. These results are very crucial for the understanding of how the density fields are calculated and evolved within a simulation box. In the §2.2, we will use the Fourier transform of the velocity field derived above to perturb the halos to their final locations. We will reference the above two equations again when discussing how our simulation boxes work.

1.3.1.2 Special Cases

Now we will try to solve the equation of motion for the perturbation (Eq. (1.28) and Eq. (1.30)) in the linear regime for the following cases:

- Isentropic initial perturbations with Adiabatic evolution: In this case we can ignore the specific entropy and set $k^2S(\vec{k}) = 0$ in the Eq. (1.30). Thus giving us the relation:

$$\ddot{\delta}_{\vec{k}} + 2\frac{\dot{a}}{a}\dot{\delta}_{\vec{k}} = \left[4\pi G\langle\rho\rangle - \frac{k^2 c_s^2}{a^2}\right]\delta_{\vec{k}} \tag{1.32}$$

We can either move back to proper units (in which case there will be no $\frac{\dot{a}}{a}\vec{x}\cdot\nabla_x$ term and the above equation will reduce to simple equation of $\ddot{\delta}_{\vec{k}} = \left[4\pi G\langle\rho\rangle - \frac{k^2 c_s^2}{a^2}\right]\delta_{\vec{k}}$, or we can just ignore the expansion of the Universe in the equation Eq. (1.32). In either case we will get an equation similar to a simple harmonic oscillator, whose solutions can be given by:

$$\delta(t) = e^{\pm\, t/\tau}\ ,\ \tau = \frac{1}{\sqrt{4\pi G\langle\rho\rangle\ -\ c_s^2 k^2}} \tag{1.33}$$

These solutions to the above equation depend on a characteristic length scale $\lambda_J = 1/k_J$. It characterizes the length scale at which the above equation transitions from oscillatory mode to growing/ decaying mode. This length is called Jeans length and is defined as,

$$\lambda_J = c_s\sqrt{\frac{\pi}{G\langle\rho\rangle}} \tag{1.34}$$

for the $\lambda < \lambda_J$, or $k > k_J$, the solution is a combination of sine and cosine waves (oscillatory solution), hence the perturbations are bound. But for the $\lambda > \lambda_J$, or $k < k_J$, the solution is now a linear combination of two exponential functions, one representing the growing mode ($e^{t/\tau}$) while the other represents the decaying mode ($e^{-t/\tau}$). For large time scales, the perturbation could grow very large and cause the gravitational collapse.

Since pressure forces in a fluid travel at the speed of sound within the fluid, which can be written as $c_s \propto \lambda_J(G\rho)^{-1/2}$, i.e., the pressure forces can cover λ_J distance within $(G\rho)^{-1/2}$ time (gravitational free-fall time). Thus for any length shorter than Jeans length $l < \lambda_J$, the pressure forces are fast enough counter balance the gravitational pull. But for the length scales larger than λ_J the pressure force can not react to the gravitational build up. Thus for the mass of fluid greater than within a sphere of Jeans length diameter, i.e., $M_J = 4\pi\rho/3 \times (\lambda_J/2)^3$, called Jeans mass, the pressure gradient could not support the self gravity, thus causing the fluid to collapse. By substituting the value of λ_J in M_J, we get, $M_J \propto \rho^{-1/2}$.

Similarly, recalling that for a non relativistic ideal gas $P \propto \rho T$, thus the sound speed $c_s = (\sqrt{\partial P/\partial \rho})_s \propto T^{1/2}$. This gives us the relationship between $M_J \propto \lambda_J^3 \propto c_s^3 \propto T^{3/2}$.

- Isentropic and pressureless fluid: In this case we will ignore even the pressure term and just look at the gravitational term. The equation of motion for perturbations for this fluid is given by,

$$\ddot{\delta}_{\vec{k}} + 2H\dot{\delta}_{\vec{k}} = 4\pi G\langle\rho\rangle\delta_{\vec{k}} \tag{1.35}$$

Where $H = \dot{a}/a$, Hubble-Lemaître constant. From our cosmological standard model, we showed,

$$H(t) \equiv \frac{\dot{a}}{a}, \quad \frac{dH}{dt} + H^2 = -\frac{4\pi G}{3}(\rho_m - 2\rho_v) \tag{1.36}$$

Where, $\rho_m \propto a^{-3}$ is the matter density and $\rho_v =$ constant. Now differentiating the above equation with t while using the fact that $d/dt = d/da \times da/dt = \dot{a}d/da = aHd/da$. Thus $d\rho/dt = aHd\rho/da \propto -3H\rho$, we get,

$$\frac{d^2H}{dt^2} + 2H\frac{dH}{dt} = 4\pi G\rho_m H \tag{1.37}$$

The above equation looks just the same as the Eq. (1.35). Thus one of the solutions of $\delta_{\vec{k}}(t) = H(t)$. For both the matter and radiation dependent Universes, it can be shown $H(t) \propto t^{-1}$. Thus one of the solutions for $\delta_{\vec{k}}(t) \propto t^{-1}$. Now, using the definition of the Wronskian, $W(t) = \delta_1\dot{\delta_2} - \delta_2\dot{\delta_1}$, if δ_1 and δ_2 are the solutions of the Eq. (1.35), one can show that

$$\begin{aligned} &\frac{dW}{dt} = -2H(t)W(t) \\ &\int \frac{dW}{W} = -2\int H(t)dt \ \propto -2\ln(a) \\ &\implies W = \delta_1\dot{\delta_2} - \delta_2\dot{\delta_1} =\propto \ a^{-2} \end{aligned} \tag{1.38}$$

Now after substituting $\delta_1 \propto t^{-1}$ in the above relation, we get the growing mode solution,

$$\delta_2 \propto H(t) \int \frac{dt'}{a^2(t')H^2(t')} \tag{1.39}$$

for an Einstein-de Sitter Universe, $\delta_2 \propto t^{2/3}$, thus the solutions to the above case are:

$$\begin{aligned} &Growing\ mode : \delta \propto H(t) \int_0^a \frac{dt'}{H^2(t')a^2(t')} \\ &Decaying\ mode : \delta \propto H(t) \end{aligned} \tag{1.40}$$

The growing mode solution sometimes is also written as,

$$\delta_2 \propto D(z) \propto \frac{g(z)}{(1+z)} \tag{1.41}$$

where $g(z)$ is the growth factor and can be approximated to Carroll et al., 1992

$$g(z) \approx \frac{5}{2}\Omega_m(z) \left[\Omega_m^{4/7}(z) - \Omega_\Lambda(z) + \left(1 + \frac{\Omega_m(z)}{2}\right)\left(1 + \frac{\Omega_\Lambda(z)}{70}\right)\right]^{-1} \tag{1.42}$$

1.3.2 Nonlinear Evolution

As the Universe continues to evolve to large time scales, it is dominated by the growing mode solution from Eq. (1.41). Thus, making perturbations bigger and bigger.

We can note that even though we assumed a linear limit in the earlier section, the density and velocity equations are nonlinear. A complete nonlinear solution can only be provided by N-body simulations. In these simulations, a discrete set of particles are used. These particles interact with each other via gravitational interactions. The equation of motions of all these particles are solved at small time steps iteratively, where the position and velocities are modified at each times stamp. However, these simulations are slow and computationally expensive. It is beneficial to rather study some approximate cases of nonlinear perturbations to develop semi-numerical models to evaluate the physics beyond the linear regime. Furthermore, these approximate solutions can provide very accurate physical understanding of the non linear evolution of the Universe.

1.3.2.1 Zeldovich Approximation: Quasi Linear Approximation

We begin our treatment of non-linear perturbation by employing the simple but very effective technique of transforming coordinates from Eulerian to Lagrangian, Yakov B. Zel'dovich, 1970 showed how the Linear Perturbation Theory (LPT) in this frame can

encompass the non-linear effects from Eulerian coordinates. In this section, we will use the Zeldovich approximation to calculate the velocity and density distribution of a volume of particles to picture its kinematics.

Let us begin by assuming the ideal fluid case of pressureless gas shown above (Newtonian approximation) and considering that enough time has passed that we are well within the growing perturbation regime, but also within the linear approximation case. For such a case, we can write a particle's actual position $\vec{r}$ in terms of time t and its Lagrangian coordinate $\vec{q} \equiv \vec{q}(t = 0)$, up to first order as,

$$\vec{r} = a(t)\vec{q} + b(t)\vec{f}(\vec{q}) \tag{1.43}$$

The first term of the above equation describes the cosmological expansion, $b(t)$ is the temporal growth function, and $\vec{f}(\vec{q})$ is the spatial distribution of the initial perturbations.

The second term in the above equation looks very similar to the linear evolution of density perturbation as derived in Eq. (1.41),

$$\delta(\vec{r}, t) = D(t)\delta_i(\vec{r}), \tag{1.44}$$

where $\delta_i(\vec{r})$ is the density perturbation at some initial time and $D(t)$ is normalized such that $D(t_i) = 1$. The density field thus grows self-similarly with time[3]. Thus, we could assume that $\vec{f}(\vec{q})$ should be some function of $\delta(\vec{x})$, while $b(t)$ should be related to $D(t)$[4]. But let's not get ahead of ourselves and try to derive the relations.

Now, let us calculate the deformation tensor of a small volume of particles centered at $\vec{q}$,

$$D_{ik} \equiv \frac{\partial \vec{x_i}}{\partial \vec{q_i}} = a(t)\delta_{ik} + b(t)\frac{\partial \vec{f_i}}{\partial \vec{q_i}}. \tag{1.45}$$

[3]Which means for any times t_1 and t_2 we get $\delta(\vec{r}, t_2) = (D(t_2)/D(t_1))\delta(\vec{r}, t_1)$.

[4]Spoiler: I got ahead of myself, if we write the position of the particle in comoving coordinates, i.e., $\vec{r} = a(t)\vec{r}$, where $\vec{r} = \vec{q} + (b(t)/a(t))\vec{f}(\vec{q})$ we can see that the growing term is related to $b(t)$ as $D(t) = (b(t)/a(t))$. The perturbations in space can be related to the perturbations in density by recalling that the only force that will carry the perturbations is the gravitational force, which depends on the masses of the interacting particles. From the Poisson equation in comoving coordinates Eq. (1.16) we can write $\nabla^2\phi(\vec{r}, a) = 4\pi G\rho^2\delta(\vec{r}, a) = 4\pi G\rho^2 D(a)\delta_i(\vec{r}) = (D(a)/a)(4\pi G\rho^3\delta_i(\vec{r}))$, now from the Eq. (1.11), Poisson : $\nabla_r^2\Phi(\vec{r}, t) = 4\pi G\rho(\vec{r}, t) = 4\pi G\rho_i a^3\delta(\vec{r})$, we get $\nabla^2\phi(\vec{r}, a) = (D(a)/a)\nabla^2\phi_i(\vec{r})$. Solving this equation, we get $\phi(\vec{r}, a) = (D(a)/a)\phi_i(\vec{r})$ up to some constant. Now recalling that the force $F(\vec{r}, a) = -\nabla\phi(\vec{r}, a) = -(D(a)/a)\nabla\phi_i(\vec{r})$ this perturbation in force is going to drive the perturbation in the position of the particles. Taking the divergence of the force, we get $\nabla \cdot F(\vec{r}, a) = -(D(a)/a)\nabla^2\phi_i(\vec{r}) = -(D(a)/a)4\pi G\rho_i a^3\delta(\vec{r})$ or $\nabla \cdot F(\vec{r}, a) \propto -D(a)\delta(\vec{r})$. The term $\vec{f}(\vec{q})$ is nothing but some sort of acceleration given by this force. Hence, we can predict that $\nabla \cdot f(\vec{r}, a) \propto -D(a)\delta(\vec{r})$.

Where δ_{ik} is the Kronecker delta, not the density. The tensor D_{ik} is also the Jacobian matrix for the transformation from $\vec{r}$ to $\vec{q}$. Since we assumed only the growing mode for perturbation, which either stretches or compresses the volume (depending on time and a(t)), thus D_{ik} is a positive definite symmetric matrix. Which means we can choose $\vec{q}$ such that the coordinate axes align along $\frac{\partial \vec{f}_i}{\partial \vec{q}_i}$ (the set of fundamental axes), we can write D_{ik} as,

$$D = \begin{Vmatrix} a(t) - \alpha b(t) & 0 & 0 \\ 0 & a(t) - \beta b(t) & 0 \\ 0 & 0 & a(t) - \beta b(t) \end{Vmatrix}. \tag{1.46}$$

This represents how a perturbed volume, which looks like a cube at t_i has transformed into a parallelepiped. Where $(-\alpha, -\beta, -\gamma)$ are the three eigenvalues of the strain tensor $\partial \vec{f}_i / \partial \vec{q}_j$. Using this, we can now calculate the density near the particle from the conservation of mass,

$$\frac{\rho(\vec{r}, t)}{\langle \rho(t) \rangle} = \left[\left(1 - \frac{b}{a}\alpha\right) \left(1 - \frac{b}{a}\beta\right) \left(1 - \frac{b}{a}\gamma\right) \right]^{-1} \tag{1.47}$$

These eigen values (α, β, γ) are functions of $\vec{q}$. We can find the maximum values for these eigenvalues before they collapse the volume. Let's look at α, if $\alpha(\vec{q}) > 0$, then for $\alpha = \alpha_m$, such that at time $t = t_c$, $a(t_c) - \alpha_m b(t_c) = 0$. This will give us the moment when $\rho \to 0$, hence collapsing the volume. Infinite density resulting from the unilateral compression along the $\alpha-$axis. It is possible that the collapse can happen along more than one direction; the probability of collapse happening along one or two directions is $\sim 42\%$, whereas for the collapse to happen along all three directions, the probability is $\sim 8\%$ (Doroshkevich and Ya B Zel'dovich, 1964). For more than one direction, the order of collapse axes depends on which axis has the largest value. The unilateral compression transforms the three-dimensional ellipsoid in $\vec{q}$-space into a flat two-dimensional sheet in $\vec{r}$, called pancakes. The bilateral compression will form the filament-like structures, the collapse along all three dimensions will create a node, and if all $(\alpha, \beta, \gamma > 0)$, then there will be no collapse and the void will be formed instead. From here onwards, we revert to our old notation $a \equiv a(t)$, where we omit the explicit mentioning of the time dependency, and we do the same with $b \equiv b(t)$.

Now, we can define our equations of motion. Taking the time derivative of Eq. (1.43), we get,

$$\begin{aligned} \vec{u} &= \dot{a}\vec{q} + \dot{b}\vec{f}(\vec{q}); \\ \frac{d\vec{u}}{dt} &= \ddot{a}\vec{q} + \ddot{b}\vec{f}(\vec{q}), \end{aligned} \tag{1.48}$$

Since we assumed only the growing mode, D was symmetric. This also implies that D and by extension $\vec{f}$ are irrational. Thus we can write $\vec{f}$ as a gradient of some potential,

$$\vec{f}(\vec{q}) = \vec{\nabla}\psi(\vec{q}) \tag{1.49}$$

Now, if we expand Eq. (1.47) and keep only the first order terms, then we can derive the relation for density perturbation δ and $\vec{f}$ as follows,

$$\delta = -\frac{b}{a}(\alpha + \beta + \gamma) = -\frac{b}{a}\vec{\nabla}.\vec{f} \tag{1.50}$$

From Eq. (1.44), we can confirm that $D = b/a$. Similarly, we can calculate the co-moving velocity $\vec{v}$ from Eq. (1.48).

$$\vec{v} = \frac{1}{a}(\dot{\vec{x}} - \frac{\dot{a}}{a}\vec{x}) = \frac{1}{a}\left(b - \frac{\dot{a}b}{a}\right)\vec{f} \tag{1.51}$$

The comoving velocity relation is consistent with the continuity relation in Eq. (1.15) ($\nabla \cdot \vec{v} = -\dot{\delta}$). From Freidmann equations in Eq. (1.2) and the growth equation for δ from Eq. (1.35), we get the following relation for b

$$\frac{\ddot{b}}{b} = -\frac{2\ddot{a}}{a} = \frac{8\pi G\rho_0}{3} \tag{1.52}$$

For $\Omega = 1$, $b \propto t^{4/3}$. Using this relation, we can now calculate the equations of motion of the particle volume from Eq. (1.48). Similar to the Eq. (1.31), we can calculate the displacement field $\vec{f}$ in Fourier space:

$$\vec{f_k} = -i\frac{\delta_k}{k^2}\vec{k}. \tag{1.53}$$

Summary

Zeldovich approximation is also a first-order perturbation theory, but in Lagrangian space. The benefit that this coordinate transformation gives is that when we linearize the equations in Eulerian coordinates, we ignore the terms like $\nabla \cdot (\vec{v}\delta)$ and $\vec{v} \cdot \nabla\vec{v}$, see Eq. (1.24), but in Zeldovich approximation, we linearize the displacement field $\vec{f}$ instead of the density δ. Since $\delta \propto \nabla \cdot f \propto \nabla \cdot \vec{v}$, thus taking into account the contribution from the velocity gradients. That's why it gives accurate results even when the fluctuations are nonlinear, hence called quasi-linear approximation. Furthermore, the Zeldovich approximation can calculate the collapse, which can be well described by the equations of motion explained in this section. It starts to fail after forming these collapsed structures, since the particles no longer move along the direction of initial velocities as assumed in this theory; rather, these particles oscillate in the deep potential created by the collapse. Thus, we must go beyond the quasi-linear approximation to study the collapsed structures.

1.3.2.2 Spherical Collapse

As the Universe continues to evolve, the perturbations get bigger and bigger, to a point where the density perturbation of the structure being formed (galaxies and galaxy clusters) is much greater than unity, $\delta \gg 1$. Such objects lie way beyond the linear and quasi-linear perturbation theories. This section will study the nonlinear gravitational collapse in $\Lambda = 0$ Universe. Since we are concerned with the Universe at high-redshift, which was matter-dominated, we can neglect the contributions from Λ.

Let's assume a spherically symmetric shell of mass M within the shell and radius r, which evolves according to the Newtonian equation, given by,

$$\frac{d^2r}{dt^2} = -\frac{GM}{r^2} \tag{1.54}$$

Assuming the mass M is constant during the collapse, hence we can integrate the above equation to get the energy relation,

$$\frac{1}{2}\left(\frac{dr}{dt}\right)^2 = \frac{GM}{r} + E \tag{1.55}$$

Where E is the specific energy. The solutions to non-trival values of E for the above equation can be given in the parametric form, which depends on the sign of E. For the collapse of the shell, $E < 0$, the solution to the above equation can be parameterized as,

$$r = A(1 - \cos\theta), \quad t = B(\theta - \sin\theta) \tag{1.56}$$

where A and B are two constants which can be determined from the initial conditions. During earlier times, when $\theta \ll 1$, we can expand the Eq. (1.56) in the powers of θ, keeping only the terms up to the order θ^5. By substituting it back into the Eq. (1.54) and Eq. (1.55), we can solve for A and B.

$$A^3 = GMB^2 \text{ and } A = GM/(-2E) \tag{1.57}$$

The values of A and B can be calculated from the initial conditions. At $\theta = \pi$, the shell attains the maximum value of radius, $r_{\max} = 2A$, at the maximum radius time $t_{\max} = \pi B$.

Usually, r and t are represented in terms of Lagrangian radius $r_l = (3M/4\pi\langle\rho(t)\rangle)^{1/3}$, perturbation δ_l and the growth factors as shown in the Eq. (1.42), $\delta_i = \delta_l(t)a_i g_i/a_t g_t$, where δ_i is the initial perturbation which evolved to time t via LPT, a_t is the scale factor at time t and g_t is the growth factor at time t. For the present time $t_0 \gg t_i$ (sufficiently early time), the growth factor around this time, $g_i \approx 1$, thus the density perturbation at time t, is written as $\delta_i \approx (\delta_l(t)g_t)a_i/a_t$. The density parameter Ω_i evolves with time t in a similar fashion, i.e., $(\Omega_i^{-1} - 1) = (\Omega_t^{-1} - 1)a_i/a_t$. For our case, at high redshift, the Universe is matter-dominated, we also make the following substitutions, $t_i^{-1} \approx \frac{3}{2}\Omega_t^{1/2}H_t(a_t/a_i)^{3/2}$ and $r_l \approx r_i a_t/a_i$. Hence, the new forms of r and t are now written as,

$$\frac{r}{r_i(t)} = \frac{1}{2} \cdot \frac{1 - \cos\theta}{\left[5\delta_i(t)/3g(t) + (1 - \Omega_i^{-1})\right]} \tag{1.58}$$

$$H_i t = \frac{1}{2\Omega_i^{1/2}} \cdot \frac{\theta - \sin\theta}{\left[5\delta_i(t)/3g(t) + (1 - \Omega_i^{-1})\right]^{3/2}} \tag{1.59}$$

The above two equations can provide us with the complete description of the evolution of the spherical shell. To understand the implications of the above equations, we can look at the following cases:

- Case 1 - Early time: For very early time, we can approximate $t \approx B\ \theta^3/6$ and $r \approx A\,\theta^2/2$. Solving for r we get $r \propto t^{2/3}$, which agrees with our linear perturbation growing mode results. The shell expands with the bubble flow, $a \propto t^{2/3}$.

- Case 2 - Turnaround: The gravitational pull of the sphere gets large enough that it stops expanding according to the Hubble flow and achieves a maximum radius of

$$r_{\max}/r_i(t) = \left[\delta_i(t)/3g(t) + (1 - \Omega_i^{-1})\right]^{-1} \tag{1.60}$$

at the turnover time t_{ta}.

$$H_i t_{ta} = \frac{\pi}{2\Omega_i^{1/2}} \left[\delta_i(t)/3g(t) + (1 - \Omega_i^{-1})\right]^{-3/2}. \tag{1.61}$$

The overdensity/ density perturbation required for the spherical shell to reach the size r_{max} at time t_{ta} is,

$$\begin{aligned}\delta_l(t_{ta}) &= \frac{3g(t_{\rm ta})}{5}\left\{\left[\frac{\pi}{2\Omega_i^{1/2}(t_{\rm ta})H(t_{\rm ta})t_{\rm ta}}\right]^{2/3} - \left[1 - \Omega^{-1}(t_{\rm ta})\right]\right\} \\ &= \frac{3}{5}\left(\frac{3\pi}{4}\right)^{2/3} \approx 1.06\end{aligned} \tag{1.62}$$

Which follows from the assumption $\Omega = 1$, and for the linear regime $H_i t_i \approx 2/3$. Using the property $\rho(t) = \langle\rho(t)\rangle(r_l/r(t))^3$, where $\langle\rho(t)\rangle$ is the background density at r_l, we can write the true density as,

$$\rho_{r_{\rm max}}(t_a) = \langle\rho(t_a)\rangle\left(\frac{r_l(t_a)}{r_{\rm max}(t_a)}\right)^3 = \left(\frac{3\pi}{4}\right)^2\langle\rho(t_a)\rangle \tag{1.63}$$

- Case 3 - Collapse: As the perturbation continues to increase, the sphere starts contracting. Under the influence of gravity alone for $\theta = 2\pi$ it entirely collapses to a singularity. The time it takes to collapse $t_{col} = 2t_{ta}$. Substituting the value of t_{col} in the Eq. (1.62), we get the overdensity requirement for the collapse as,

$$\begin{aligned}\delta_l(t_{col}) &= \frac{3g(t_{\rm col})}{5}\left\{\left[\frac{\pi}{\Omega_i^{1/2}(t_{\rm col})H(t_{\rm col})t_{\rm col}}\right]^{2/3} - \left[1 - \Omega^{-1}(t_{\rm col})\right]\right\} \\ &= \frac{3}{5}\left(\frac{3\pi}{2}\right)^{2/3}(\Omega(t_{col})^{0.0185} \approx 1.686(\Omega(t_{col})^{0.0185}\end{aligned} \tag{1.64}$$

Note, that all the factors in the Eq. (1.62) at t_{col}, depends only on $\Omega(t)$, hence $\delta_l(t_{col})$ is also a function of $\Omega(t)$, although the dependency is very weak. Similar to the turnover case, the results here are extrapolated from the linear regime.

- Case 4 - Viriaization: However, the collapse in principle does not occur. This is because the above equations are invalid for very low values of r. As the shell collapses, the particles in the mass shell could cross the shell, and hence our

assumption of constant mass fails. Furthermore, as the mass shell starts to collapse, the work done by the gravitational potential energy in collapsing raises the kinetic energy of the particles within the shell. Due to which, at a specific size, the kinetic energy (K) is big enough to balance the gravitational potential (U) and form something called virialized objects. This usually happens when $U = -2K$ (the virial objects, also shown in §1.3.6, are formed when the shell size has reduced to 1/2 of the maximum value. In our model, this occurs at $\theta = 3\pi/2$ with $\delta_l(t_{vir}) = 1.4$.

The Virial theorem states that for a bound system in equilibrium, the time-averaged kinetic energy ($\langle K \rangle$) of the system is related to the time-averaged potential energy $\langle U \rangle$ of the same system. For potential energy of the form $U \propto r^n$, the virial theorem is written as,

$$\begin{aligned} &\langle K \rangle = -\frac{n}{2}\langle U \rangle \\ &\text{for gravitational potential, where n = -1} \\ &\langle K \rangle = -\frac{1}{2}\langle U \rangle \end{aligned} \tag{1.65}$$

This is a very crucial theorem for the structure formation under quasi-static equilibrium.

Let us try to quickly calculate the true nonlinear over-density of the virialized objects. Assuming $\Lambda = 0$ and a matter-dominated Universe, let's consider a sphere of mass M, which has reached the maximum expansion at some time t_{max} after which it starts to collapse under its gravity. At the maximum expansion, the total energy of the system is given by the internal energy $E = -3GM^2/5r_{max}$. Since it's the turnaround point, the kinetic energy is zero. When the sphere starts to collapse, the energy spent in the work done by gravity to reduce the size, is converted into the kinetic energy of the particles making up this system[5]. The sphere eventually reaches a quasi-static equilibrium state, where the gravitational collapse is supported by the kinetic energy of the particles within the system. The properties of such objects are described by the virial theorem, Eq. (1.65). We will look at the different virial quantities of such a quasi-static object in the §1.3.6.

[5]The primary idea of the collapse is in redistribution of the gravitational potential energy. For baryons it is easy to see how the kinetic energy comes into the picture. But for DM the gravitational interactions alone can redistribute the energy via the process called "Shell Crossing" and "Violent Relaxation" (Lynden-Bell, 1967).

For now, let's try to calculate the over-density relation for such objects. Assuming that the object still has the uniform distribution of mass, the potential energy at the final virial radius r_{vir} is thus given by $U = -3GM^2/5r_{vir}$. According to the virial theorem, $U = 2E = -2K$, where K is the kinetic energy of the system. This implies that $r_{vir} = r_{max}/2$. Usually, different literature assumes different values of time it takes for the system to virialize, one can either assume it to be $t(\theta = 3\pi/2)$ in the Eq. (1.59) or $t(\theta = 2\pi) = 2t_{max}$. Here we will use $t_{vir} = 2t_{max}$ as in Mo et al., 2010. The overdensity can thus be written as,

$$\begin{aligned} 1 + \Delta_{vir} &= \frac{\rho_{t_{vir}}}{\langle \rho_{t_{vir}} \rangle} = \frac{\rho_{t_{tmax}} (r_{max}/r_{vir})^3}{\langle \rho_{t_{2tmax}} \rangle} \\ &= \frac{\rho_{t_{tmax}}}{\langle \rho_{t_{tmax}} \rangle} \frac{\langle \rho_{t_{tmax}} \rangle}{\langle \rho_{t_{2tmax}} \rangle} \left(\frac{r_{max}}{r_{vir}} \right)^3 \end{aligned} \tag{1.66}$$

Recall from the Eq. (1.63), $\rho_{t_{tmax}}/\langle \rho_{t_{tmax}} \rangle = (3\pi/4)^2$, and $\langle \rho_{t_{tmax}} \rangle / \langle \rho_{t_{2tmax}} \rangle = 4$, we get $\Delta_{vir} = 18\pi^2 \approx 178$.

The above relation can be generalized for $\Omega_m \neq 1$ (Bryan and Norman, 1998) using the following approximation,

$$\begin{aligned} \Delta_{vir} &\approx 18\pi^2 + 82x - 39x^2 \quad \text{for} \quad \Omega_R = 0 \\ \Delta_{vir} &\approx 18\pi^2 + 60x - 32x^2 \quad \text{for} \quad \Omega_\Lambda = 0 \end{aligned} \tag{1.67}$$

where $x = \Omega(t_{vir}) - 1$.

1.3.3 Halo Mass Function

In the above section, we described the spherical collapse of matter (DM and baryonic) into virialized objects called DM halos. These DM halos are crucial for the formation of galaxies and stars. In this section, we will calculate the Halo Mass Function (HMF) to formulate how these halos are populated and distributed. We will present both the Press-Schecter Press-Schechter (PS) (Press and Schechter, 1974) and Excursion-Set (ES) (Bond et al., 1991) formalisms.

Before we begin to describe the individual formalism, we first need to set the stage for these formalisms to work upon. Let's define the density field at a given time by considering the overdensity field evolving linearly with time t, defined as $\delta(\vec{x}, t) = \delta_0(\vec{x})D(t)$. Where $\delta_0(\vec{x})$ is the overdensity field linearly evolved to the time t using

the normalized growth factor $D(t)$, as shown in Eq. (1.42). From the Eq. (1.64), the critical overdensity field, at which the collapse would occur to form a virialized object is $\delta(\vec{x},t) > \delta_c = 1.686$ or equivalently $\delta_0(\vec{x}) > 1.686/D(t) \equiv \delta_c(t)$. The $\delta(\vec{x},t)$ is then convolved with a smoothing window function $W(\vec{x};R)$, which can connect these density fields to the mass. The $W(\vec{x};R)$ is a window function with the characteristic filter size R corresponding to a mass M. The dependency of $W(\vec{x};R)$ on M depends on the kind of window function we use. In Chapter 2, we will demonstrate the working of 21cmFAST using a k-sharp top hat window function. The smoothed overdensity then looks like,

$$\delta_s(\vec{x};R) = \int d^3y\, W(|\vec{x}-\vec{y}|;R)\delta(\vec{y}). \tag{1.68}$$

From here onward, we will start distinguishing between PS and ES.

1.3.3.1 Press-Schecter Formalism

In PS Formalism, we fix the size of the filter and convolve the density at each pixel with $W(\vec{x};R)$, subject to the periodic boundary condition, and then check the collapsed condition. i.e., $\delta_s > \delta_c$. Thus, the PS formalism states that the probability $\delta_s > \delta_c$ is the same as the fraction of total mass present in the form of halos at time t, with mass greater than M. If we assume $\delta_0(\vec{x})$ is Gaussian random field, hence $\delta_s(\vec{x})$, also becomes a Gaussian random field, we can define the probability that $\delta_s > \delta_c$ as,

$$P(\delta > \delta_c(t)) = \frac{1}{\sqrt{2\pi}\sigma(M)} \int_{\delta_c(t)}^{\infty} \exp\left[-\frac{\delta_s^2}{2\sigma^2(M)}\right] d\delta_s = \frac{1}{2}\,\mathrm{erfc}\left(\frac{\delta_c(t)}{\sqrt{2}\sigma(M)}\right) \tag{1.69}$$

where,

$$\sigma^2(M) = \langle \delta_s^2(\vec{x};R)\rangle = \int dk \frac{k^2 P(k)}{2\pi^2} |\tilde{W}_R(k)|^2 \tag{1.70}$$

is the mass variance of the smoothed density field with the power spectrum P(k) of the density perturbations, and $\tilde{W}_R(k)$ is the $W(\vec{x};R)$ in Fourier space. The $\sigma(M)$ describes the typical size of the overdensity.

There is one caveat: since the PS formalism has a fixed filter scale R, this could under-count the halo mass fraction. Consider regions that may look under-dense on small scales, but are part of an overdense region on a bigger scale, will not be considered as part of the halo. This is called the cloud-in-cloud problem. It can be shown by taking $M \to 0$, making $\sigma(M) \to \infty$ (as the $W(\vec{x};R)$ decreases to smaller mass-scales,

$\tilde{W}_R(k)$ broadens, hence increasing $\sigma(M)$). The probability from Eq. (1.69) suggests, $P(\delta > \delta_c(t)) \to 1/2$. Suggesting that even when the fluctuations are very large (large $\sigma(M)$) only half of the total mass in the Universe collapses into virial objects. To correct for this, Press and Schechter introduced an ad hoc fudge factor of 2, assuming that the mass in the under-dense region will eventually be accreted into overdense regions, and doubling their mass. Thus, the final probability $F(> M)$ of the fraction of mass present within the collapsed objects with mass greater than M, is now $F(> M) = 2P(\delta_c(t))$.

The number density of collapsed objects within the mass range $M \to M + dM$ is,

$$\begin{aligned} n(M,t)\,dM &= \frac{\langle\rho\rangle}{M}\frac{\partial F(>M)}{\partial M}\,dM \\ &= 2\frac{\langle\rho\rangle}{M}\frac{\partial \mathcal{P}[\delta > \delta_c(t)]}{\partial M}\,dM = \sqrt{\frac{2}{\pi}}\frac{\langle\rho\rangle}{M^2}\frac{\delta_c}{\sigma}\exp\left(-\frac{\delta_c^2}{2\sigma^2}\right)\left|\frac{d\ln\sigma}{d\ln M}\right|dM \end{aligned} \tag{1.71}$$

This is known as the PS mass function. The only time-dependent term in the above equation is $\delta_c(t)$, and the mass-dependent term is $\sigma(M)$. We can then define a variable $\nu = \delta_c(t)/\sigma(M)$, thus rewriting the Eq. (1.71) as,

$$n(M,t)\,dM = \frac{\langle\rho\rangle}{M^2} f_{\rm PS}(\nu)\left|\frac{d\ln\nu}{d\ln M}\right| dM \tag{1.72}$$

where,

$$f_{ps}(\nu) = \sqrt{\frac{2}{\pi}}\nu\exp -\nu^2/2. \tag{1.73}$$

Finally, the halo mass function (HMF), which describes the number density of dark matter halos as function of their mass, is given by,

$$\boxed{\frac{dn}{dM} = \frac{\langle\rho\rangle}{M^2} f_{PS}(\nu)\left|\frac{d\ln\nu}{d\ln M}\right|.} \tag{1.74}$$

However, on the high mass end, the simplistic nature of the PS formalism (assuming a spherical collapse model) does not perform well (Peacock, 1998). We can instead use the Sheth-Tormen (ST) mass function (Sheth and Tormen, 2002), which utilizes the ellipsoidal collapse model to encapsulate the anisotropies. It is given by,

$$\boxed{\frac{dn}{dM} = A\frac{\langle\rho\rangle}{M^2}[1 + (a\nu^2)^{-p}] f_{PS}(\nu)\left|\frac{d\ln\nu}{d\ln M}\right|.} \tag{1.75}$$

The constants A, a, and p can be calibrated to match N-body simulations.

1.3.3.2 Excursion Set Theory

In semi-numerical simulations, the ES theory is usually employed to calculate the HMF. After evolving the initial random Gaussian density fields to the time t for a simulation box of size L. We gradually vary the size of the filter scale R starting from $R \approx L$. This is the primary difference between PS and ES formalism. The mass variance now is also a function of R, for the sake of clarity, we will follow the same notation as in (Mo et al., 2010) and call it S, thus,

$$S(M,R) \equiv \sigma^2(R) = \int \frac{dk}{k} \frac{k^3 P(k)}{2\pi^2} |\tilde{W}(k;R)|^2. \tag{1.76}$$

After re-smoothing the density fields at every stage of R, we check for the collapsed condition criteria $\delta(S) \geq \delta_c$. Now, if we look at a pixel at some distance $\vec{x}$, as we change R, the value for overdensity $\delta_S(\vec{x})$ gets renewed, since all the pixels had random initial overdensity, this new value is also random. If we look at the value of $\delta_S(\vec{x})$ as a function of S, it resembles a random walk. Thus, we instead look at the probability when the $\delta_S(\vec{x})$ crosses the δ_c for the first time. Redefining the Eq. (1.69) in terms of "first crossing function" $f(S)$. Where $f(S)$ is a statistical distribution function at which a random walk of $\delta_S(\vec{x})$ crosses δ_c for the first time. Mathematically, it is defined as,

$$\boxed{f(S) = \frac{1}{\sqrt{2\pi}} \frac{\delta_c}{S^{3/2}} \exp\left(-\frac{\delta_c^2}{2S}\right).} \tag{1.77}$$

The HMF can then similarly be defined by replacing $F(>M)$ with $f(S)$,

$$\boxed{\begin{aligned} \frac{dn}{dM} &= \frac{\langle\rho\rangle}{M} f(S) \left|\frac{dS}{dM}\right| \\ &= \frac{\langle\rho\rangle}{M} \frac{1}{\sqrt{2\pi}} \frac{\delta_c}{S^{3/2}} \exp\left(-\frac{\delta_c^2}{2S}\right) \left|\frac{dS}{dM}\right|. \end{aligned}} \tag{1.78}$$

Again, by using the ellipsoidal collapse correction from (Sheth and Tormen, 2002), we can write the first crossing function as:

$$f_{ST}(S) = \frac{1}{\sqrt{2\pi}} \frac{|T(S)|}{(S)^{3/2}} \exp\left[-\frac{B(S)^2}{2S}\right] \tag{1.79}$$

With

$$B(S) = \sqrt{a}\delta_c(z)[1 + \beta(a\nu)^{-\alpha}] \tag{1.80}$$

denoting the barrier of the ellipsoidal collapse. The quantity $T(S)$ is given by,

$$T(S) = \sum_{i=0}^{n} \frac{S^i}{i!} \frac{\partial^i B(S)}{\partial S^i} \tag{1.81}$$

The new HMF with the ellipsoidal correction will then be,

$$\begin{aligned} \frac{dn}{dM} &= \frac{\langle\rho\rangle}{M} f_{ST}(S) \left|\frac{dS}{dM}\right| \\ &= \frac{\langle\rho\rangle}{M} \frac{1}{\sqrt{2\pi}} \frac{|T(S)|}{(S)^{3/2}} \exp\left[-\frac{B(S)^2}{2S}\right] \left|\frac{dS}{dM}\right| \end{aligned} \tag{1.82}$$

1.3.4 Collapsed Fraction

A very important quantity in all the studies of structure formation is the collapsed fraction f_{coll} (Press and Schechter, 1974; Bond et al., 1991; Lacey and Cole, 1993; Sheth and Tormen, 1999). f_{coll} is defined as the fraction of matter that has collapsed to form bound structures within a given mass range. It also represents the probability that a randomly chosen region of the Universe will have $\delta(\vec{x}, t) > \delta_c$ at a given mass scale and time. Thus, quantifies the fraction of matter that has undergone gravitational collapse within the specified mass range. In terms of the HMF, it is given by:

$$f_{coll}(M_{min}, t) = \frac{1}{\langle\rho_m\rangle} \int_{M_{min}}^{\infty} M \frac{dn}{dM}\, dM \tag{1.83}$$

For the PS formalism, the collapsed fraction is nothing but the probability $F(> M)$:

$$f_{coll}(M_{min}, t) = \text{erfc}\left(\frac{\delta_c(t)}{\sqrt{2}\sigma(M)}\right) \tag{1.84}$$

For ES, it is given by the integral of the first crossing function as shown in the Eq. (1.77)

$$f_{\text{coll}}^{\text{EST}}(> M) = \int_{S(M)}^{\infty} f(S')\, dS' \tag{1.85}$$

Similarly, we can get the ellipsoidal corrected collapsed fraction by integrating the first cross function $f_{ST}(S)$ from the Eq. (1.79),

$$f_{\text{coll}}^{\text{ST}}(> M) = \int_{S(M)}^{\infty} f_{ST}(S')\, dS' \tag{1.86}$$

1.3.5 Conditional quantities

As explained above, the major difference between PS and ES is that the smoothing filter scale in ES is not fixed. Hence, when calculating the above quantities, i.e., first crossing function, mass function, HMF, and collapsed fraction, which were integrated over the entire volume of the simulation box, we may want to average them out based on smaller regions. Let's say we select N_{cell} number of grid cells within the large box. Let the mass of this set be M_{cell}; hence, what we would want are the conditional quantities, which are given by (Sheth and Tormen, 2002),

$$n(M_h|M_{cell};\delta_{L,0}) = \frac{\langle\rho_m\rangle}{M_h} f(S|S_{cell};\delta_{L,0}) \frac{dS}{dM_h} \tag{1.87}$$

Where $\delta_{L,0}$ is the initial density contrast of the cell linearly extrapolated to $z = 0$. The function $f(S|S_{cell})$ is similar to the one we calculated above. Hence, we can just modify all our ES quantities with the conditional ES quantities and get the following results:

- Conditional first crossing fraction:

$$f(S|S_{cell};\delta_{L,0}) = \frac{1}{\sqrt{2\pi}} \frac{\delta_c(z) - \delta_{L,0}}{(S - S_{cell})^{3/2}} \exp\left[-\frac{[\delta_c(z) - \delta_{L,0}]^2}{2(S - S_{cell})}\right] \tag{1.88}$$

- Conditional ellipsoidal first crossing fraction:

$$f(S|S_{cell};\delta_{L,0}) = \frac{1}{\sqrt{2\pi}} \frac{|T(S|S_{cell})|}{(S - S_{cell})^{3/2}} \exp\left[-\frac{[B(S) - \delta_{L,0}]^2}{2(S - S_{cell})}\right] \tag{1.89}$$

  With

$$B(S) = \sqrt{a}\delta_c(z)[1 + \beta(a\nu)^{-\alpha}] \tag{1.90}$$

$$T(S|S_{cell}) = \sum_{i=0}^{n} \frac{(S_{cell} - s)^i}{i!} \frac{\partial^i[B(S) - \delta_{L,0}]}{\partial s^i} \tag{1.91}$$

- Conditional collapsed fraction:

$$f_{coll}(M_{min}|S_{cell};\delta_{L,0}) = \int_{S_{cell}}^{s_{min}} ds f(S|S_{cell};\delta_{L,0}) \tag{1.92}$$

  where $s_{min} = \sigma^2(M_{min})$.

1.3.6 Virial Objects

In the earlier section, we defined the virial theorem, Eq. (1.65), and showed that for the spherical collapse, the collapsing body will reach a quasi-static equilibrium to form a stable structure, see §1.3.2.2. The critical density Δ_{vir} required for such an object to form in a matter-dominated Universe was $\Delta_{vir} \approx 18\pi^2$. While, the modified critical overdensity for $\Omega_m + \Omega_\Lambda = 1$ from the Eq. (1.67) (Bryan and Norman, 1998) is given by,

$$\Delta_{vir} \approx 18\pi^2 + 82x - 39x^2 \tag{1.93}$$

with $x = \Omega_m^z - 1$ evaluated at redshift z, and Ω_m^z is given by,

$$\Omega_m^z = \frac{\Omega_m(1+z)^3}{\Omega_m(1+z)^3 + \Omega_\Lambda + \Omega_k(1+z)^2}. \tag{1.94}$$

Now, within these viral objects, the baryonic gas can further collapse to form galaxies and galactic clusters. Unlike dark matter, the baryonic perturbations are affected by the pressure term along with the gravitational interaction. As we saw in §1.3.2.2.

From the Eq. (1.66), for matter dominated Universe, the critical density $\Delta_{vir} = 18\pi^2 \gg 1$, the density relation gives us,

$$\Delta_{vir} = \frac{\rho_{t_{vir}}}{\langle \rho_{t_{vir}} \rangle} \tag{1.95}$$

where, $\langle \rho_{t_{vir}} \rangle$ is the background density. Which we can assume to be critical density ρ_c (see §1.3.2.2). Thus for a halo of mass M and radius r_{vir}, and redefining $\rho_{t_{vir}} \equiv \rho_{halo}$, we can rewrite the above equation as,

$$\rho_{halo} = \Delta_c \cdot \rho_c = \frac{3M}{4\pi r_{vir}^3} \tag{1.96}$$

Rewriting the above equation fro r_{vir},

$$r_{vir}^3 = \frac{3M}{4\pi \Delta_c \cdot \rho_c} \tag{1.97}$$

Now substituting for $\rho_c(z) = \frac{3H_0^3}{8\pi G}[\Omega_m(1+z)^3 + \Omega_\Lambda + \Omega_k(1+z)^2] = \frac{3H_0^3}{8\pi G}\frac{\Omega_m(1+z)^3}{\Omega_m^z}$ in the above equation, we get

$$r_{vir} \propto M \left[\frac{\Omega_m}{\Omega_m^z}\Delta_c\right]^{-1/3} (1+z)^{-1} \tag{1.98}$$

This can finally be written in physical units as (Barkana and Loeb, 2001),

$$r_{vir} = 0.784 \left(\frac{M}{10^8\, h^{-1}\, M_\odot}\right)^{1/3} \left(\frac{\Omega_m}{\Omega_z}\frac{\Delta_c}{18\pi^2}\right)^{-1/3} \left(\frac{1+z}{10}\right)^{-1} h^{-1}\,\mathrm{kpc}. \tag{1.99}$$

Now equating the kinetic energy of the particles $K = \frac{3}{2}\frac{M}{\mu m_p} k_B T_{\rm vir}$ with the total internal energy $U = -3GM^2/5r_{vir}$, according to the virial theorem, we get the following result for the T_{vir} (Barkana and Loeb, 2001),

$$T_{\rm vir} = \frac{\mu m_p}{5k_B}\frac{GM}{R} = 1.98 \times 10^4 \left(\frac{\mu}{0.6}\right) \left(\frac{M}{10^8\, h^{-1}\, M_\odot}\right)^{2/3} \left(\frac{\Omega_m}{\Omega_z}\frac{\Delta_c}{18\pi^2}\right)^{1/3} \left(\frac{1+z}{10}\right) \mathrm{K}. \tag{1.100}$$

Inverting the above equation to get the mass M as a function of viral temperature T_{vir},

$$M(T_{vir}, z) = 10^8 h^{-1} M_\odot \left(\frac{T_{vir}}{1.98 \times 10^4 K}\right)^{3/2} \left(\frac{\mu}{0.6}\right)^{-3/2} \left[\left(\frac{\Omega_m}{\Omega_m^z}\right)\left(\frac{\Delta_c}{18\pi^2}\right)\right]^{-1/2} \left(\frac{1+z}{10}\right)^{-3/2}. \tag{1.101}$$

1.4 Reionization

Soon (in cosmological sense) after the halos have formed and the gas within these halos cooled down to form stars, these stars started to radiate energetic photons ($\mathrm{E_{ph}} \geq 13.6\mathrm{eV}$) which eventually ionized the HI in the ambient medium. These photons could travel further beyond the host halos and ionize the IGM, which was predominantly neutral. Thus, embarking on the era of reionization. During this era, the Universe went through a phase transition from being predominantly neutral to ionized. In this section, we will work on the physics of reionization and derive the equations that will be useful for comprehending reionization as a whole, along with this thesis. *The derivations in this section are inspired from the discussions in Tirthankar Roy Choudhury, 2022*

1.4.1 The Gas Cools Down

In the previous §1.3.2.2, we have looked into the idea of how some tiny perturbations in the matter density of the Universe could lead to the formation of virialized structures,

like halos. Because of their deep potential wells, these gravitationally bound dense halos can attract the baryons to form dense gas clouds. Once attracted, the gas particles will subsequently settle into the virial equilibrium state as defined in §1.3.6. The gas particles in the halos will move around with the circular velocity $\sim v_c \equiv \sqrt{\frac{GM}{R_{\rm vir}}}$. Due to which they will heat to the temperatures $\sim T_{\rm vir}$. This process is known as shock heating[6].

For a halo at $z = 10$, and assuming $\Delta_{\rm vir} = 18\pi^2$ (1.67) for a matter dominated Universe, density of the gas within a virialized halo is $\rho_b \sim 10^{-25}$gm cm^{-3} (see equations: Eq. (1.99), Eq. (1.100), and Eq. (1.101)). The number density for this case (assuming $\mu \sim 1$) $n_b \sim 6 \times 10^{-2}$cm^{-3}. Which is insufficient for the star formation. Hence, the gas must first cool down and condense for the stars to form. Cooling of the gas at high redshift usually occurs via collisional excitation of atoms, followed by radiative cooling by photon emission. Basically, converting the kinetic energy of photons into the radiation that can escape the halo. Thus reducing the total energy of the system and, hence, cooling it. As shown in Padmanabhan, 2000 for a gas composed primarily of hydrogen and helium, the collisional excitation of HI and helium shows peaks around 2×10^4 K and 9×10^4 K. But this process drops rapidly for the gas temperature $T \lesssim 10^4$ K. Thus, for $T_{\rm vir} \lesssim 10^4$ K, the gas will not be able to cool down and hence the minimum mass required for the halos to form stars at high redshift $\sim 10^8 M_\odot$[7], see Eq. (1.101).

Thus to conclude, in order for the gas to cooldown within these halos, there exists a minimum threshold on the mass of the halos that can support star formation, $\rm M_{min}$. The value of the $\rm M_{min}$ depends upon the composition of the gas within the halos. In §1.6 we demonstrate how we apply this threshold in our simulations.

1.4.2 And the Stars are born

The cooling of gas leads to a decrease in pressure, due to which the gas starts to collapse under the gravitational pressure. Since the Jeans mass ($M_j \propto T^{3/2}$) is proportional to the temperature, this makes smaller masses collapse under gravitational pressure. When the gas collapses, the density increases, further reducing the Jeans mass $M_j \propto \rho^{-1/2}$ and

[6]The Temperature and mass relation for a viral object are shown in the equation in Eq. (1.101).

[7]This minimum mass criteria is valid only for high redshift where the Universe is primarily made of hydrogen and helium. At low redshifts, where heavy metals are present, the cooling can also occur at lower temperatures. Furthermore, the other way cooling can occur at low temperatures is through the rotational and vibrational levels of molecules like H_2 and CO. The presence of H_2 can lower the cooling temperature to $T_{\rm vir} \sim 500$ K, corresponding to halo masses $M \sim 10^6 M_\odot$. These halos are known as mini halos and can be important sources of stars at high redshifts. But, given the complexity of forming H_2 and the formation of the stars leading to suppression in molecular cooling due to the feedback process, makes a good case to ignore them for the time being.

causing further collapse. Consequently, the large clouds start fragmenting into smaller size clumps. This eventually forms stars depending on when and how the fragmentation process stops (Benedetta Ciardi and Ferrara, 2005; Dayal and Ferrara, 2018).

The collapsing gas can form stars of mass $\mathrm{M_{star}}$ under the probability distribution given by the Initial Mass Function (IMF). The IMF gives the relative number of stars within a given mass range. At high redshifts, due to poor cooling in the absence of metals, the fragmentation stopped early, forming very massive stars (Bromm, 2013), called Population III or PopIII stars. Due to their heavy mass, these stars were short-lived, but were a great source of ionizing photons. Eventually, with the metal enrichment of the medium, facilitating the cooling mechanism, the next generation of stars, Population II or PopII stars, could form. The IMF of popII stars is well distributed with a significant contribution coming from $\mathrm{M_{star}} \sim M_\odot$ (Mo et al., 2010). Let us now try to derive some results to calculate the ionizing efficiency of stars within a halo.

Let us assume that a fraction $\mathrm{f_{star}}$ of baryonic gas formed stars. Assuming a very simplistic model, we can define the total stellar mass within the halo as,

$$\mathrm{M_{star}} = \mathrm{f_{star}}(\mathrm{M_h}, t) \left(\frac{\Omega_b}{\Omega_m}\right) \mathrm{M_h}, \tag{1.102}$$

where, $\mathrm{M_h}$ is the halo mass. We can then define the star formation rate as,

$$\dot{\mathrm{M}}_{\mathrm{star}}(\mathrm{M}, t, t_{\mathrm{form}}) = \mathrm{f_{star}}(\mathrm{M_h}, t_{\mathrm{form}}) \left(\frac{\Omega_b}{\Omega_m}\right) \mathrm{M_h}\ \Lambda_{\mathrm{star}}(t - t_{\mathrm{form}}; \mathrm{M_h}, t_{\mathrm{form}}), \tag{1.103}$$

where, t_{form} is the galaxy formation time and $\Lambda_{\mathrm{star}}(t - t_{\mathrm{form}}; \mathrm{M_h}, t_{\mathrm{form}})$ is the star formation profile.

For continuous star formation, we can assume $\Lambda_{\mathrm{star}}(t; \mathrm{M_h}, t_{\mathrm{form}})$ to be constant. While the short burst of stars we can write $\Lambda_{\mathrm{star}}(t; \mathrm{M_h}, t_{\mathrm{form}}) = \delta_D(t)$.

1.4.3 They Started Radiating

Next, we would like to compute the ionizing radiation from stars, which is the major driver for the reionization. We do so by using population synthesis models. According to the population synthesis model, the O-type stars with mass $\sim 15 - 90 M_\odot$, are the major contributors to the ionizing photons budget due to their high surface temperatures. Due to their high mass, these stars are short lived ($\sim 10^6$ yrs) hence their profiles can be approximated with $\Lambda_{\mathrm{star}}(t; \mathrm{M_h}, t_{\mathrm{form}}) = \delta_D(t)$. Lastly, the contribution to the ionizing photons is much higher for these low metallicity stars due to inefficient cooling.

With this in mind, we can write the total luminosity of a halo of mass $\mathrm{M_h}$ as,

$$\begin{aligned} L_\nu(\mathrm{M_h}, t, t_{\rm form}) &= \int_{t_{\rm form}}^{\infty} dt'\, \dot{\mathrm{M}}_{\rm star}(\mathrm{M_h}, t', t_{\rm form})\, l_\nu(t - t'; \mathrm{M_h}, t_{\rm form}) \\ &= \mathrm{f_{star}}(\mathrm{M_h}, t_{\rm form}) \left(\frac{\Omega_b}{\Omega_m}\right) M \int_{t_{\rm form}}^{\infty} dt'\, \Lambda_{\rm star}(t' - t_{\rm form})\, l_\nu(t - t'; \mathrm{M_h}, t_{\rm form}) \end{aligned} \tag{1.104}$$

Where, $l_\nu(t; \mathrm{M_h}, t_{\rm form})$ is the specific luminosity, and is defined as the luminosity per unit mass of stars formed for a burst of star formation at $t = 0$ inside a halo of mass $\mathrm{M_h}$, and formation time $t_{\rm form}$. Using the total luminosity, we can define the ionizing radiation, obtained by dividing the luminosity by the photon energy $h\nu$ and integrating it above $\nu = \nu_{\rm ion}$, where $\nu_{\rm ion}$ is the minimum ionizing photon frequency.

$$\begin{aligned} \dot{N}_\gamma(\mathrm{M_h}, t, t_{\rm form}) &= \int_{\nu_{\rm ion}}^{\infty} d\nu\, \frac{L_\nu(\mathrm{M_h}, t, t_{\rm form})}{h\nu} \\ &= \mathrm{f_{star}}(\mathrm{M_h}, t_{\rm form}) \left(\frac{\Omega_b}{\Omega_m}\right) M \int_{t_{\rm form}}^{\infty} dt'\, \Lambda_{\rm star}(t' - t_{\rm form}) \\ &\quad \times \int_{\nu_{\rm ion}}^{\infty} d\nu\, \frac{l_\nu(t - t'; \mathrm{M_h}, t_{\rm form})}{h\nu} \end{aligned} \tag{1.105}$$

Now, assuming the lifespan of stars is much shorter than the age of the Universe, we can approximate the above integral as,

$$\dot{N}_\gamma(\mathrm{M_h}, t, t_{\rm form}) = \mathrm{f_{star}}(\mathrm{M_h}, t_{\rm form})\; N_{\gamma/b}(\mathrm{M_h}, t_{\rm form}) \left(\frac{\Omega_b}{\Omega_m}\right) M\; \Lambda_{\rm star}(t - t_{\rm form}). \tag{1.106}$$

where, $N_{\gamma/b}$ is the total ionizing photons per unit mass of the stars, produced by a halo and is defined as,

$$N_{\gamma/b}(\mathrm{M_h}, t_{\rm form}) = \int_0^{\infty} dt\; \int_{\nu_{\rm ion}}^{\infty} d\nu\; \frac{l_\nu(t; \mathrm{M_h}, t_{\rm form})}{h\nu} \tag{1.107}$$

Thus writing, the integral over l_ν, as;

$$\int_{\nu_{\rm ion}}^{\infty} d\nu\; \frac{l_\nu(t; \mathrm{M_h}, t_{\rm form})}{h\nu} = N_{\gamma/b}(\mathrm{M_h}, t_{\rm form})\; \delta_D(t), \tag{1.108}$$

and integrating over t to get the equation showed in Eq. (1.106).

1.4.4 Escaping Photons

The ionizing photons produced by the stars within the halo must escape to the IGM to ionize it. Not all the photons produced may escape into the IGM. Hence, we will instead have some fraction of the photons escaping, as only these photons participate in reionization. We quantify this with a parameter called escape fraction, $\mathrm{f_{esc}}$. Thus modifying the equation eq:Ndot with the $\mathrm{f_{esc}}$ to get,

$$\dot{N}_\gamma(\mathrm{M_h}, t, t_{\mathrm{form}}) = \mathrm{f_{esc}}(\mathrm{M_h}, t_{\mathrm{form}})\ \mathrm{f_{star}}(\mathrm{M_h}, t_{\mathrm{form}})\ N_{\gamma/b}(\mathrm{M_h}, t_{\mathrm{form}}) \left(\frac{\Omega_b}{\Omega_m}\right) M\ \Lambda_{\mathrm{star}}(t - t_{\mathrm{form}}). \tag{1.109}$$

With this relation, we can calculate the ionizing photon rate as a function of halo mass $\mathrm{M_h}$. We can simplify it further by assuming the short burst model for the $\Lambda_{\mathrm{star}}(t - t_{\mathrm{form}})$, i.e., the duration of star formation is much smaller than the age of the Universe, and substituting $\Lambda_{\mathrm{star}}(t - t_{\mathrm{form}}) \approx \delta_D(t - t_{\mathrm{form}})$, we get

$$\dot{N}_\gamma(\mathrm{M_h}, t, t_{\mathrm{form}}) = \mathrm{f_{esc}}(\mathrm{M_h}, t_{\mathrm{form}})\ \mathrm{f_{star}}(\mathrm{M_h}, t_{\mathrm{form}})\ N_{\gamma/b}(\mathrm{M_h}, t_{\mathrm{form}}) \left(\frac{\Omega_b}{\Omega_m}\right) M\ \delta_D(t - t_{\mathrm{form}}). \tag{1.110}$$

We can now integrate the above equation for $t \geq t_{\mathrm{form}}$ to get the total ionizing photons. Since the star formation and starburst happen over a very short period, we get the following form for the total amount of escaping photons:

$$N_\gamma(\mathrm{M_h}, t) = \mathrm{f_{esc}}(\mathrm{M_h}, t)\ \mathrm{f_{star}}(\mathrm{M_h}, t)\ N_{\gamma/b}(\mathrm{M_h}, t) \left(\frac{\Omega_b}{\Omega_m}\right) M. \tag{1.111}$$

N_γ turns out to be the product of the all the three unknowns; $\mathrm{f_{esc}}$, $\mathrm{f_{star}}$ and $N_{\gamma/b}$, making these parameters highly degenrate with each other. Thus, instead of modelling them individually, we can instead look at their combination and define it as the new and only reionization parameter. This parameter is called the reionization efficiency parameter (ζ) and is defined as follows,

$$\zeta(\mathrm{M_h}, t) = \frac{\mathrm{f_{esc}}(\mathrm{M_h}, t)\ \mathrm{f_{star}}(\mathrm{M_h}, t)\ N_{\gamma/b}(\mathrm{M_h}, t)\ \mathrm{m_p}}{1 - Y}, \tag{1.112}$$

where Y is helium weight fraction (usually taken to be 0.24). Thus simplifying N_γ to

$$N_\gamma(\mathrm{M_h}, t) = \zeta(\mathrm{M_h}, t)\ (1 - Y) \left(\frac{\Omega_b}{\Omega_m}\right) \frac{M}{m_p}. \tag{1.113}$$

The parameter ζ is usually treated as a free parameter in reionization models. In Chapter 2, we will return to ζ again and use a similar form as described above, with some slight modifications, and show how we calculate this term in our simulations. Nonetheless, the physical intuition and the meaning of this parameter remain the same.

We can now integrate the equation for the total number of photons escaping from a halo, Eq. (1.105), with the HMF from Eq. (1.74), Eq. (1.78) or Eq. (1.75) to get the mean number of ionizing photons from all the halos that can support star formation (halos with mass $\mathrm{M_h} \geq \mathrm{M_{min}}$) per unit comoving volume, which is given by,

$$\begin{aligned}\bar{n}_\gamma(t) &= \int_{M_{\rm min}}^{\infty} dM\, N_\gamma(\mathrm{M_h}, t)\, n(\mathrm{M_h}, t) \\ &= \frac{1-Y}{m_p}\left(\frac{\Omega_b}{\Omega_m}\right)\int_{M_{\rm min}}^{\infty} dM\, \zeta(\mathrm{M_h}, t)\, \mathrm{M_h}\, n(\mathrm{M_h}, t)\end{aligned} \tag{1.114}$$

Assuming ζ to be independent of $\mathrm{M_h}$, then

$$\begin{aligned}\bar{n}_\gamma(t) &= \frac{1-Y}{m_p}\left(\frac{\Omega_b}{\Omega_m}\right)\zeta(t)\int_{M_{\rm min}}^{\infty} dM\, \mathrm{M_h}\, n(\mathrm{M_h}, t) \\ &= \frac{1-Y}{m_p}\left(\frac{\Omega_b}{\Omega_m}\right)\zeta(t)\ \langle\rho\rangle_m\, f_{\rm coll}(\mathrm{M_{hmin}}, t) \\ &= \bar{n}_H\, \zeta(t)\, f_{\rm coll}(\mathrm{M_{hmin}}, t),\end{aligned} \tag{1.115}$$

Where $\langle\rho\rangle_m$ is the mean comoving matter density, and $f_{\rm coll}$ is the collapsed fraction described in the Eq. (1.83). The mean hydrogen number density for the comoving volume is then written as,

$$\bar{n}_H = \frac{\langle\rho\rangle_H}{m_p} = \frac{(1-Y)\Omega_b}{\Omega_m}\frac{\langle\rho\rangle_m}{m_p}. \tag{1.116}$$

Thus, from the last term of the Eq. (1.115), under all the approximations made so far, the ratio of mean ionizing photon to hydrogen density, or the mean ionizing ($\bar{Q}_{\rm HII}$)/ neutral fraction ($\bar{\mathrm{x}}_{\rm HI}$), assuming every escaping ionizing photon ionizes a hydrogen atom, is given by the product of ionizing photon efficiency and the collapsed fraction, $\zeta\ f_{\rm coll}$. This is a very remarkable and valuable result. We will use this result to calculate the ionized pixels within our simulations as well.

Accounting for the halo mass $\mathrm{M_h}$ dependency of ζ we get the similar results,

$$\begin{aligned}\langle\zeta\ f_{\rm coll}(\mathrm{M_{hmin}}, t)\rangle &= \frac{\bar{n}_\gamma(t)}{\bar{n}_H} \\ &= \frac{1}{\langle\rho\rangle_m}\int_{M_{\rm min}}^{\infty} dM\ \zeta(\mathrm{M_h}, t)\ \mathrm{M_h}\ \frac{dn(\mathrm{M_h}, t)}{dM}.\end{aligned} \tag{1.117}$$

Finally, we can define the global mean emissivity of ionizing photons by differentiating the above quantity with respect to time, i.e.,

$$\epsilon(t) = \dot{\bar{n}}_{\gamma}(t) = \bar{n}_H \frac{d\langle \zeta \ f_{\rm coll}(\mathrm{M_{hmin}}, t)\rangle}{dt}. \tag{1.118}$$

The emissivity can be understood as the average number of photons per unit time and per unit comoving volume available for hydrogen reionization. It is one of the most important quantities in modeling reionization.

1.4.4.1 Feedback Process

The feedback processes act as an intermediary between the IGM and the galaxies. These processes can quench the production of ionizing photons within the galaxies and also affect the growth of the galaxy. In this section, we will very briefly describe these process; for greater detail, one can refer to (Benedetta Ciardi and Ferrara, 2005; Dayal and Ferrara, 2018)

- Chemical feedback: This refers to the change in metallicity of the gas composition within the galaxy. When the first heavy stars die (via supernovae), they dump heavier metals formed in their core into the near interstellar medium. This increases the metallicity of the next generation of stars, suppressing ionizing photon production.
- Mechanical feedback: The supernovae can also heat the interstellar medium, thus suppressing the star formation. In some cases, they can eject the gas into the IGM. The simplest way to model it is via M_{min}.
- Radiation feedback: If the energy of an ionizing photon $E_{ph} > 13.6$ eV, the excess energy goes into the electron's kinetic energy, thus raising the temperature of the IGM. This heating can prevent the cooling of gas in lighter halos. It also lowers the accretion rate of the gas within, thus halting the growth of the galaxy.

1.4.5 Ionizing the Universe

Here we will take a big leap in the continuity. When the photons escape to the IGM and start ionizing the HI gas and possibly changing the thermal state of the IGM, one should begin with the radiative transfer equations for a complete understanding of the

reionization process. Unfortunately, the discussion of radiative transfer is beyond the scope of this thesis, and instead of deriving the necessary results, we will quote the results directly where they seem fit. The readers are advised to refer to the following texts for self-study (Padmanabhan, 2000; Mo et al., 2010; Peacock, 1998; Narlikar, 2002; Tirthankar Roy Choudhury, 2022). Nonetheless, the omission of the radiative transfer section will not hamper the understanding of the work carried out in this thesis, and it is left up to the reader.

We begin by assuming a simplistic form of the rate of ionized fraction within an ionized region surrounded by an optically thick ambient medium (such that all the radiation is absorbed locally, which is a fair approximation for the Universe near the EoR with high HI, called local absorption),

$$\frac{dQ_{\rm HII}}{dt} = \frac{\dot{\bar{n}}_\gamma}{\bar{n}^{\rm II}_{\rm HII}} - Q_{\rm HII}\,\frac{\bar{n}^{\rm II}_{\rm HI}}{\bar{n}^{\rm II}_{\rm HII}}\Gamma^{\rm II}_{\rm HI}. \tag{1.119}$$

Here $\bar{n}^{\rm II}_{\rm HII}$ is the average ionized hydrogen number density, $\bar{n}^{\rm II}_{\rm HI}$ is the average HI number density, and $\Gamma^{\rm II}_{\rm HI}$ is the hydrogen photoionization rate, in the ionized region. The superscript II represents quantities in the ionized region. $\dot{\bar{n}}_\gamma$ is the emissivity and $\bar{n}^{\rm II}_{\rm HII}$ is the average ionized hydrogen number density. The $\bar{n}^{\rm II}_{\rm HII}$ can be present in the ionized regions in the form of high-density clumps of gas, which can remain neutral even after absorbing all the ionizing photons. These self-shielded objects can be seen in the quasar spectra as Lyman Limit Systems (LLS); we will study their effects on the quasar spectra in Chapter 4 as well. These objects usually have very high recombination rates.

The hydrogen photoionization rate in the ionized region depends upon the ionizing source and is given by,

$$\Gamma^{\rm II}_{\rm HI} \equiv 4\pi \int_{\nu_{\rm ion}}^{\infty} \frac{d\nu}{h\nu}\sigma_H(\nu)\bar{J}^{\rm II}_\nu = 4\pi \int_{\nu_{\rm ion}}^{\infty} \frac{d\nu}{h\nu}\sigma_H(\nu)\left(\frac{\bar{\tilde{J}}^{\rm II}_\nu}{a^3}\right), \tag{1.120}$$

where $\bar{\tilde{J}}^{\rm II}_\nu$ is the average comoving intensity/ source in the ionized region, and $\sigma_H(\nu)$ is the absorption cross-section of the photoionization rate. For our initial assumption of local absorption, one can show that the mean intensity $\bar{J}^{\rm II}_\nu$ is homogeneous within the ionized region.

One can picture the Eq. (1.119) as the amount of ionizing photons produced within the ionized region can either increase the total number of ionized fraction (the term on the left-hand side) or can be absorbed by the neutral gas within the ionized region (the second term on right-hand side). We have already studied the first term on the

right-hand side of the Eq. (1.119). However, to study the second term, we need some information about the residual HI density $\bar{n}_{\rm HI}^{\rm II}$. We begin by writing the evolution equation of HI density within the IGM as,

$$\frac{dn_{\rm HI}(t,\mathbf{x})}{dt} = -\Gamma_{\rm HI}(t,\mathbf{x})n_{\rm HI}(t,\mathbf{x}) + a^{-3}\ \alpha_R(T(t,\mathbf{x}))n_{\rm HII}(t,\mathbf{x})n_e(t,\mathbf{x}), \tag{1.121}$$

The global $n_{\rm HI}$ density will decrease as more and more HI is ionized. The first term on the right-hand side of the above equation represents the decrease in $n_{\rm HI}$ due to photoionization of the gas[8]. The second term on the right-hand side of the above equation represents the increase in the $n_{\rm HI}$ due to the recombination of HII with the free electrons. The $n_e(t,\mathbf{x})$ thus represents the free electron density, $\alpha_R(T)$ is the temperature-dependent recombination rate, and a^3 is the factor to convert the densities from comoving to proper units. Although, the above equation is for any point in the IGM, hence we dropped the superscript II, but solving it for the HII regions is more relevant for the study of reionization.

Before we move further, we want to draw some connection between $\Gamma_{\rm HI}$, the photoionization rate, and the recombination time scale t_{rec}, the average timescale in which the free electron and HII recombine to form HI, at the photoionization equilibrium. In general $\Gamma_{\rm HI} \propto t_{rec}^{-1}$, during photoionization equilibrium and is expressed as, $\Gamma_{\rm HI} \sim \frac{1}{t_{rec}}\left(\frac{\rm n_e\ n_{HII}}{\rm n_{HI}}\right)$. For the time scales we are interested in ($\gg t_{rec}$ or $(\Gamma_{\rm HI}^{-1})$), the system settles down into the quasi-equilibrium state, known as the photoionization equilibrium. In this case, the left-hand side of the $dn_{\rm HI}(t,\mathbf{x})/dt$ term is zero and the solution to the Eq. (1.121) is given by,

$$\Gamma_{\rm HI}(t,\mathbf{x})n_{\rm HI}(t,\mathbf{x}) \approx \alpha_R(T(t,\mathbf{x}))n_{\rm HII}(t,\mathbf{x})n_e(t,\mathbf{x}). \tag{1.122}$$

The approximation of very short $\Gamma_{\rm HI}^{-1}$ at high redshift is a tricky subject. Especially near the self-shielded regions or the edges of the ionized bubbles, the ionizing photons are completely absorbed, causing the transmitted ionizing radiation to weaken. The $\Gamma_{\rm HI}$ in these cases be very small (Kušmić et al., 2022). Hence, one needs to be careful when using this approximation. Nevertheless, we are safe to make this approximation for the case where we are interested in the average properties of the ionized regions. Given that the photoionization equilibrium condition is valid, and given that the local

[8]We use the terms gas and hydrogen interchangeably, although even at high redshift, there could be other elements like helium. However, for the course of this thesis, we are only concerned with hydrogen reionization.

absorption approximation allows us to assume homogeneous mean intensity within the ionized regions, we can integrate the Eq. (1.122) and normalize with the total ionized region column $V_{\rm HII}$ to get,

$$\Gamma^{\rm II}_{\rm HI} \int_{V_{\rm HII}} \frac{d^3x}{V_{\rm HII}} n_{\rm HI}(t,\mathbf{x}) = a^{-3}\ \alpha_R(T) \int_{V_{\rm HII}} \frac{d^3x}{V_{\rm HII}} n_{\rm HII}(t,\mathbf{x}) n_e(t,\mathbf{x}). \tag{1.123}$$

The above equation can further be rewritten in terms of a new quantity called the clumping factor,

$$\Gamma^{\rm II}_{\rm HI}\ \bar{n}^{\rm II}_{\rm HI} = a^{-3}\ \alpha_R(T)\ \mathcal{C}\ \bar{n}^{\rm II}_{\rm HII}\ \bar{n}^{\rm II}_e, \tag{1.124}$$

The clumping factor characterizes the fluctuations in the ionized regions of the IGM. It is defined as follows,

$$\begin{aligned}\mathcal{C} \equiv & \frac{\int_{V_{\rm HII}} (d^3x/V_{\rm HII})\, n_{\rm HII}(t,\mathbf{x})\, n_e(t,\mathbf{x})}{\left[\int_{V_{\rm HII}} (d^3x/V_{\rm HII})\, n_{\rm HII}(t,\mathbf{x})\right]\left[\int_{V_{\rm HII}} (d^3x/V_{\rm HII})\, n_e(t,\mathbf{x})\right]} \\ = & \frac{\overline{n_{\rm HII}\, n_e}^{\rm II}}{\bar{n}^{\rm II}_{\rm HII}\, \bar{n}^{\rm II}_e}\end{aligned} \tag{1.125}$$

The clumping factor arises due to the density^2 dependency of the recombination term. To evaluate the $\mathcal{C}$, we must know both the density structure of the IGM and how the ionization fronts propagate in the high-density regions. This is not so easy to model. One should note that, during the calculations of $\mathcal{C}$, we assumed uniform temperature across the ionized region (isothermal case) and took the $\alpha_R(T)$ out of the integral. In principle, there could be a temperature gradient within the ionized region, but usually the temperature dependence of $\alpha_R(T)$ is relatively weak $\sim T^{-0.7}$, hence we can ignore that part.

Now, substituting the result from the Eq. (1.124) to the Eq. (1.119), we can write the rate of change of $Q_{\rm HII}$ as,

$$\frac{dQ_{\rm HII}}{dt} = \frac{\dot{n}_\gamma}{n^{\rm II}_{\rm HII}} - Q_{\rm HII} a^{-3}\ \mathcal{C}\alpha_R(T) n^{\rm II}_e \tag{1.126}$$

Finally, we can make more substitutions from the properties of the ionized regions. We can write $n^{\rm II}_{\rm HII} \approx \bar{n}_H$ (assuming almost all the hydrogen atoms have ionized), and $n^{\rm II}_e = \chi_{\rm He} \times n^{\rm II}_{\rm HII} \approx \chi_{\rm He} \times \bar{n}_H$, where $\chi_{\rm He}$ accounts for the additional contribution of free electrons from helium[9]. Lastly, substituting for the $\dot{n}_\gamma$ from the Eq. (1.118), we get

[9]let Y be the mass fraction of helium atoms, then the fraction of electrons present in helium =

$$\frac{dQ_{\rm HII}}{dt} = \frac{d\langle \zeta f_{\rm coll} \rangle}{dt} - Q_{\rm HII}\ \chi_{\rm He}\ a^{-3}\mathcal{C}\alpha_R(T)n_H. \tag{1.127}$$

This equation can be solved for the ζ and $\mathcal{C}$ by looking at the reionization history of the Universe. If we keep every other parameter fixed and only increase the value of ζ, increasing the ionizing photon budget in the Universe, we end up with a Universe that reionizes pretty early, also referred to as the early reionization model. Whereas if we keep everything fixed and only vary $\mathcal{C}$, hence increasing the clumpiness, either the density or the amount of dense HI systems, which end up increasing the overall recombination rate of the Universe. This slows down the reionization process, giving us late reionization models.

Clarification

Even though we calculated the ionization efficiency using the stars, the equations and working principle of our prescription of reionization models remain the same even if we substitute stars with galaxies. What we actually calculated was the ionizing photon budgets from all the bright stars in a galaxy. Furthermore, the structure of galaxies and stars does not matter for our modeling, which uses semi-numerical techniques. What matters is correctly associating ζ with the halo mass. Since ζ is usually derived using N-body simulations and later used in our models, we don't have to worry about the complicated modeling of galaxies. This is also one of the reasons, 21cmFAST can get away by modeling the density fields via DM halos.

$Y/4$ (or $2Y/4$), given the ionization state of the helium. If n_H is the number density of hydrogen, then the number density of helium is $= Y/(1-Y) \times n_H$. Thus, the total number of electrons coming from the contribution of both hydrogen and helium $\chi_{\rm He} = (1 + Y/4(1-Y))n_H$. For singly ionized helium with $Y = 0.24$, $\chi_{\rm He} \approx 1.08$, while for doubly ionized helium $\chi_{\rm He} \approx 1.16$.

Summary

After the formation of virialized objects such as dark matter halos, the baryonic gas within these halos (given that $M_{halo} > M_{min}$ to facilitate the cooling) can further collapse by fragmenting into clumps of much denser objects, such as stars. In this section, we showed how the first generation of massive stars (Pop III), capable of producing ionizing photons, will drive the reionization in the IGM. We mapped the contribution of each halo to the reionization budget by calculating the reionization efficiency parameter ζ given by Eq. (1.112). Finally, in the Eq. (1.127), we showed how the ionized bubbles carved by these ionizing photons grow.

1.5 Observing the Reionization

In the previous section, we developed an understanding of how the neutral gas in the IGM changes during the evolution of reionization. We will now see how we can actually measure/ calculate the various quantities mentioned in §1.4. There are a few ways to study the reionization topology observationally, but for the sake of this thesis, we will mention only a handful of them.

1.5.1 CMB

As described in §1.2, after the matter decoupled from the radiation, the photons scattered from the surface of last scattering, called CMB photons. These photons started to stream freely throughout the Universe. During reionization, CMB photons, when scattered off by the free electrons present in the IGM via Thomson scattering, polarize the CMB power spectrum. The CMB optical depth of Thomson scattering relates to the ionized fraction in the following way,

$$\tau_{\rm el} = c\ \sigma_T \int_{t_0}^{t_{\rm LS}} dt\ \bar{n}_e(t)\ a^{-3}(t) = c\ \sigma_T\ \bar{n}_H \int_{t_0}^{t_{\rm LS}} dt\ \chi_{\rm He}(t)\ Q_{\rm HII}(t)\ a^{-3}, \tag{1.128}$$

where, $\bar{n}_e = \chi_{\rm He} Q_{\rm HII}\ \bar{n}_H$ is the free electron number density averaged over all volume, σ_T is the cross-section for Thomson scattering, and $t_{\rm LS}$ is the cosmic time during the surface of last scattering. Hence, by measuring $\tau_{\rm el}$ we can probe the integrated reionization history. To calculate $\tau_{\rm el}$, we study the large-scale polarization produced by the Thomson scattering on the CMB angular power spectrum.

1.5.2 21cm Line

The ground state of a HI atom consists of two hyperfine states (i.e., the triplet when the spin states of the proton and electron are parallel, and the singlet when the spin states are antiparallel), with the antiparallel alignment typically representing a slightly lower energy state than the parallel alignment. At the thermal equilibrium, the spins of electrons and the protons in the HI tend to be antiparallel due to the influence of the electromagnetic interaction. However, quantum mechanics dictates that there is a probability for the spins to be in either parallel or antiparallel configurations. Due to this probabilistic nature, if the atom transitions from triplet to singlet, it produces a photon having a rest wavelength of 21 cm or a rest frequency of 1420 MHz.

This signal is directly proportional to the HI density. The effects of the 21 cm signal can be seen against the CMB by comparing the spin temperature T_s (which is directly coupled to the gas temperature) against the CMB temperature, T_γ. If $T_s > T_\gamma$, then the 21 cm signal can be seen on the emission spectra, whereas if $T_s < T_\gamma$, then the 21 cm signal can be seen on the absorption spectra (Bowman et al., 2018; Singh et al., 2022).

This signal can provide a complementary view of our Universe by directly studying the HI distribution. The globally averaged signal carries the global signature of thermal and ionization histories in the Universe. There are multiple ongoing efforts worldwide to detect it, namely, Low-Frequency Array (LOFAR) (B. Gehlot et al., 2019; Florent G Mertens et al., 2020), Murchison Widefield Array (MWA) (Barry et al., 2019; Cathryn M Trott et al., 2020), Giant Metrewave Radio Telescope (GMRT) (Paciga et al., 2013), and Hydrogen Epoch of Reionization Array (HERA) (DeBoer et al., 2017; HERA Collaboration et al., 2023).

1.5.3 Quasar spectra

In order to study the Universe during reionization, we need some bright sources of light from that era. Quasars, being one of the most luminous non-transient objects in the Universe, satisfy this requirement effectively.

After the discovery of the first quasars as bright point-like emitters (similar to stars in brightness but at much larger distances), they were initially called “quasi-stellar radio sources”(Hazard et al., 1963; Schmidt, 1963). Sooner, with the discovery of more sources with similar spectral features in the rest-UV regime, but not so prevalent radio emission (Sandage, 1965), they all were clubbed into a single category of “quasi-stellar objects”,

"QSOs", or quasars. The extreme brightness of quasars is due to the active super-massive black hole of mass $M_{BH} \sim 10^{7-10} M_{\odot}$ at the center of the galaxy, accreting the gas around it. As the gas is accreted, it is compressed and heated. Due to the temperature gradient along the radius of the disk, it emits radiation in all wavelengths. With the advancement in modern technology and the launch of the Euclid telescope, the anticipation of finding more and more high-redshift quasars is ever so high.

In the Fig. 1.2 we show a typical quasar spectrum at the redshift $z = 1.34$ (Charlton and Churchill, 2001). Once the light rays emitted by quasars enter the IGM, they can interact with the gas in the IGM through various processes, e.g., Thomson scattering, Rayleigh scattering, Compton scattering, etc. Thus acting as a tool to probe the Universe. Let $F_\nu(t_0)$ be the Spectral Energy Distribution (SED) of the quasar reaching us at time $t = t_0$. Let the intrinsic SED be F_ν. Assuming there's no emission along the line of sight of the quasar, then the Flux reaching us is given by

$$F_\nu(t_0) = F_\nu a^2 \exp\left(-\tau(\nu)\right) \tag{1.129}$$

where $\tau(\nu)$ is the optical depth of the medium. This is a very generic relation for any light; when travelling through a medium, it suffers attenuation from the medium. In the following sections, we will discuss how this light can be attenuated and how to calculate its respective optical depth. Although one can dive deeper into the properties of the quasars and how they affect the surrounding IGM, most of it will be covered in the upcoming chapters. Hence, we will rather discuss it where it is relevant. For instance, other than stars, quasars can also ionize the Universe and modify the ionized bubbles around their host halos. In Chapter 2 we will calculate and show the effect of ionized bubbles and discuss the quasar lifetime effects. In Eq. (3.5), we explain and discuss the quasar proximity zones. Hence, we will skip those details for now.

1.5.4 $Ly - \alpha$ Absorption

The Lyman-alpha ($Ly - \alpha$) transition refers to the transition of a hydrogen atom's electron from the second energy level $n = 2$ to the ground state $n = 1$, emitting a photon with a wavelength of 1216 Å (or a frequency of $\sim 2.47 \times 10^{15} Hz$). As discussed in the section §1.5.3, the HI density of the IGM along the line of sight of quasar spectra will attenuate the flux from the quasar as shown in Eq. (1.129). Here, the σ_α is the absorption cross section of the Lyα transition.

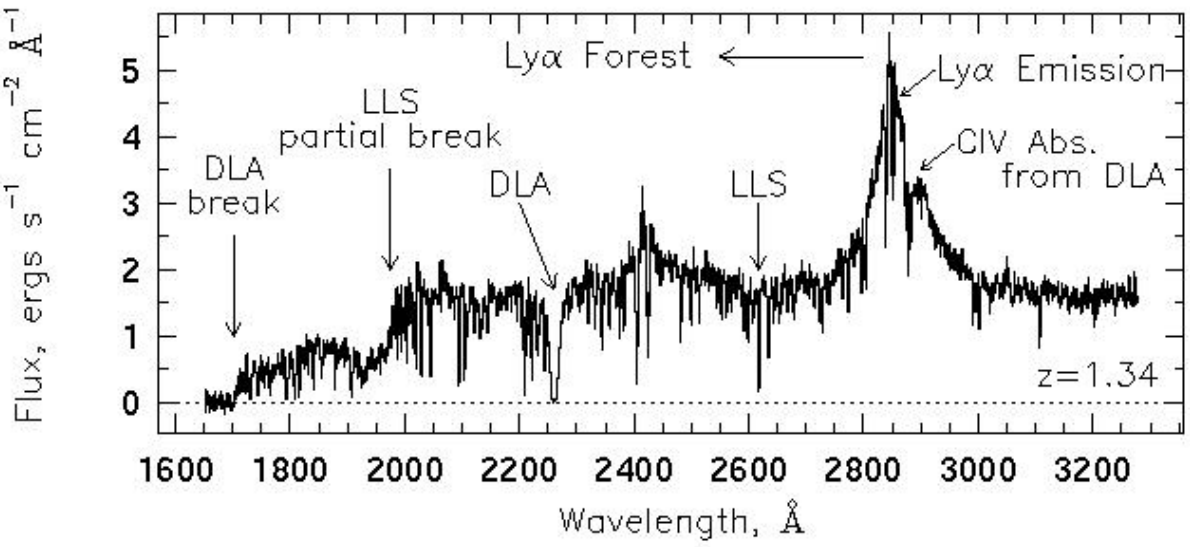


Figure 1.2. A typical quasar spectra at redshift $z = 1.34$. The absorption line bluewards of the $Ly-\alpha$ represents the $Ly-\alpha$ forest. The DLA (Damped Lyman Alpha) systems show very wide absorption features. The LLS (Lyman Limit Systems) shows sharp breaks in the spectra as all of the ionizing photons are absorbed.

The profile of σ_α is given by the Rayleigh scatter cross-section of photons with frequency $\nu = \nu_\alpha$ (which is a Lorentzian profile) modified by the thermal motion of absorbing HI atoms along the line of sight, given by the Maxwell-Boltzmann distribution (which is a Gaussian profile), results in a Voigt profile. The Voigt profile at the resonance resembles a Gaussian core, but has a long Lorentzian tail far from the resonance. The cross-section is written as follows:

$$\sigma_\alpha(\nu) = \frac{f_\alpha\, \pi e^2}{m_e c}\phi(\nu - \nu_\alpha), \tag{1.130}$$

where $f_\alpha = 0.416$ is the Lyα oscillator strength and $\phi(\nu - \nu_\alpha)$ is the Voigt profile.

Thus, for the photons with frequency ν traveling through a medium with HI number density given by $\mathrm{n_{HI}}$. The optical depth is written as

$$\tau_\alpha(\nu) = \int \mathrm{n_{HI}}\sigma(\nu)dr = \int \mathrm{n_{HI}}\frac{f_\alpha\, \pi e^2}{m_e c}\phi(\nu - \nu_\alpha)dr, \tag{1.131}$$

for Lyα absorption around the resonance, the Voigt profile can be approximated with a Dirac delta function. We can convert ν into observed frequency, $\nu_o = \nu\,(1+z)$, and the infinitesimal length element dr, to infinitesimal redshift dz, using the following relation from FLRW cosmology,

$$\frac{dr}{dz} \equiv \frac{c}{H(z)(1+z)} = \frac{c/H_0}{(1+z)\sqrt{\Omega_\mathrm{m}(1+z)^3 + \Omega_\Lambda}}. \tag{1.132}$$

Now substituting for ν_o and dz the Eq. (1.131) can be integrated to give

$$\tau_{GP}(z_\alpha) = \frac{\pi e^2}{m_e c} f_\alpha \frac{n_{\rm HI}(z_\alpha)}{\nu_\alpha H(z_\alpha)}, \tag{1.133}$$

where z_α is the redshift when any photon blueward of Lyα is redshifted to Lyα frequency. The τ_{GP} is also called the Gunn-Peterson optical depth. The Gunn-Peterson effect refers to the absorption of photons by HI in the IGM at Lyα the resonance. We can substitute for $H(z_\alpha)$ and $n_{\rm HI} = n_{H,0}\,{\rm x_{HI}}\,(1+z)^3$, where $\rm x_{HI}$ is the neutral fraction and $n_{H,0}$ is the number density today, we can re-write the above equation as (Gunn and Peterson, 1965; Barkana and Loeb, 2001)

$$\tau_{GP}(z) \approx 1.8 \times 10^5 {\rm x_{HI}} \left(\frac{\Omega_b h^2}{0.022}\right) \left(\frac{1+z}{7}\right)^3 \left(\frac{H_0}{70\ {\rm km/s/Mpc}}\right) \tag{1.134}$$

Where Ω_b is the baryon density parameter. We can see that even for ${\rm x_{HI}} = 10^{-4}$, $\tau_{GP} \approx 10$. Thus insinuating that even with very low $\rm x_{HI}$ all of the Lyα photons will be absorbed. We see this feature in the quasar spectra at high redshifts, where the flux blueward of Lyα is completely absent. This is called the Gunn-Peterson trough. If the $\rm x_{HI}$ is very low (which it is for low redshifts), the frequencies blueward of Lyα can eventually be redshifted to Lyα frequency and absorbed at that redshift. Hence, we see many absorption lines, corresponding to Lyα at different redshifts in the quasar spectra. This series of absorption lines is called the Lyman-alpha forest. In the Figure 1.3, the first three plots show the dense Lyα forest absorption blueward of Lyα. Whereas, in the fourth plot, we see the Gunn-Peterson trough.

The observed Lyα absorption features can be used to trace the s density and opacity fluctuations at different cosmic epochs. This can quickly be shown from the photoionization equilibrium condition (this is a valid assumption at low redshifts when the IGM is highly ionized). Now for such case, $n_{\rm HI} \ll n_H$, $n_{\rm HII} \approx n_H$, $n_e = \chi_{\rm He} n_H$, and from 1.122, we can write

$$n_{\rm HI} = \frac{\chi_{\rm He}}{a^3} \frac{\alpha_R(T)}{\Gamma_{\rm HI}} n_H^2 \tag{1.135}$$

The recombination rate as a function of Temperature (Rauch, 1998) and the Temperature-density power-law relation (Hui and Gnedin, 1997) can be approximated as,

$$\alpha(T) = 4.2 \times 10^{-13} {\rm cm^3 s^{-1}} \left(\frac{T}{10^4 {\rm K}}\right)^{-0.7}. \tag{1.136}$$

and,

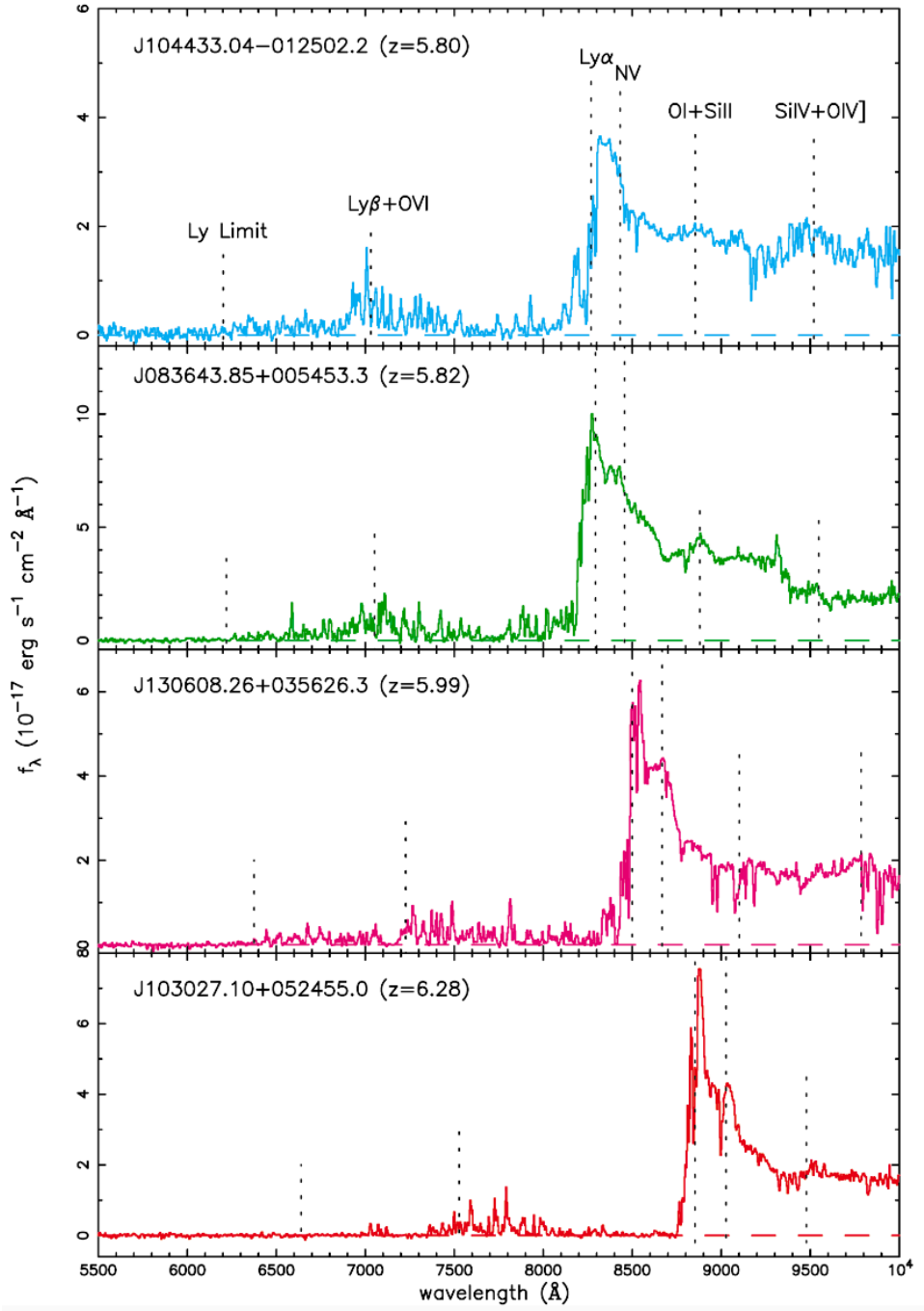


Figure 1.3. The first three panels of this figure show the quasars spectra showing $Ly-\alpha$ forest, while the last panel shows the Gunn-Peterson trough. The trough is strongly visible the closer to the EoR we move. This image is taken from (R. H. Becker et al., 2001).

$$T = T_0(1+\delta_b)^{\gamma-1}, \tag{1.137}$$

where δ_b is the baryonic overdensity contrast. Now, combining all the above terms with the Eq. (1.131) to get,

$$\tau_\alpha(\delta_b) = 5.01 \left(\frac{\chi_{\rm He}}{1.08}\right)\left(\frac{1-Y}{0.76}\right)^2\left(\frac{\Omega_b h^{3/2}}{0.0269}\right)^2\left(\frac{1+z}{6.5}\right)^6\left(\frac{9.23}{H(z)/H_0}\right) \tag{1.138}$$

$$\times\left(\frac{T_0}{10^4\ \rm K}\right)^{-0.7}\left(\frac{10^{-12}\ \rm s^{-1}}{\Gamma_{\rm HI}}\right)(1+\delta_b)^\beta, \tag{1.139}$$

where $\beta = 2.7-\gamma$. This gives us the relation between the Lyα optical depth and the baryonic matter overdensity contrast, given we know the values for $\Gamma_{\rm HI}$, T_0, and γ.

1.5.5 Damping wings

As we showed in Section §1.5.4, the Lyα absorption cross-section is so high that all the photons near the resonance of Lyα are completely absorbed for high-redshifts. Thus leaving us with the Gunn-Peterson trough. This restricts the applicability of Lyα transitions. However, we also saw that the general profile of the Lyα transmission is a Voigt profile. The Voigt profile, away from the resonance, has a long Lorentzian tail. Hence, for the case where $x_{\rm HI}$ is very high, we can expect some absorption from the tail of the Lyα profile, dampening the incoming flux. This tail, since seen in the absorption spectra, is called the Lyα damping wing. The damping wings can also be observed on the red side of the Gunn-Peterson trough (Miralda-Escudé, 1998), even if the Lyα forest on the blue side is fully opaque (absorbed).

In figure Fig. 1.4 (F. Wang et al., 2020), on the top panel, we see the full spectrum of quasar (J0252-0203, at redshift z = 7.00), with a fit on the red side using the principal component analysis (PCA). On the blue side, we see the prediction made using PCA. The bottom panels show the zoomed-in version near Lyα. On the bottom left panel, we see how the quasar spectrum is being dominated by the Gunn-Peterson trough on the blue side. But not with a sharp cut, we instead see the signature of the $Ly-\alpha$ damping wing, which is being highlighted on the bottom right panel. We also see the damping effects on the right side. This is how the damping wing signature looks.

Following the work in Miralda-Escudé, 1998, we will now try to calculate the Lyα damping wing profile caused by the homogeneous neutral IGM. The Rayleigh scattering cross-section profile of the Lyα resonance line by HI for a photon with angular frequency ω is written as(see the section §23, "Resonance line shape" in (Peebles, 1993)),

$$\sigma(\omega) = \frac{3\lambda_\alpha^2\Lambda^2}{8\pi}\frac{(\omega/\omega_\alpha)^4}{(\omega-\omega_\alpha)^2+(\Lambda^2/4)\,(\omega/\omega_\alpha)^6}\ . \tag{1.140}$$

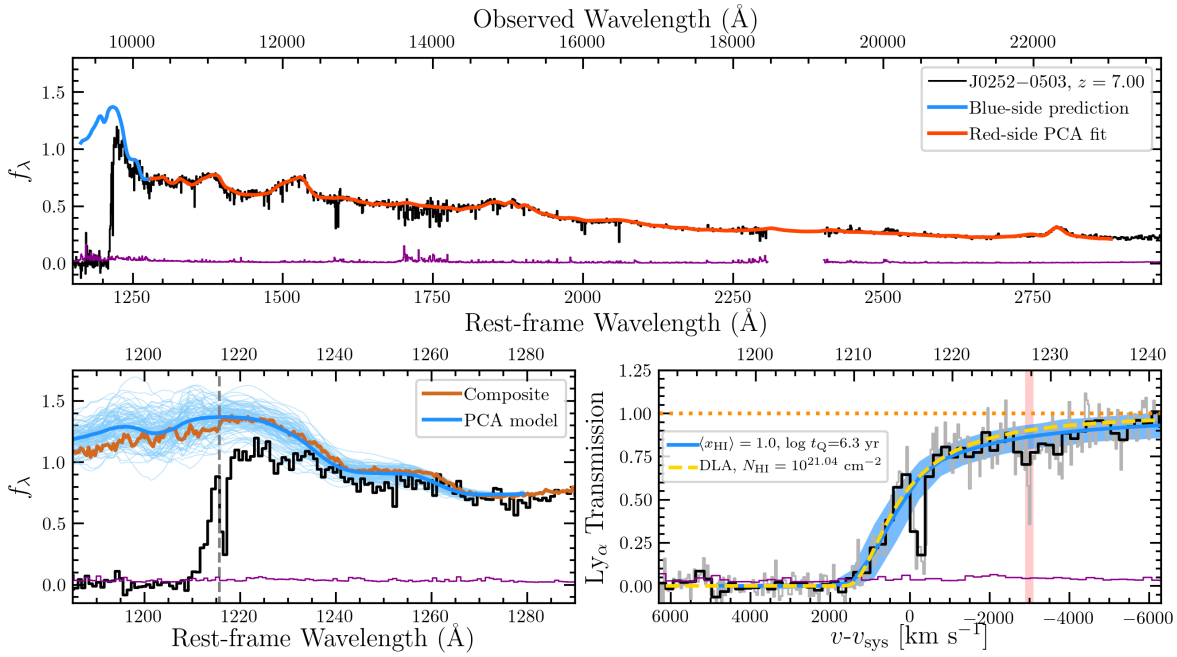


Figure 1.4. This figure illustrates what the damping wing profile looks like in the high redshift quasar spectrum, closer to the EoR. The bottom panels show how the Gunn-Peterson trough is not a sharp cut, and the red side of Lyα has the damping wing imprint. This quasar under study is J0252-0203, at redshift z = 7.00. This image is taken from (F. Wang et al., 2020).

Where $\Lambda = 6.25 \times 10^8 \mathrm{s}^{-1}$ is the decay constant for Lyα resonance, and $\omega_\alpha = 2\pi\nu_\alpha = 2\pi c/\lambda_\alpha$. For the tail part ($\|\omega - \omega_\alpha\| \gg \Lambda$) we can ignore $(\omega/\omega_\alpha)^6$. Assuming constant $n_{\rm HI} \equiv n_0$ for the IGM at $z_n < z < z_s$, where z_s is the redshift of the source and negligible HIdensity for $z < z_n$, the optical depth at observed wavelength $\lambda = \lambda_\alpha(1+z_s) + \Delta\lambda$ from the Eq. (1.131) can then be written as,

$$\begin{aligned}\tau(\Delta\lambda) &= \int_{z_n}^{z_s} \frac{dz}{1+z}\,\frac{c}{H(z)}\,n_0(1+z)^3\,\sigma\left(\frac{\omega}{\omega_\alpha}\right), \quad \text{with } \frac{\omega}{\omega_\alpha} = \frac{(1+z)}{(1+z_s)(1+\delta)} \\ &= \int_{z_n}^{z_s} \frac{dz}{1+z}\,\frac{c}{H(z)}\,n_0(1+z)^3\,\frac{3\lambda_\alpha^2\Lambda^2}{8\pi\omega_\alpha^2}\,\frac{(\omega/\omega_\alpha)^4}{(\omega/\omega_\alpha - 1)^2}\end{aligned} \tag{1.141}$$

where $1+\delta = \lambda/[\lambda_\alpha(1+z_s)]$ or $\lambda_\alpha/\lambda = \omega/\omega_\alpha = [(1+\delta)(1+z_s)]^{-1}$. The Gunn-Peterson optical depth from the Eq. (1.133) is $\tau_G P(z_s) = 3\lambda_\alpha^3\Lambda n_0/[8\pi c H(z_s)]$, where $\Lambda = \frac{8\pi^2 e^2}{m_e c\lambda^2} f$, and using $H(z) \propto (1+z)^{3/2}$, we get the following integral:

$$\tau(\Delta\lambda) = \frac{\tau_0 R_\alpha}{\pi}\int_{z_n}^{z_s}\frac{dz}{1+z}\left(\frac{1+z}{1+z_s}\right)^{11/2}(1+\delta)^{-4}\left[\frac{1+z}{(1+z_s)\,(1+\delta)} - 1\right]^{-2}. \tag{1.142}$$

Where $R_\alpha = \Lambda/4\pi\nu$. Now substituting $x = (1+z)/[(1+z_s)(1+\delta)]$ we get the following form,

$$\tau(\Delta\lambda) = \frac{\tau_0 R_\alpha}{\pi} (1+\delta)^{3/2} \int_{x_1}^{x_2} \frac{dx\, x^{9/2}}{(1-x)^2} \ , \tag{1.143}$$

The result of the integral is:

$$\int \frac{dx\, x^{9/2}}{(1-x)^2} = \frac{x^{9/2}}{1-x} + \frac{9}{7}x^{7/2} + \frac{9}{5}x^{5/2} + 3x^{3/2} + 9x^{1/2} - \frac{9}{2}\log\frac{1+x^{1/2}}{1-x^{1/2}} \ . \tag{1.144}$$

1.5.5.1 Modified Damping Wing Profile

In our work, we use the modified version of the above equation (D. Mortlock, 2016; Kist et al., 2024), where the following approximation for $\sigma(\omega)$ is used:

$$\sigma(\omega) = \frac{1}{4}\frac{3\lambda_\alpha^2\Lambda^2}{8\pi} \frac{(\omega/\omega_\alpha)^2}{(\omega-\omega_\alpha)^2 + (\Lambda^2)\,(16\pi^2)} \quad \text{if} \quad |\omega-\omega_\alpha| \ll \omega_\alpha \tag{1.145}$$

yielding the following result:

$$\int \frac{dx\, x^{1/2}}{4(1-x)^2} = \frac{x^{1/2}}{4(1-x)} + \frac{1}{8}\log\frac{1+x^{1/2}}{1-x^{1/2}} \ . \tag{1.146}$$

The difference between these two models is, in Eq. (1.146) the cross-section near the wings is approximated by a simple Lorentzian profile, whereas Eq. (1.144) inherently assumes a two-level hydrogen model, in which the bound electron is confined to the ground state ($n = 1$) and the excited state ($n = 2$) or the 1s and 2p states. The Eq. (1.146) is more accurate in the low wavelength regime, while the Eq. (1.144) is more accurate in the long wavelength regime. Since most of the absorption happens when the photons first enter the IGM, Eq. (1.146) is a better approximation (D. Mortlock, 2016).

1.6 Simulation techniques

We have seen some observational probes in the previous section. We would now like to see the available technology that we can employ to have a better theoretical understanding of the observations. The modeling of the reionization epoch can be broadly classified into three categories:

- Analytical/ Semi-Analytical Models: Analytic or Semi-analytic uses approximate analytic or parametric equations to solve the problem on hand. These methods often replace the complex motion of multi-particle systems with statistical models. That's why the early semi-analytical models (Wyithe and Loeb, 2003; T Roy

Choudhury and Ferrara, 2006; Pritchard et al., 2010; Mitra et al., 2011; Mitra et al., 2012; Mitra et al., 2015) have been reasonably successful in constraining the globally averaged quantities. The global properties of reionization can be derived by solving the Eq. (1.126), a first-order differential equation. Combining the Eq. (1.124) with $\Gamma_{\rm HI} \sim \frac{1}{t_{rec}} \left(\frac{\rm n_e \ n_{HII}}{\rm n_{HI}} \right)$ and substituting it in Eq. (1.126), we can further simplify to into

$$\frac{dQ_{HII}}{dt} = \frac{\dot{n}_{ion}}{\bar{n}_H} - \frac{Q_{HII}}{\bar{t}_{rec}}$$

(Tirthankar Roy Choudhury, 2022). It simplifies the overall equation at the expense of inhomogeneities in the ionization. These techniques are ideal for exploring a large parametric space of models.

- Full Radiation Hydrodynamic Simulations: As the name suggests, these simulations are the combination of the radiative transfer (RT) simulations, which aim to solve the full RT equations to calculate very precise interactions of photons propagating through a medium, and the full set of hydrodynamic equations (Eq. (1.9), Eq. (1.10), Eq. (1.11)) in cosmological settings. These are coupled with gravity to simulate the exact evolution of dark matter and gas required for the structure formation. The governing equation for the RT models is,

$$\frac{\partial I_\nu}{\partial t} + \frac{c}{a}\hat{\mathbf{n}} \cdot \vec{\nabla} I_\nu - \frac{\dot{a}}{a}\nu\frac{\partial I_\nu}{\partial \nu} + 3\frac{\dot{a}}{a} I_\nu = -c\kappa_\nu I_\nu + c j_\nu, \tag{1.147}$$

where I_ν is the specific intensity, κ_ν is the absorption coefficient, j_ν is the emission coefficient and $\hat{n}$ is the direction of the unit vector. The following are some research groups working on these simulations (Mellema et al., 2006; Iliev et al., 2006; Trac and Cen, 2007; Ghara et al., 2015; Ocvirk et al., 2016; Katz et al., 2019; Ocvirk et al., 2020; Kannan et al., 2022; Garaldi et al., 2022; Puchwein et al., 2023). Due to their high complexity, these simulations are computationally demanding and time-consuming. For our work, which requires running a grid of models for the comprehensive study of a wide range of parameter space, these simulations are not feasible.

- Semi-Numerical Simulations: Semi-numerical simulations provide a middle ground. These simulations combine analytical models with numerical methods to simulate the formation and evolution of cosmic structures such as galaxies, dark matter halos, and the IGM. They use either the empirical results derived from observations

or full radiation-hydrodynamic simulations to replace the complicated physics to set up the analytic framework and then employ numerical techniques like ES to solve for the desired results. Hence, they can capture most of the required physics without compromising much with efficiency. One such semi-numerical simulation tool, 21cmFAST (Mesinger and S. R. Furlanetto, 2008; Mesinger et al., 2011), which we used in this work, has been discussed in detail in §2.2, along with the theory of its working. Some other examples of semi-numerical simulations for reionization are: SIMFAST21 - (Santos et al., 2010), zreion - (Battaglia et al., 2013), DRAGONS - (Mutch et al., 2016; Geil et al., 2016), ASTRAEUS - (Hutter et al., 2021), and SCRIPT - (Tirthankar Roy Choudhury and Paranjape, 2018).

1.7 Structure of the thesis

In this thesis, we aim to understand and decode the topology of the Universe at high redshift, especially during the EoR $z = 7$. To view the Universe around that period, we generate extensive sets of large and patchy reionization boxes using 21cmFAST, subjected to varying astrophysical, IGM, and reionization source parameters. Within each box, defined by a unique combination of parameters, we locate halos of various masses, divided into specific mass bins. We assume these halos can host quasars and draw a random sightline through these quasars. We collect all these sightlines, and assuming they are the line of sight towards us, we study their properties when they travel through the box. These halo-sightline pairs will appear in every chapter and are the fundamental tools for all our analysis. We also keep the initial and boundary conditions constant between every model.

In Chapter 2, we use these models to study the dependency of the reionization topology on the varying astrophysical, IGM, and reionization source parameters by studying their signatures on Lyα damping wings. We use the aforementioned halo-sightline pairs and calculate the Lyα damping wing optical depth along them. We then study the properties of their median signal, M(DW), and the sightline-to-sightline 68-percentile scatter width around the median, ΔSW_{68}. The M(DW) and ΔSW_{68} are then used to show the nature of the dependency of the reionization topology on the model parameters.

In Chapter 3, we use similar models as described above and the M(DW), and ΔSW_{68} profiles calculated above to quantify the constraining power of the damping wings on

the model parameters using the Fisher Information Matrix (FIM). We also forecast the number of quasars needed to conduct our damping wing analysis robustly.

In Chapter 4, we calculate the quasar transmission flux for the ionizing photons at redshift $z = 6$ instead of calculating the damping wing profiles. We then study the dependency of the ionizing photon transmission flux on the model parameters and qualitatively look for the parameter space that can reproduce the short Mean Free Path (MFP) for ionizing photons at $z = 6$.

Finally, in Chapter 5, we summarize all our significant results and discuss the future prospects of the different projects carried out in this thesis.

2

Damping Wings Statistics

The work in this chapter has been presented and published in the AAS Journals under (Sharma et al., 2025). Additional details regarding the operating principle of 21cmFAST have been added to fit the thesis narrative. I carried out the scientific tasks and wrote all the texts with the support and guidance provided by the co-authors Fred Davies, Prakash Gaikwad, Fahad Nasir, and Sarah Bosman.

Abstract

The damping wing signatures in high-redshift quasars have proven instrumental in studying the epoch of reionization. With the upcoming Euclid mission set to discover many more quasars, it is crucial to explore what this new set of quasars might reveal not only about the reionization history but also its topology. The topology should influence the shape and variation of quasar-damping wing signals across sightlines. We use 21cmFAST to generate patchy reionization models in cosmological volumes with diverse astrophysical parameters for the ionizing sources. We examine the median, M(DW) and the sightline-to-sightline variation ΔSW_{68} for an ensemble of damping wing signals. We find that the neutral fraction x_{HI}, quasar lifetime t_q, quasar host halo mass M_{qso}, and minimum DM halo mass (that can support star formation) M_{min} significantly impact the M(DW). Parameters tied to the reionization topology, like x_{HI} and M_{min}, strongly affect ΔSW_{68}, compared to t_q. Our findings highlight that quasar damping wings are sensitive to M_{min}, a key variable linking reionization topology, history, and feedback processes. We also explore the convergence of damping wing signals and ionized bubble sizes with box size. We present a suite of models to assess the ability of future quasar samples to constrain astrophysical model parameters and the additional systematic uncertainty on the neutral fraction incurred when fixed to a single fiducial value.

Contents

2.1 Introduction

At redshifts close to the EoR, nearly all the flux corresponding to the Lyα wavelength is absorbed even when the HI fraction (x_{HI}) is as low as 10^{-4} (Gunn and Peterson, 1965). However, as we saw in **??**, the absorption cross-section of Lyman transitions exhibits a long-tail behaviour. Hence, we can instead observe the absorption around these tails far from the peak, the so-called "damping wing" (see §2.3.3) (Miralda-Escudé, 1998). These damping wing signals in the spectra of high-redshift quasars have been used to constrain the global neutral fraction, x_{HI}. Studies such as (Daniel J Mortlock et al., 2011; Bradley Greig et al., 2017; Frederick B. Davies et al., 2018; F. Wang et al., 2020; Bradley Greig et al., 2022) showed that the $x_{HI} = 0.5$ at $z \sim 7.0 - 7.5$. Most of the damping wing analysis were used to constrain mostly the neutral fraction, x_{HI}, (Bradley Greig et al., 2017; Frederick B. Davies et al., 2018; Charlotte A Mason et al., 2018). While works like (Kist et al., 2024) tried to constrain both the x_{HI} and the quasar lifetime t_q. In (Bradley Greig et al., 2019), it was hinted that the choice of the source model could shift the neutral fraction inference, and therefore should be included in the damping wings study. However, it was not emphasized much, as it was believed that it may not be important when considering the damping wings from individual quasars. The uncertainty on the neutral fraction from a single quasar – largely due to cosmic variance – was large. Hence, it becomes useful to consider an ensemble of damping wings spectra to explore the source parameter dependency of the reionization topology using the damping wings. Furthermore, the single-source models are inadequate in incorporating the effect

of reionization topology on the intrinsic scatter of the damping wing signal. Thus demanding a need for a comprehensive analysis of an ensemble of damping wing signals with a wider choice of parameters, covering both the astrophysical and source parameters. As a result, this chapter aims to examine the dependence of the median damping wing profile, M(DW), and the sightline-to-sightline scatter, ΔSW_{68}, on a set of source and astrophysical parameters. Unless stated otherwise, we will use the term "astrophysical parameters" when discussing the source, IGM, and astrophysical parameters.

In §2.2, we begin by first defining the architecture of the 21cmFAST simulation and how it carries out the theory discussed in the §1.3. We then discuss the parameter space and Lyα damping wing calculations in our models. Finally, we end the §2.3 with the summary of all the parameters used, and the properties of our simulation boxes. In §2.4, we demonstrate the results from our models and summarize the behaviour of damping wings subjected to the changes in the selected batch of parameters. In §2.5, we present our understanding of these results and the interesting features we observe. Lastly, in Appendix A.1, we perform various convergence tests for the damping wing signal across different box sizes, to determine the optimum box size for our study. We assume a flat ΛCDM cosmology throughout our work, based on the results from *Planck* (Aghanim et al., 2020), with cosmological parameters $h = 0.68$, $\Omega_m = 0.3$, $\Omega_b = 0.045$, and $\sigma_8 = 0.8$.

2.2 21cmFAST

21cmFAST is a semi-numeric simulation tool that produces 3D cosmological boxes of various physical fields in the early universe (Mesinger et al., 2011; Murray et al., 2020; Muñoz et al., 2022; Qin et al., 2020; Park et al., 2019). For our case, we generated boxes with density and ionization fields subject to our range of astrophysical parameters. 21cmFAST uses evolved IGM density, which is generated by perturbing the initial linear density field using second-order perturbation theory (Scoccimarro, 1998). This field is then smoothed onto a lower-resolution grid using an ES approach (S. R. Furlanetto et al., 2004). The ionized fields are computed from the evolved IGM density field by comparing the number of photons, integrated over time and produced by both UV and X-ray sources, to the number of neutral atoms within regions of decreasing filters of radius R. The size of these filters starts from the maximum photon horizon, R_{mfp}, down to the individual pixel resolution of a single cell, R_{cell}. The cells are flagged as fully ionized if the condition in the equation (Bradley Greig and Mesinger, 2018):

$$\zeta f_{\rm coll}(x, z, \mathrm{R}, \mathrm{M_{min}}) \geq 1 \tag{2.1}$$

Here $f_{\rm coll}$ represents the fraction of collapsed matter residing within halos in the region R with mass greater than $\mathrm{M_{min}}$ (Press and Schechter, 1974; Bond et al., 1991; Lacey and Cole, 1993; Sheth and Tormen, 1999). Effectively, $\mathrm{M_{min}}$ sets the minimum mass of halos that can support star formation.

Below this threshold mass of $\mathrm{M_{min}}$ the star formation is inefficient, thus quenching the amount of ionizing photons produced by such galaxies. The quenching of star formation may arise due to Supernova feedback, photoheating feedback, or inefficient cooling (Paul R. Shapiro et al., 1994; Mark L. Giroux et al., 1994; Hui and Gnedin, 1997; Barkana and Loeb, 2001; Springel and Hernquist, 2003; Mesinger and Dijkstra, 2008; Okamoto et al., 2008; Sobacchi and Mesinger, 2013b; Sobacchi and Mesinger, 2013a). In our simulations, this suppression is estimated using a redshift-independent duty cycle (Park et al., 2019).

$$f_{\rm duty}(\mathrm{M_h}) = \exp\left(-\frac{\mathrm{M_{min}}}{\mathrm{M_h}}\right) \tag{2.2}$$

Thus, for halos with a mass close to $\mathrm{M_{min}}$, only a fraction $f_{\rm duty}$ of them are forming stars with an efficiency of $f_{\rm star}$, the rest do not. This parameter, in conjunction with the $\mathrm{x_{HI}}$ controls the ionized bubble size distribution within our simulation boxes.

2.2.1 Density And Ionized Fields Calculation

In 21cmFAST, the evolved density fields are calculated by evolving the initial density fluctuations forward in time using perturbation theory. Specifically, 21cmFAST typically employs second-order perturbation theory by default, which provides a more accurate description of the nonlinear evolution of density fluctuations compared to LPT. Users can choose to switch to LPT for faster calculations by setting "USE_2LPT" to false when defining initial conditions (we discussed LPT in §1.7).

The simulation begins by generating initial conditions that describe the density fluctuations in the universe. These initial density fluctuations are represented as a Gaussian random field, with the mean and covariance functions determined by the underlying cosmological parameters and the initial power spectrum. The mathematical representation of Gaussian random fields involves specifying the mean and covariance functions.

- Mean Function: The mean function typically assumes a homogeneous and isotropic universe, where the mean density is constant across space. Therefore, the mean function $\mu(\mathbf{x})$ is usually a constant value.

- Covariance Function: The covariance function sets the correlation between density fluctuations at different points in space. For Gaussian random fields, the covariance $C(\mathbf{x}, \mathbf{x}')$ function between two points $\vec{x}$ and $\vec{x'}$ is determined using the matter power spectrum $P(k)$ through Fourier transforms:

$$C(\mathbf{x}, \mathbf{x}') = \int \frac{d^3k}{(2\pi)^3} P(k) e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} \tag{2.3}$$

Once the mean and covariance functions are specified, the Gaussian random field can be generated by drawing samples from a multivariate Gaussian distribution with the mean and covariance functions as parameters. These are the overdensities we require for the perturbation as shown in Eq. (1.28), Eq. (1.30), Eq. (1.32), and Eq. (1.35). *21cmFAST generates the density and velocity initial conditions in Lagrangian space and not in Eulerian space.* These density fields are then evolved to the desired redshift using first-order perturbation theory from (Yakov B. Zel'dovich, 1970), which is the Zel'dovich approximation (see §1.3.2.1). Before we proceed further, it is crucial to mention that the perturbation approach of 21cmFAST follows the DM evolution and does not capture virialized structure. There is no distinction between baryons and DM in 21cmFAST. This approximation works for large-scale structure, where the pixel scale is comparable to the size of the DM halos, as the baryons mostly follow the gravitational potential of DM on those scales. The complication arises when the baryons have to collapse, but we don't deal with it here.

After obtaining the density field, we follow the excursion-set formalism as discussed in §1.3.3, i.e., we take the density field perturbed to the desired redshift and smoothen it over with a filter function, as shown in Eq. (1.68). The filter scale is then varied from the size of the entire box to the 1 voxel resolution in discrete steps of a factor of 1.1. 21cmFAST uses a sharp k-space top hat filter to smooth the density field over different spatial scales. The sharp k-space top hat smoothing filter in k space is given by $W(k, R)$ is given by:

$$\widetilde{W}_k(kR) \propto \begin{cases} 1, & k \geq k_{max} \\ 0, & k < k_{max} \end{cases} \tag{2.4}$$

where k_{max} corresponds to the Nyquist frequency of the simulation grid (Nyquist, 1928; Shannon, 1949; Mesinger et al., 2011). The choice of this filter function is exciting because the distribution of change in the density $\Delta\delta_s$ is independent of δ_s and depends only on the variance $S(M) \equiv \sigma^2(k)$ (see the Eq. (1.76));

$$S \equiv \sigma^2(R) = \int \frac{dk}{k} \frac{k^3 P(k)}{2\pi^2} |W(kR)|^2 \tag{2.5}$$

Which means, as we change the filter scale from $R \to R - dR$, the change in δ_c follows a Markovian random walk. Note that the change in δ_c represents the Markovian random walk feature only for the k-sharp top hat filter. Now, the next step in ES theory, after smoothing the density field, is to calculate the first crossing point. For the filter scale R, if $(\delta_R > \delta_c)$, where $\delta_c \approx 1.686$, critical density for the spherical collapse, then we assign a halo of mass $M = (4\pi/3)\langle\rho\rangle R^3$, where $\langle\rho\rangle$ is the mean matter density, to the central pixel of that filter.

In §1.3.3 we defined the first crossing function for ES and modified ES in detail. The analytic form of first crossing function for ES is given by,

$$f(\sigma^2)d\sigma^2 = \frac{1}{\sqrt{2\pi}} \frac{\delta_c}{S^{3/2}} \exp\left(-\frac{\delta_c^2}{2S}\right) d\sigma^2 \tag{2.6}$$

where δ_c is the critical density threshold for collapse. The HMF is related to the first crossing by Eq. (1.78). We rewrite the equation here again for completeness's sake:

$$\frac{dn}{dM} = \frac{\langle\rho\rangle}{M} f(S) \left|\frac{dS}{dM}\right| \tag{2.7}$$

In , as we saw earlier, the halos are calculated by running the smoothing filter across the box with monotonically decreasing size. At every filter, the collapse condition is checked, and when the collapsed condition is satisfied for the first time, the mass is assigned to the center of the pixel corresponding to the filter scale. One can substitute the above equation with the ST mass function as well. The basic idea will remain the same.

Once these halos are located (their mass and coordinates are found) using the LPT. Their positions need to be corrected, especially across different redshifts. Since PS and ES were derived keeping LPT in mind. 21cmFAST uses the linear fields to locate halos. These halos are then displaced using the gradient of the velocity fields from the Zeldovich approximation. The first order approximation of velocity gradient used in 21cmFAST is calculated by taking the line-of-sight derivative of the linear velocity field from Eq. (1.31) (Re-quoting the velocity field here for the sake of convenience):

$$\vec{v}(\vec{k}) = \frac{i\vec{k}}{k^2}\dot{\delta}(\vec{k}) = \frac{i\vec{k}}{k^2}\dot{D}(z)\delta(\vec{k}) \tag{2.8}$$

The velocity gradient used in 21cmFAST is then given by,

$$\frac{dv_r}{dr}(\vec{k}, z) = ik_r v_r(\vec{k}z) \approx -\frac{k_r^2}{k^2}\dot{D}(z)\delta(\vec{k}) \tag{2.9}$$

The computed velocity fields are then used to perturb the positions of the halos to the Eulerian space. By applying these velocity-induced displacements, 21cmFAST ensures that the halo positions are accurately reflected. This is a very brief summary of the halo finder algorithm of 21cmFAST. This is also how 21cmFAST calculates the evolved density fields.

So far, we haven't had to use the analytic expression of the first crossing function. But in order to calculate the collapsed fraction $f_{\rm coll}$, we will have to rely on the analytic calculation since we are not resolving the individual halos down to tiny scales. After smoothing the density field at a given filter scale R_{cell}, the conditional collapsed fraction is calculated using the conditional first-crossing distribution (see Eq. (1.88)),

$$f_{\rm coll}(\delta_{\rm cell}, S_{\rm cell}; z) = \bar{N}\,\mathrm{erfc}\left[\frac{\delta_c(z) - \delta_{\rm cell}}{\sqrt{2\left(S_{\rm min} - S_{\rm cell}\right)}}\right] \tag{2.10}$$

We write $f_{\rm coll}$ instead of the first crossing function, because they are both the same for the PS formalism. The value of $\bar{N}$ is used, which matches the mean ST collapsed fraction from N-body simulations. Thus giving us the amount of matter that has collapsed into the halos, with halo mass $M \geq \mathrm{M_{min}}$, where $\mathrm{M_{min}}$ is the threshold above which the halos can form stars. The value of this collapsed fraction, in conjunction with the ionizing efficiency ζ, defined in §1.4.4, subject to the condition in Eq. (3.24) determines the ionized field in 21cmFAST.

2.3 Model Calculations

In this section, we describe the models we used to calculate our astrophysical parameters, how they are calculated in 21cmFAST, and how they control the reionization topology. In the §2.3.3, we describe the form of Lyα damping wing optical depth we used in our calculations. In the sections §2.3.4 we summarize all the parameters along with their fiducial values and ranges we used in our reionization models, and finally, in the §2.3.5 we establish the properties of our simulation boxes.

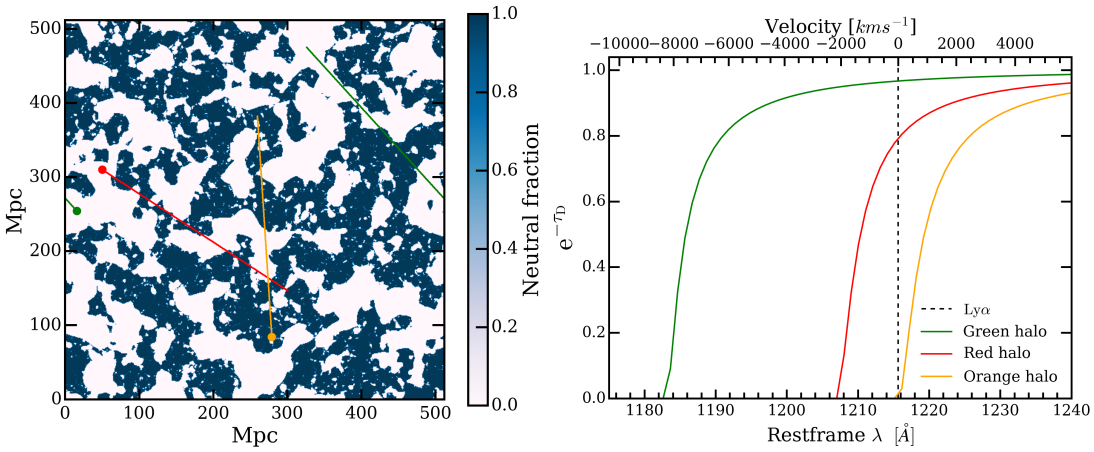


Figure 2.1. The upper panel shows a sample slice of the ionized box, along with three different halos (red, green, and orange) located at various positions within the box. The trajectories of random sightlines passing through these halos are also displayed. We show the damping wing signals corresponding to the respective sightlines in the lower panel.

2.3.1 Simulation Parameters

The standard 21cmFAST approach considers a simplified model for calculating the stellar mass of galaxies and the number of ionizing photons in our simulations. The stellar mass of a galaxy, M_{star}, can be expressed as a function of its host halo mass, M_h as follows (Park et al., 2019; Kuhlen and Faucher-Giguère, 2012; Dayal et al., 2014; Behroozi and Silk, 2015; Mitra et al., 2015; Mutch et al., 2016; Sun and S. R. Furlanetto, 2016; Yue et al., 2016) (see §1.4.2)

$$M_{star}(M_h) = f_{star}\left(\frac{\Omega_b}{\Omega_m}\right) M_h. \tag{2.11}$$

Here f_{star} denotes the fraction of galactic baryons present in stars. The f_{star} also has a power law dependency on halo mass (Park et al., 2019; Behroozi and Silk, 2015), which comes from the fact that gas is being exchanged between the IGM and the halo due to feedback processes:

$$f_{star} = f_{star,10}\left(\frac{M_h}{10^{10}M_\odot}\right)^{\alpha_{star}}, \tag{2.12}$$

where $f_{star,10}$ is the normalization factor representing the fraction of galactic gas in stars normalized to a halo mass of $10^{10}M_\odot$, and α_{star} represents the power law index of the dependence of f_{star} on halo mass.

We can express the other important quantity, f_{esc}, which governs the fraction of ionizing photons escaping from star-forming galaxies into the IGM, in a manner similar

to f_{star} (Park et al., 2019),

$$f_{esc} = f_{esc,10} \left(\frac{M_h}{10^{10} M_\odot} \right)^{\alpha_{esc}}, \tag{2.13}$$

where $f_{esc,10}$ is the normalization factor representing the fraction of ionizing UV escape fraction normalized to $10^{10} M_\odot$ halo mass fraction, and α_{esc} represents the power law dependency of f_{esc} on halo mass. Both f_{star} and f_{esc} have a physical upper limit of ≤ 1 by definition. This parameter plays a critical role in understanding the effects of feedback processes. Given the uncertainty surrounding both the nature and the precise value of f_{esc}, it can be used to calibrate the value of the neutral fraction when varying other source parameters.

Together these quantities tell us how much stellar mass is present in any galactic DM halo and how many ionizing photons they eject into the IGM. We can now calculate the ionizing UV efficiency of a galaxy (Bradley Greig and Mesinger, 2017) (see Eq. (1.112)):

$$\zeta(M_h) = 30 \left(\frac{f_{esc}}{0.12} \right) \left(\frac{f_{star}}{0.05} \right) \left(\frac{N_{\gamma/b}}{4000} \right) \left(\frac{1.5}{1 + n_{rec}} \right), \tag{2.14}$$

where $N_{\gamma/b}$ is the number of ionizing photons per baryon produced in the stars and n_{rec} is the number of times hydrogen atom recombines. For our case, the rate of recombination is so small in the IGM compared to Hubble expansion that we can neglect n_{rec} in the last term.

In the following, we will abbreviate the normalization terms $f_{star,10}$ and $f_{esc,10}$ as f_{star} and f_{esc}, respectively, unless stated otherwise.

We aim to compare the effects of the aforementioned parameters on reionization topology at a fixed neutral fraction. In our models, we adjust f_{esc} to calibrate the desired global neutral fraction for a given combination of parameters. We expect f_{star} to be degenerate with f_{esc}, something which we also see in our results, although a subtle distinction remains due to differences in the halo mass corresponding to the upper limit value of 1. All of the parameters described thus far are astrophysical in nature and can be tuned in our simulations, thereby impacting the reionization topology on a global scale. Table Table 2.1 lists these parameters, along with their respective ranges and fiducial values used in this study.

2.3.2 Quasar Model

After obtaining the reionization topology, we conduct our analysis by selecting random halos of a specific mass and observing a line of sight toward each halo, assuming it hosts

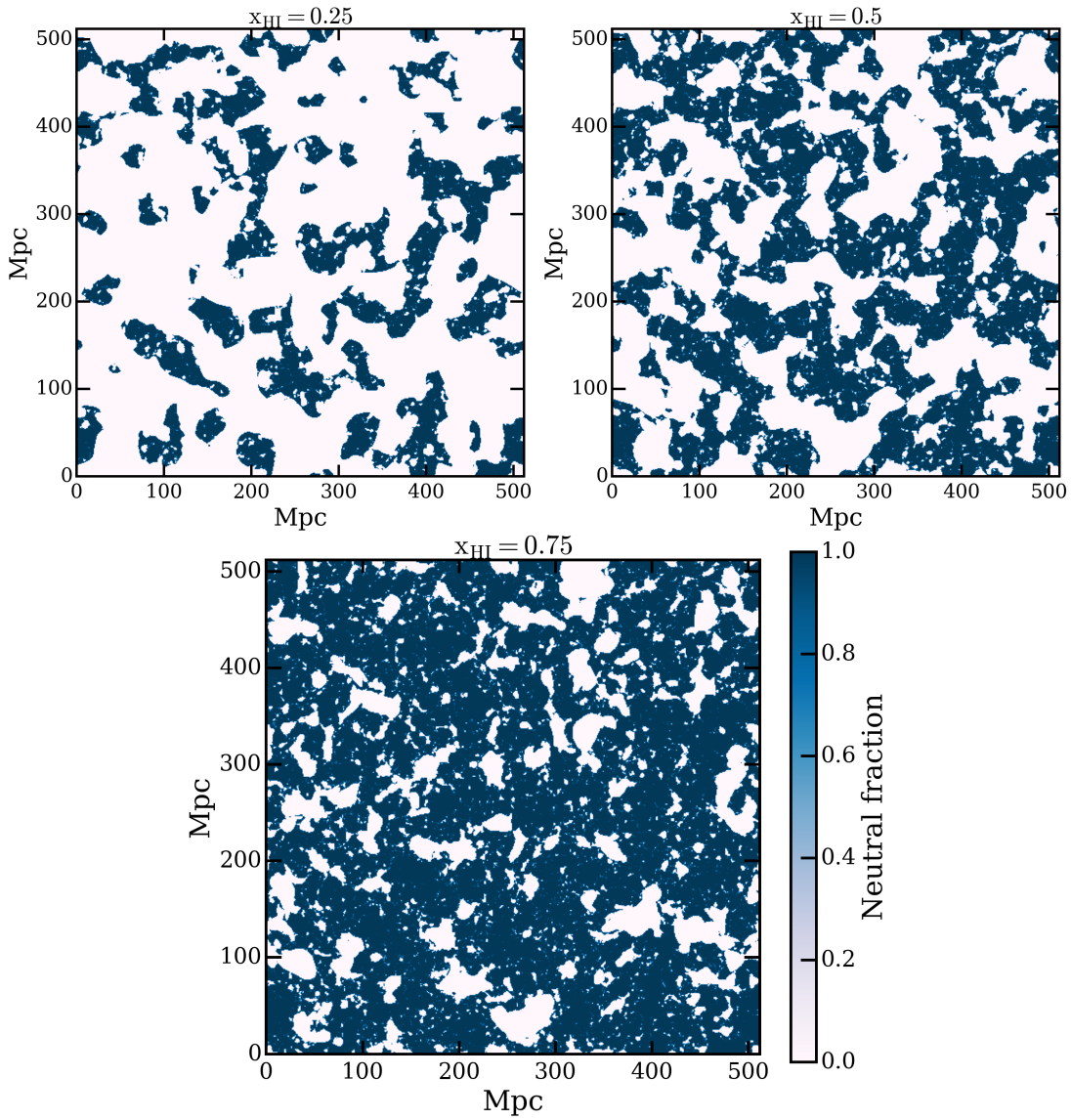


Figure 2.2. The first panel shows the slice of the ionized box with $x_{HI} = 0.25$, the second panel represents $x_{HI} = 0.5$, while the third panel represents $x_{HI} = 0.75$.

a quasar. In contrast to the calculation of the ionization field described above, for this, we require instantiating discrete halos within the semi-numerical simulation box. We use the method from (Mesinger and S. Furlanetto, 2007) to locate DM halos using ES theory on the initial conditions. These halo positions are then perturbed using velocity fields, calculated via LPT, to obtain the corrected locations at the desired redshift. Although the ionized field we derive is based solely on the density field, we use this halo catalog to position our quasars. This approach enables us to include lower-mass halos in our damping wing analysis, which would otherwise require much higher particle resolution. Given that quasars are likely to reside in halos with masses around $10^{12} M_{\odot}$, we explored a range from $10^{9} M_{\odot}$ to $10^{12} M_{\odot}$ to fully capture the dependence on host halo mass

(e.g. Pizzati et al., 2024; Eilers et al., 2024) and probe the regime of damping wings that affect lower-mass galaxies as well (Keating et al., 2024).

Since we are observing a line of sight towards a quasar, a sufficiently luminous quasar could influence the reionization topology in its vicinity, potentially changing the strength of its observed damping wing. To study this effect, we assume a quasar with a lifetime t_q, residing in a halo of mass M_{qso}. During its active lifetime, the quasar carves out an ionized bubble around it, $R(t_q)$. The ionizing photon emission rate for luminous quasars at redshift $z \geq 7$ is $\dot{N}_{ph} \simeq 10^{57} s^{-1}$ (Daniel J Mortlock et al., 2011). The Strömgren radius of this bubble is expressed as (Paul R Shapiro and Mark L Giroux, 1987; Cen and Haiman, 2000):

$$R_{(t_q)} = \left(\frac{3\dot{N}_{ph} t_q}{4\pi \langle n_H \rangle} \right)^{1/3}, \tag{2.15}$$

where $\langle n_H \rangle$ is the average number density of HI within the sphere.

The above equation assumes a homogeneous reionization process, whereas reionization is inherently inhomogeneous. Moreover, quasar emission is likely anisotropic, meaning the equation does not fully capture the overall shape of the ionized bubble. However, since we are primarily concerned with ionizing photons along the line of sight, we can use the relation to estimate the expansion of the ionized bubble by considering the density and neutral fraction distribution along the line of sight, assuming spherical dilution of the photon flux. Specifically, we calculate the total number of HI atoms on a spherical shell of radius R_{t_q} and a thickness of 1 pixel. To ionize these hydrogen atoms, an equivalent number of photons is required. Therefore, we equate the total number of photons produced by the quasar over its lifetime, t_q, with the number of hydrogen atoms encountered on a spherical shell, incorporating variations in density and neutral fraction along our line of sight:

$$\dot{N}_{ph} t_q = \int_0^{R_{ion}} n_H(r) 4\pi r^2 dr. \tag{2.16}$$

The integration is performed over the voxels along the line of sight corresponding to the radius of the bubble, neglecting any recombination within this region, which should be negligible for the quasar lifetimes we consider. Here, $n_H(r)$ represents the number density of hydrogen atoms in the voxel spanning the distance between r and $r + dr$. We thus obtain the radius R_{t_q}. Quasar activity modifies the reionization topology locally, up to the extent of R_{t_q}. We assume the ionizing photon emission rate, $\dot{N}_{ph}$, remains constant,

meaning the quasar lifetime, t_q, is the sole parameter controlling the distance to the ionization front. Notably, the case of $t_q = 0$ is analogous to a sightline originating from a typical star-forming galaxy, which produces significantly fewer ionizing photons than a luminous quasar, especially for small host halo masses ($M_h \leq 10^{11} M_\odot$). The effects of quasar lifetime are incorporated *a posteriori* into our calculation, specifically during the computation of random skewers through the simulation box. They are evaluated exclusively along the line of sight for computational efficiency.

2.3.3 Lyα Damping Wings

We calculate the transmitted flux from the total Lyα damping wing optical depth τ_D at an observed wavelength of $[\lambda_{\rm obs} = \lambda_\alpha(1+z)]$ (where $\lambda_\alpha = 1215.67$ Å is the *Lyα* restframe wavelength) along the length of our random skewer (treated as the line of sight), originating from a halo at redshift z_s. This is done by summing the contribution of τ_D from each HI patch encountered along the length of the skewer using the approximation (Mesinger and S. R. Furlanetto, 2008), (see §1.5.5 for the derivation of the damping wing profile)

$\tau_D(z) = \frac{\tau_{\rm GP} R_\alpha}{\pi} \sum_i \left\{ {\rm x_{HI}}(i)(1+\delta(i)) \left(\frac{1+z_{b_i}}{1+z}\right)^{3/2} \times \left[I\left(\frac{1+z_{b_i}}{1+z}\right) - I\left(\frac{1+z_{b_i}}{1+z}\right)\right]\right\}$, where $\tau_{\rm GP} \approx 7.16 \times 10^5 [(1+z_s)/10]^{3/2}$ is the Gunn-Peterson optical depth of the IGM (Gunn and Peterson, 1965), and $R_\alpha = \Lambda/(4\pi\nu_\alpha)$, where $\Lambda = 6.25 \times 10^8 s^{-1}$ is the decay constant of Lyα at resonance and $\nu_\alpha = 2.47 \times 10^{15} Hz$ is the frequency of the Lyα transition. ${\rm x_{HI}}(i)$ is the HI fraction in the $i^{\rm th}$ patch, and $\delta(i)$ is the matter overdensity in that patch. Lastly, the integration term I is defined as (D. Mortlock, 2016; Kist et al., 2024),

$$I(x) = \frac{1}{4}\left[\frac{x^{1/2}}{1-x} + 2\log\left|\frac{1-x^{1/2}}{1+x^{1/2}}\right|\right]. \tag{2.17}$$

In our calculation of optical depth, we do not include the absorption inside the proximity zone and are looking at the IGM damping wing alone, as our primary concern is to study the effects of reionization topology.

2.3.4 Summary of Parameters

In this subsection, we summarize all the parameters used in the simulation. The first type is the "Source parameters", which can be tuned in the simulation and have a global effect on reionization topology. These include $f_{\rm star}$, $\alpha_{\rm star}$, $\alpha_{\rm esc}$, and $M_{\rm min}$. As stated previously, the upper limit of $f_{\rm star}$ is ≤ 1. The key distinction between $\alpha_{\rm star}$ and $\alpha_{\rm esc}$ lies in their respective ranges: $\alpha_{\rm star}$ has a positive range, reflecting the fact that

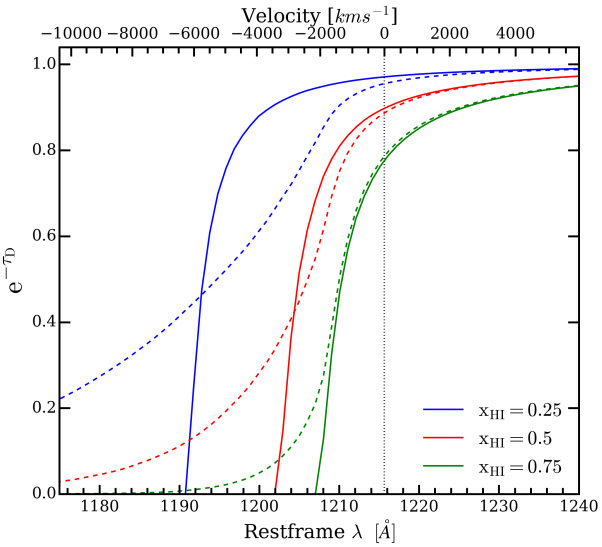


Figure 2.3. The median (solid line) and mean (dashed line) transmitted flux for a given neutral fraction. The mean and median are calculated over 10,000 randomly distributed sightlines originating at the halos of mass $M_{qso} \approx 4 \times 10^{11} M_\odot$.

Parameter	Range	Fiducial model	Parameter type
x_{HI} (Global mean neutral fraction)	$(0.25, 0.75)$	0.5	IGM parameter
M_{min} (Minimum halo mass to support star formation)	$(2M_p, 200M_p)$	$20M_p$	Source parameter
α_{esc}	$(-1, 0)$	-0.5	Source parameter
α_{star}	$(0, 1)$	0.5	Source parameter
$\log_{10} f_{star}$	$(-2, -0.25)$	-1.125	Source parameter
t_q (Quasar lifetime)	(0 Myr, 30 Myr)	1 Myr	Quasar parameter
M_{qso} (Host halo mass)	$(\approx 1 \times 10^9 M_\odot, \approx 4 \times 10^{12} M_\odot)$	$\approx 4 \times 10^{11} M_\odot$	Quasar parameter

Table 2.1. Astrophysical Parameters Space: This table lists all the parameters used in our study along with their respective ranges, fiducial values, and parameter types.

higher-mass halos tend to have more gas available for star formation, while α_{esc} has a negative range, indicating that high-mass halos possess deeper potential wells which prevent the formation of low-density channels through which ionizing photons can escape. Our parameter ranges are consistent with the constraints outlined in (Park et al., 2019), following calibration to UV Luminosity Function (LF). Finally, the range of M_{min} is determined based on the average mass of a pixel (M_p) within the simulation box. In our fiducial case, the pixel mass is $M_p \approx 10^{8.28} M_\odot$. Thus, we set M_{min} slightly above this value ($= 2 \times M_p$) to avoid numerical issues from unresolved halos and extended the range up to $200 \times M_p$. The fiducial value for M_{min} is approximately $10^{9.58} M_\odot$ ($= 20 \times M_p$).

The second parameter type is the "IGM parameter", which is indirectly adjusted by varying f_{esc} in our simulation. The sole constraint on f_{esc} is that $f_{esc} \leq 1$. For x_{HI}, we chose a range of $(0.25, 0.75)$, with a fiducial value of 0.5 at redshift 7.

Lastly, the third type is the "Quasar parameters", which include the lifetime of quasar activity (t_q) and the mass of the halo hosting the quasar (M_{qso}). The lower limit of $t_q = 0$

Myr (inactive quasar) represents galaxy spectra for low-mass halos, while the upper limit of $t_q = 30$ Myr. The studies done in (Morey et al., 2021; Eilers et al., 2021) suggest the average lifetime of a quasar ~ 1 Myr for a population of quasars around $z \geq 6$. Hence, for our fiducial model, we set $t_q = 1$ Myr.

The range of M_{qso} is determined after the simulation has run and the halo catalog has been generated, but typical values of M_{qso} range from $10^9 M_\odot$ to $10^{12} M_\odot$. The selection of M_{qso} and grid parameters is discussed in the next subsection.

2.3.5 Grid Parameters

To analyze the signature of damping wings, we generate a grid of simulation boxes with reionization topologies governed by the parameters in Table 2.1, using 21cmFAST. Each box in the grid has a side length of 512 Mpc, with 2048^3 number of grids and a grid of 512^3 for the ionized field. Periodic boundary conditions are applied to all boxes. After generating the simulation boxes and halo catalog, we select a sample of 6-7 evenly distributed mass bins, depending on the total number of available bins, ranging from 10^9 to $10^{12} M_\odot$.

For each mass bin, we require a host halo for the quasar and a randomly directed sightline originating from the halo. We achieve this by randomly sampling 10,000 halos from each bin and drawing a skewer in a random direction from each halo. This random skewer serves as our line of sight, along which we evaluate the damping wing signal. In cases where the number of halos in the mass bin is fewer than 10,000 (for very massive halos), we loop over the existing halo sample with different random sightlines to generate 10,000 skewers. Each skewer extends over a length of 300 Mpc. Along these skewers, we compute the Lyα damping wing optical depth using Equation §2.3.3 and evaluate the transmission flux. This entire set of mass bins, along with 10,000 skewers for each bin, constitutes one model.

The selection of halos and sightlines remains consistent across different models, with only the parameters discussed above being varied. For our analysis, we plot the mean and median transmission flux for the observed wavelength, for each mass bin and model. We also examine the sightline-to-sightline scatter within the 68^{th} percentile region around the median, which we call ΔSW_{68}, to study the impact of these parameters and cosmic variance. In all our results we display the plots for halo mass $M_{qso} \approx 4 \times 10^{11} M_\odot$ (fiducial value of M_{qso}) unless stated otherwise.

Lastly, we tested the effect of changes in reionization topology, due to varying box

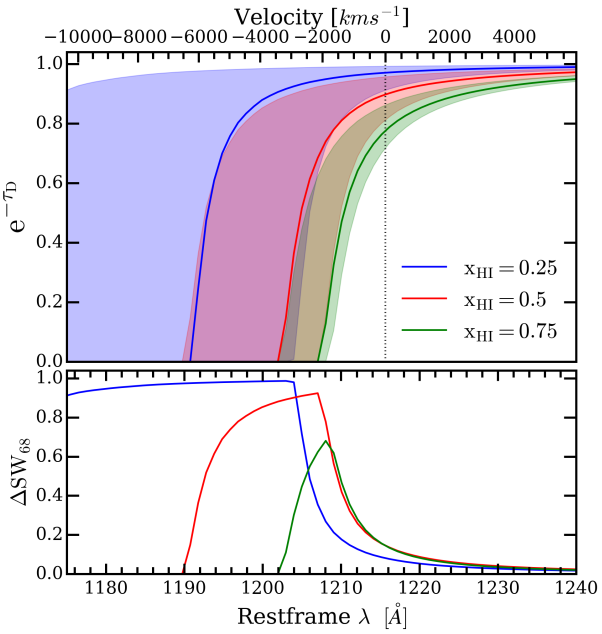


Figure 2.4. The M(DW) signals for all three different values of x_{HI} calculated over 10,000 randomly distributed sightlines originating at the halos of mass $M_{qso} \approx 4 \times 10^{11} M_{\odot}$. The shaded region in the upper panel represents the 68 percentile scatter of the damping wings around the M(DW). The lower panel shows the width of the 68 percentile region, also called scatter with ΔSW_{68}. The higher the value of x_{HI} the higher the effect of damping on the transmission flux. Also, the higher the value of x_{HI} the lower (more constrained) the width of the ΔSW_{68}.

sizes, on the damping wings, while maintaining a consistent grid resolution. The box sizes tested range from [(256 Mpc)^3, (512 Mpc)^3, (640 Mpc)^3, (768 Mpc)^3, (896 Mpc)^3], with the ICs and evolved field cell sizes scaled proportionally, i.e., ICs = 4 × box size with a cell size of 1 Mpc. For each box size, we ran a set of six different realizations of the initial conditions to minimize uncertainties arising from the random distribution of the initial Gaussian fields. The goal of these tests was to determine the box size beyond which the average damping wing profiles converge. The results and discussions of these tests are presented in Appendix A.1.

2.4 Results

In this Section, we present our findings from running the grid of models. All simulations were conducted with boxes of 512 Mpc in length at redshift 7.0. Unless otherwise specified, the parameter values used in our case studies are based on the fiducial model Table 2.1.

For illustration purposes, the upper panel of Fig. 2.1 displays a sample slice of the ionized box from our fiducial model. On this panel, the red (green and orange) dots rep-

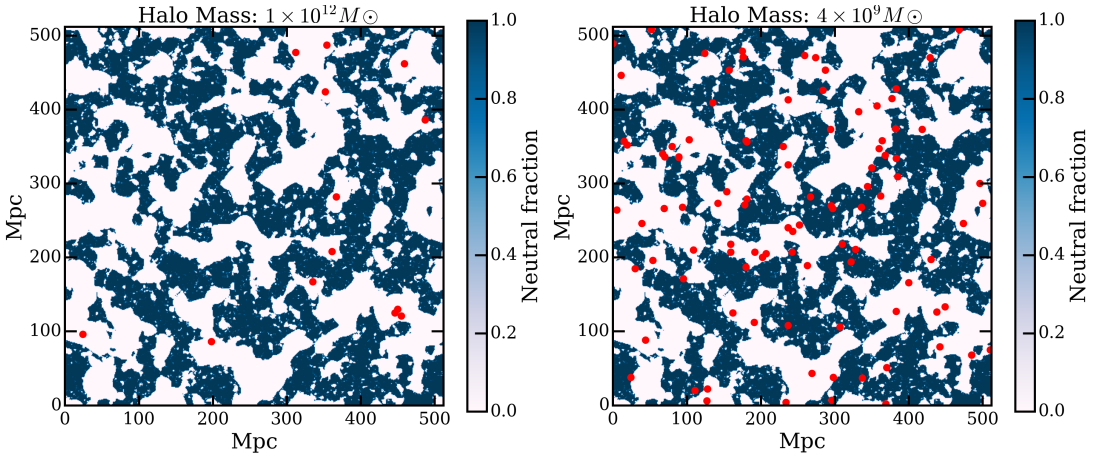


Figure 2.5. The distribution of high-mass ($\approx 9 \times 10^{11} M_\odot$) (left plot) and low-mass halos ($\approx 4 \times 10^{9} M_\odot$) (right plot). The high-mass halos reside in ionized regions, whereas the low-mass halos are distributed all over the box

resent halos, and the corresponding red (green and orange) lines indicate the trajectory of light from the quasar along the line of sight. In the lower panel of the same Figure, we show the resulting damping wing signal from the respective color-coded sightline.

As expected, we observe the strongest damping of the transmitted flux for the orange halo, which is located in a predominantly neutral region and encounters several neutral islands along the sightline. In contrast, the green and red halos, situated in ionized regions, experience less damping. However, while both the red and green halos reside in ionized regions, the sightline from the red halo traverses more neutral islands compared to that of the green halo. As a result, the red sightline experiences relatively greater damping than the green one. This illustrates the dependence of the Lyα signal on both the local environment of the host halo and the global environment along the sightline.

In Fig. 2.2 we present the slices of the ionized box for 0.25, 0.50, and 0.75 mean global neutral fraction to illustrate how the reionization topology changes with the x_{HI}. In Fig. 2.3 we show the *mean* transmitted flux for the same values of x_{HI}. Unlike the median transmission flux, which represents a typical damping wing spectra, the mean transmission flux averages multiple damping wing profiles that terminate at various distances, effectively stacking them on top of one another. As a result of this stacking and the substantial cosmic variance, a long-tail behavior is observed in the mean signal. This tail is particularly prominent in cases with greater scatter, such as $x_{HI} = 0.25$.

Some studies have examined this stacking of profiles in detail (Ďurovčíková et al., 2024). However, for our purposes, we are more interested in examining the individual damping wing profiles and their variation with reionization topology. Therefore, we focus

on the M(DW), which better captures the appearance of a typical damping wing, rather than the mean. Furthermore, the M(DW) is more robust for high-redshift quasars, as the number of sightlines is limited. From this point onward, unless stated otherwise, our analysis will be based primarily on the M(DW).

In Fig. 2.4, we present the M(DW) signals for three different values of x_{HI} (0.25, 0.5, 0.75), originating from halos with fiducial M_{qso}. The shaded region represents the 68th percentile scatter of the damping wings around the M(DW). As expected, for larger values of x_{HI}, the sightlines encounter more neutral regions, resulting in significantly more damping. The shaded region illustrates the variability of the damping wings within the 68th percentile around the M(DW).

We also observe that for higher values of x_{HI}, it is more probable for a sightline to intersect a neutral island earlier in its journey. Conversely, for lower values of x_{HI}, the sightline can travel a greater distance before encountering any neutral islands. This trend is reflected in the scatter width (ΔSW_{68}) of the damping wing profiles, shown in the lower panel of Fig. 2.4. Notably, ΔSW_{68} increases as the neutral fraction decreases, particularly at larger distances.

The most massive halos tend to reside deep within ionized regions, as these high-mass halos correspond to the peaks in the density field, which host higher concentrations of both high- and low-mass halos. In contrast, low-mass halos are more evenly distributed throughout the simulation box. This distribution is evident in Fig. 2.5. The left panel shows that high-mass halos ($\approx 9 \times 10^{11} M_{\odot}$) are biased towards ionized regions, while the right panel illustrates that low-mass halos ($\approx 4 \times 10^{9} M_{\odot}$) are dispersed throughout the box. This bias is also reflected in the damping wing signals for these halos. In Fig. 2.6, we observe overall weaker damping for sightlines originating in massive halos, consistent with their tendency to be located in larger ionized regions.

Fig. 2.7 illustrates the effect of t_q on the damping wings. The first three plots show the location of the halo and the corresponding sightline, ordered by increasing t_q values [0, 1, 30] Myr. The red-shaded region represents the size of the ionized bubble. In the fourth plot, we compare the damping wing profiles for the quasar-off case ($t_q = 0$ Myr) with the quasar-on cases ($t_q = 1$ and 30 Myr). The $t_q = 0$ case is particularly relevant for damping wing signals originating from galaxies (e.g. (Umeda et al., 2024; Keating et al., 2024)).

In Fig. 2.8 we explore the effects of quasar lifetime for two extreme cases ($t_q = 0$ and $t_q = 30$ Myr) on an ensemble of halos with the fiducial M_{qso}. As expected, quasar lifetime significantly impacts the M(DW), being the second most dominant factor after

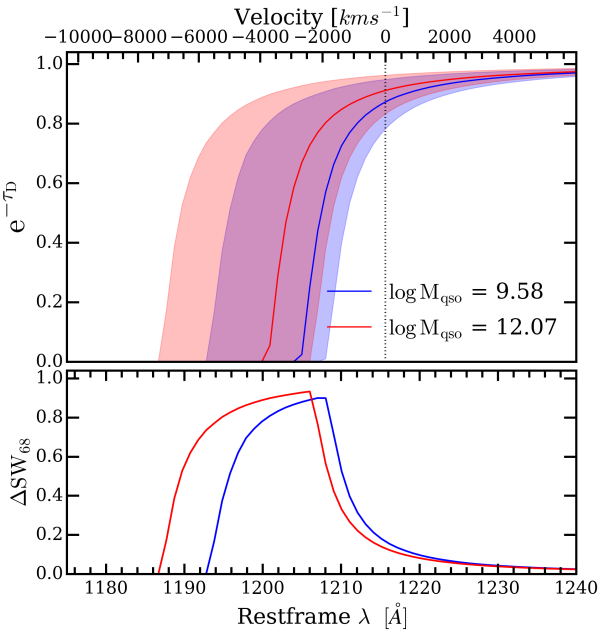


Figure 2.6. Similar to Fig. Fig. 2.4, we show the variation in the M(DW) signal and ΔSW_{68} as a function of the mass of halo hosting quasar (M_{qso}). Quasars living in more massive halos suffer less damping.

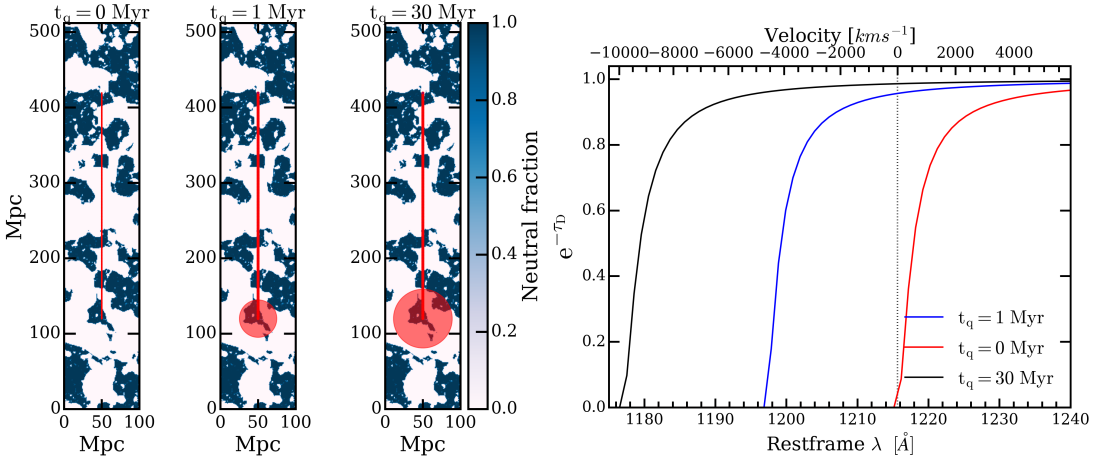


Figure 2.7. The effect of quasar lifetime on damping wings. The upper three panels show the halos and the sightlines for $t_q = [0, 1.30]$ Myr from left to right, with the shaded region representing the size of the respective ionized bubble carved by the quasar. The lower panel shows the effect of quasar activity on the damping wings. The longer the quasar is active, the bigger the ionized bubble and hence less damping is observed.

the neutral fraction. This raises the issue of degeneracy between quasar lifetime and neutral fraction when examining any particular damping wing profile (e.g. Frederick B. Davies et al., 2018). We note that if t_q is longer than 30 Myr, the quasar carves out a large enough ionized region that the damping wing signal is almost entirely erased, even in a fully neutral IGM. Therefore, for practical measurements, we restrict $t_q \leq 30 \times 10^6$ years. Despite the significant variation in the M(DW) signal across this range of t_q, the

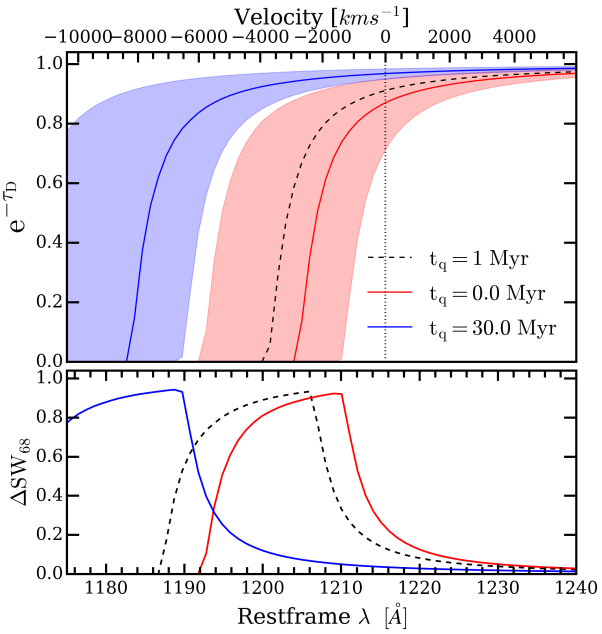


Figure 2.8. Similar to Fig. Fig. 2.4, the upper panel shows the damping wing signals for $t_q = 0$ Myr (red) and $t_q = 30$ Myr (blue). The dotted curve represents the fiducial model with $t_q = 1$ Myr.

width of the scatter at fixed distance from the ionization front remains largely unchanged, indicating that t_q has a minimal effect on the scatter. We discuss this further in §2.5, where we evaluate the effective values of x_{HI} required to reproduce the M(DW) signals for different t_q. This behavior can be attributed to the fact that quasar activity primarily alters the reionization topology locally, while the scatter arises from the distribution of HI beyond the edge of the local bubble.

We also explore the impact of varying M_{min}, which dictates the minimum halo mass capable of supporting star formation. For higher values of M_{min}, only the most massive halos contribute to the reionization topology. Since these halos are far fewer in number compared to their lower-mass counterparts, the reionization topology is primarily governed by larger ionized bubbles. In Fig. 2.9, we present slices of the ionized boxes for $M_{min} = 10^{8.58} M_\odot$ and $M_{min} = 10^{10.58} M_\odot$. As expected, the ionized field for $M_{min} = 10^{8.58} M_\odot$ features finer ionized bubbles and extended neutral regions. In contrast, the slice corresponding to $M_{min} = 10^{10.58} M_\odot$ exhibits larger, more coarsely distributed ionized regions. Consequently, for $M_{min} = 10^{10.58} M_\odot$, we anticipate an overall reduction in the damping effect.

In Fig. 2.10, we observe that the change in reionization topology due to varying M_{min} has a noticeable impact on the damping wing profiles. Specifically, for $M_{min} = 10^{10.58} M_\odot$, there is indeed less damping compared to $M_{min} = 10^{8.58} M_\odot$, as the neutral islands are

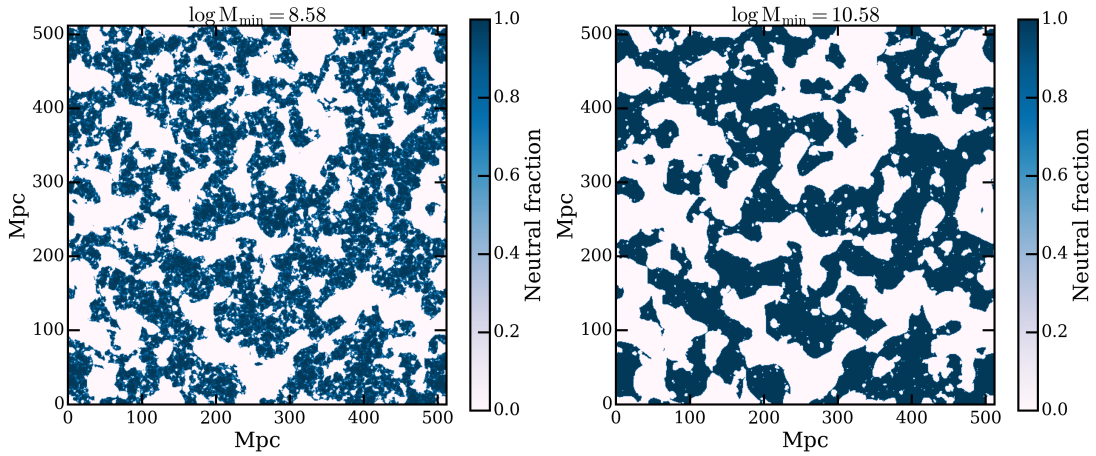


Figure 2.9. The slice of the ionized box with $M_{min} = 10^{8.58} M_\odot$ and $M_{min} = 10^{10.58} M_\odot$ from left to right respectively. On the left panel, we see the dominance of smaller and finer bubbles. Whereas, the topology is dominated by much larger and coarser bubbles on the right panel.

smaller. For the same reason, and as discussed in the case of x_{HI}, we observe more scatter in the damping wings for $M_{min} = 10^{10.58} M_\odot$. Although the $M_{min} = 10^{8.58} M_\odot$ box appears more patchy, the likelihood of encountering a neutral region along any random sightline is higher in this box compared to the $M_{min} = 10^{10.58} M_\odot$ box. This is because the $M_{min} = 10^{8.58} M_\odot$ box contains larger and more widely distributed neutral islands than the $M_{min} = 10^{10.58} M_\odot$ box.

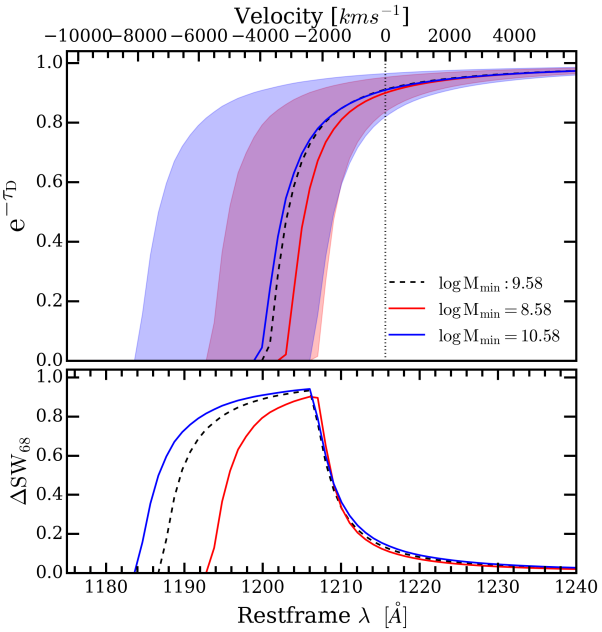


Figure 2.10. Similar to Fig. Fig. 2.4, we show the effect of M_{min} on damping wing signals. For $M_{min} = 10^{10.58} M_\odot$ (blue) where bigger bubbles govern the reionization topology, there is less damping compared to the $M_{min} = 10^{8.58} M_\odot$ (red). In the lower panel, we see that the $M_{min} = 10^{10.58} M_\odot$ has much higher ΔSW_{68} than the $M_{min} = 10^{8.58} M_\odot$.

Finally, in Fig. 2.11 we present a comprehensive overview showcasing the effects of all the model parameters on the damping wings. We have already explored the impact of x_{HI}, t_q, and M_{min} in the previous Sections. For the remaining parameters—α_{star}, α_{esc}, and f_{star}—we do not observe any significant effect on the M(DW) or the scatter. This lack of impact is likely due to these parameters being degenerate with another crucial factor, the escape fraction f_{esc}, which is used in the calculation of the ionizing efficiency of high-redshift galaxies, as expressed in the Eq. (3.25).

The ionizing efficiency ζ, which governs the global ionization state of the intergalactic medium (IGM), is directly influenced by f_{esc}. Since α_{star}, α_{esc}, and f_{star} primarily affect galaxy formation and star formation rates, their impact is indirectly reflected in f_{esc}. Thus, changes in these parameters are absorbed by adjustments in f_{esc}, which controls the total number of ionizing photons escaping into the IGM, leading to a muted effect on the observed damping wings.

In conclusion, while x_{HI}, t_q, and M_{min} play a direct and substantial role in shaping the damping wing profiles, parameters like α_{star}, α_{esc}, and f_{star} being degenerate with f_{esc} for the calculation of x_{HI} do not show any variations in our plots.

2.5 Discussion and Conclusion

The results above have established an understanding of how each parameter affects the M(DW) profile and the sightline-to-sightline scatter of the damping wings around the median. We observed that parameters influencing the global ionization topology, such as x_{HI} and M_{min}, significantly impact the scatter of the damping wings. In contrast, t_q, which alters the local ionization topology, does not substantially affect the overall scatter. In this Section, we aim to further discuss this behavior and explore the possibility of handling the degeneracy between x_{HI} and t_q.

As noted in the previous Section, the damping wing signals vary rapidly with changes in both x_{HI} and t_q, raising the issue of degeneracy in the M(DW). In Fig. 2.4, we observe that variations in the global mean neutral fraction (x_{HI}) have a pronounced effect on both the M(DW) and the scatter. This is expected since any change in x_{HI} necessitates an alteration in the overall reionization topology of the box. However, if the quasar is switched on for a time t_q, it only modifies the neutral fraction within the region described by Equation Eq. (3.26), which is a local phenomenon. Therefore, this local modification contributes minimally to the scatter of longer damping wings, which

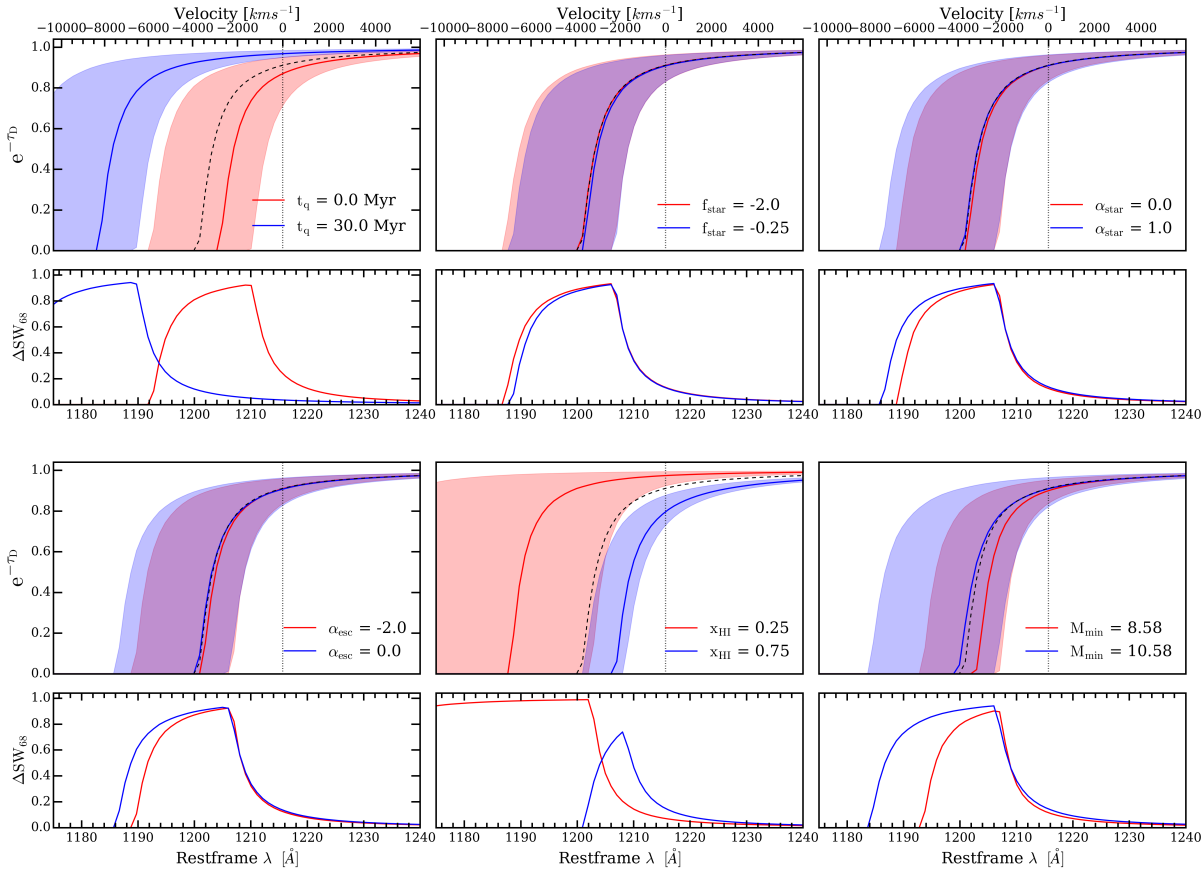


Figure 2.11. A comprehensive overview of the damping wing signals resulting from variations in all the parameters, along with their respective scatter and ΔSW_{68}. x_{HI}, t_q, and M_{min} have a significant effect on both the M(DW) profile and the scatter. The scatter for all the plots is calculated with 10,000 randomly distributed sightlines originating at the halos of mass $M_{qso} \approx 4 \times 10^{11} M_{\odot}$.

extend well beyond the radius of the ionized bubble.

We can potentially use this difference in impact to explore the degeneracy between x_{HI} and t_q. As illustrated in Fig. 2.12, when we match the effective value of x_{HI} to obtain the same M(DW) for different values of t_q, we are unable to recover the scatter. Furthermore, we observe that the scatter induced by x_{HI} evolves much more rapidly than that caused by t_q. Thus, for an ensemble of damping wings, studying the scatter in conjunction with the M(DW) may provide a more robust constraint on both of these parameters.

We also observe that M_{min} has a noticeable effect on the M(DW) signals and a stronger effect on the scatter. As shown in Fig. 2.9, the strong variation in scatter arises from the fact that changes in M_{min} affect the bubble size distribution around the halos. With larger and coarser bubbles, the probability of encountering a neutral island along a random sightline varies significantly with direction. A skewer can travel much longer distances before encountering any neutral patch, thus leading to a strong correlation with

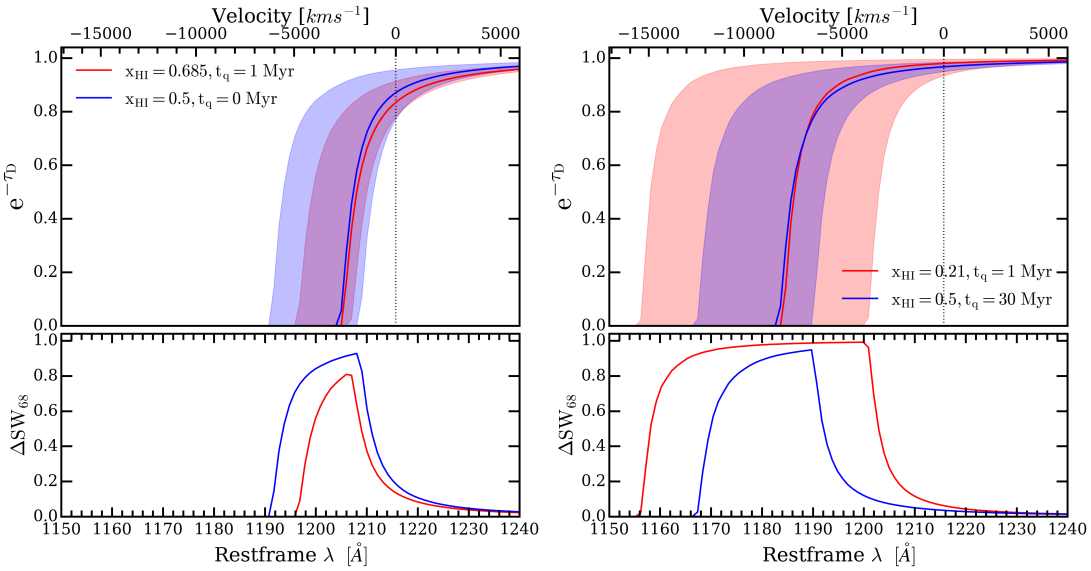


Figure 2.12. The upper panel of the first plot shows the effective value of $x_{HI} = 0.685$ to match the mean signal from $t_q = 0$ Myr. The upper panel of the second plot shows the effective value of $x_{HI} = 0.21$ to match the mean signal from $t_q = 30$ Myr. The shaded region on both plots represents the 68 percentile scatter from the M(DW). The bottom panels show the width of their respective scatter. We see that when we try to match the M(DW), the scatter varies rapidly with x_{HI} as opposed to t_q.

the scatter. This suggests that the statistics of damping wing absorption features may provide a novel probe of the ionizing output of the lowest mass galaxies, which can be regulated both by internal (supernovae, winds) and external (photoionization) feedback processes, as well as the escape fraction of ionizing photons.

Finally, in the Appendix Appendix A.1, we establish that the damping wing profiles computed using a simulation box length of 512 Mpc converge within the 10% error limit with our fiducial parameters, and hence are sufficient for our studies.

With our study, we have shown that the Lyα damping wings are a very powerful tool to not just constrain the neutral fraction (x_{HI}), but also the properties of the source model. We have established that other than x_{HI}, the quasar lifetime (t_q), the mass of the host halo (M_{qso}), and the minimum mass of the halo that can support star formation (M_{min}) can have significant effect on the damping wing profile. We also looked at the sightline-to-sightline scatter of damping wings and showed how this scatter can be used to somewhat break the degeneracy between t_q and x_{HI}. This is because quasar activity primarily alters the reionization topology locally, while the scatter arises from the distribution of HI beyond the edge of the local bubble. In the near future we are expected to see more than just a few $z > 7$ quasars. The Euclid mission is poised to uncover hundreds of luminous high-redshift quasars (Barnett et al., 2019; Schindler et al., 2023). Thus

making this systematic uncertainty not only relevant but possible to measure.

Take away points

- The Lyα damping wings are sensitive to x_{HI}, t_q, M_{min} and M_{qso}.
- x_{HI} changes both the M(DW) and ΔSW_{68} signal
- t_q mainly changes the shape of M(DW), but only shifts the ΔSW_{68}. Thus, this feature can be used to distinguish between the effects of x_{HI} and t_q on damping wing spectra
- M_{min} controls the distribution of bubble sizes among the permitted star-forming halos

3

Fisher Matrix Analysis

The work in this chapter has been submitted and is under review in the AAS Journals at the time of writing this thesis. I carried out all the scientific results presented in this chapter, with the support and guidance provided by the co-author Fred Davies. I also wrote all the texts in this paper, except for the proximity zone section, which was written by co-author Fred Davies.

Abstract

The reionization topology near the epoch depends on a multitude of parameters, namely x_{HI}, M_{min}, t_q, and M_{qso}. The effect on reionization topology as a function of these parameters can be seen through the changes in the median damping wing profile and the sightline-to-sightline scatter ΔSW_{68} of a set of quasar spectra. Combining these two observables can provide strong constraints on these parameters. In this work, we derived the constraints on the abovementioned parameters using the Fisher Information Matrix (FIM). We showed that the constraints provided by only 64 quasars at redshift 7, $x_{HI} = 0.5^{+0.02}_{-0.02}$, $M_{min} = 8.78^{+0.53}_{-0.53}$, $\log t_q/\mathrm{yr} = 6.0^{+0.12}_{-0.12}$, and $\log M_{qso}/M_\odot = 11.52^{+0.32}_{-0.31}$ are comparable to the results from other observables like 21cm signal. We also studied the dependencies of these constraints as a function of the number of quasars, the spectral noise, and the continuum noise. We also looked at the changes in the nature of these constraints at multiple redshifts and luminosities of the sources. We discovered that the overall constraints on x_{HI}, t_q, and M_{min} improve when we go lower in redshift or look at the fainter sources due to the decreasing effects of proximity zones.

Contents

3.1 Introduction

The previous chapters paint the picture of how Lyα damping wings are a crucial resource to study the universe during the epoch of reionization. While a good amount of literature exists, where people tried to constrain x_{HI} and t_q using a single quasar damping wing spectrum (Kist et al., 2024). We showed in Chapter 2 how the ensemble of damping wings can also be used to effectively study the source parameters, namely M_{min} (minimum mass of the halo that can support star formation) and M_{qso} (quasar's host halo mass). We argued that the quasar's sightline-to-sightline variation within the ensemble, together with median damping wing spectra, is a potential source to effectively constrain and help relax the degeneracy between all our selected parameters. Hence, in this chapter, we set out to quantitatively analyze the spectra of an ensemble of damping wings to derive the constraints on our astrophysical parameters filtered from the analysis of Chapter 2. We adopt the Fisher-matrix or FIM analysis to study the constraining power of the median damping wing profile for an ensemble of quasars (M(DW)), and 68-percentile sightline-to-sightline scatter width around the median (ΔSW_{68}) over the aforementioned set of astrophysical parameters.

To understand how this analysis works, we will first set up the theoretical framework of the Fisher Information and derive the intuition behind its ability to constrain the model parameters based on the observational data.

In §3.2, we derive the form of FIM we used in our calculations, and provide basic intuitions behind the working of FIM. In §3.3, we briefly describe the models for the filtered parameter space that we will focus on in this work. In §3.4, we derive the

required observables for the calculations of FIM from the ensemble of damping wings resulting from the grid of models presented in Chapter 2. Then, in §3.5, we show the constraints we get from our analysis, how robust these constraints are, and discuss the nature of these constraints. In §3.6, we explore how the constraints obtained above vary with factors such as the number of quasars ($N_{quasars}$), spectral noise, and continuum noise. We use this information to calculate the lower bound on $N_{quasars}$ required to conduct our analysis successfully. We assume a flat ΛCDM cosmology throughout our work, based on the results from *Planck* (Aghanim et al., 2020), with cosmological parameters $h = 0.676$, $\Omega_m = 0.309$, $\Omega_b = 0.0489$, and $\sigma_8 = 0.810$.

3.2 Fisher Information Formalism

The FIM or simply Fisher Matrix is a statistical tool that helps determine the amount of information any set of data contains about the model parameter, or how effectively an experiment or observation can constrain model parameters. It is beneficial for predicting the precision of future experiments or observations while they are still in the design phase. Let us begin by deriving the relationship between the likelihood function and the FIM to get the intuition behind using FIM to constrain the model parameters.

3.2.0.1 Curvature of Likelihood function

Let $\mathcal{L}(\lambda)$ be the likelihood function of our data set, modeled by some parameter λ. We can Taylor expand our likelihood function around some fiducial value (λ_0) of the parameter for which the likelihood function assumes maxima. Then,

$$\mathcal{L}(\lambda) = \mathcal{L}(\lambda_0) + \frac{\partial \mathcal{L}}{\partial \lambda}\bigg|_{\lambda_0} (\lambda - \lambda_0) + \frac{1}{2!}\frac{\partial^2 \mathcal{L}}{\partial \lambda^2}\bigg|_{\lambda_0} (\lambda - \lambda_0)^2 + \quad ... \tag{3.1}$$

The first derivative is zero since the likelihood is maximum at λ_0. If we ignore the higher-order terms, we can rewrite the above equation as,

$$\mathcal{L}(\lambda) \approx \mathcal{L}(\lambda_0) + \frac{1}{2}\frac{\partial^2 \mathcal{L}}{\partial \lambda^2}\bigg|_{\lambda_0} (\lambda - \lambda_0)^2 \tag{3.2}$$

The above equation implies that the likelihood function has a parabolic shape as a function of λ. Instead of looking at the likelihood function, we can rather look at the log of the likelihood function, as it also peaks at the same value. In that case, the shape of the likelihood function resembles a Gaussian distribution. Thus,

$$\ln \mathcal{L}(\lambda) \approx \ln \mathcal{L}(\lambda_0) + \frac{1}{2} \frac{\partial^2 \ln \mathcal{L}}{\partial \lambda^2}\bigg|_{\lambda_0} (\lambda - \lambda_0)^2 \tag{3.3}$$

We can simplify the above equation to,

$$\mathcal{L}(\lambda) = N_0 \; e^{\frac{1}{2} \frac{\partial^2 \ln \mathcal{L}}{\partial \lambda^2}\Big|_{\lambda_0} (\lambda-\lambda_0)^2}. \tag{3.4}$$

Let's call $\mathcal{F} = -\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda^2}\Big|_{\lambda_0}$ and compare the relation Eq. (3.4) with a Gaussian distribution $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, with a mean value ($\mu$) of the parameter at λ_0, then we see that the inverse of $\mathcal{F}$ represents the variance (σ^2) (or uncertainty) of the parameter λ.

$\mathcal{F}$ represents nothing but the curvature (Hessian for a more general case) of the log-likelihood function, which is equivalent to the inverse of the variance/uncertainty of the parameter λ (Fisher, 1925). Thus, the higher the $\mathcal{F}$ value, the lower the uncertainty in our parameter estimation. We will later try to derive the formal definition of the Fisher matrix and compare it with the result in the Eq. (3.4) to complete our picture.

3.2.0.2 General Form of Fisher Information Matrix

We can generalize the above equations for a multivariate Gaussian, for a set of parameters $\lambda_\alpha = \{\lambda_1, \lambda_2 ... \lambda_r\}$. In that case, under some regularity conditions (Schervish, 2012) (refer to the appendix E of Ly et al., 2017 for a discussion on the regularity conditions and see §3.2.0.3), for an observable $\mathcal{O} \equiv \mathcal{O}(\lambda_\alpha)$, we can define the FIM as follows,

$$\begin{aligned} \mathcal{F}_{ij} &= -\int_{\mathcal{O}} \left(\frac{\partial^2}{\partial \lambda_i \partial \lambda_j} \ln f(\mathcal{O}|\lambda_\alpha) \right) p_{\lambda_\alpha}(\mathcal{O}) d\mathcal{O} \\ &= -E\left[\frac{\partial^2}{\partial \lambda_i \partial \lambda_j} \ln f(\mathcal{O}|\lambda_\alpha) \Big| \lambda_\alpha \right]. \end{aligned} \tag{3.5}$$

Here $f(\mathcal{O}|\lambda_\alpha)$ is any probability distribution function that is used to estimate the parameters λ_α, for any data-set $\mathcal{O}$, in the Eq. (3.4) it was the likelihood function. The integral (expectation value) is done over different realizations of data. The above equation represents a more general representation of the FIM. As discussed above, the inverse of the FIM now will represent the covariance matrix of all our parameters. According to the Cramér-Rao theorem, the inverse of FIM is the lower bound to the covariance matrix provided by any unbiased estimator, i.e., FIM provides the minimum

uncertainties on the estimation of parameters, given a data set (observations). The derivation of the Cramér-Rao theorem and the form of FIM used in the Eq. (3.5) are discussed later in the §3.2.0.3. Let's continue and derive the FIM using χ^2 as our log-likelihood function. The χ^2 is given as

$$\ln \mathcal{L}(\lambda_\alpha) = \ln f(\mathcal{O}|\lambda_\alpha) \propto -\frac{\chi^2(\lambda_\alpha)}{2} = \sum_{k=1}^{N} \frac{1}{2} \left(\frac{\mathcal{O}_k(\lambda_\alpha) - \mathcal{O}_k^{obs}}{\sigma_k} \right)^2 \tag{3.6}$$

Here, $\mathcal{O}(k, \lambda_\alpha)$ are the theoretical values generated from our modeling, $\mathcal{O}^{obs}(k)$ is the observed value, and σ_i is the error/ uncertainty in the observed values. Let's now calculate the Fisher information of the χ^2 for a simple one-parameter case.

$$\begin{aligned} \frac{1}{2}\frac{\partial^2 \chi^2}{\partial \lambda^2} &= \sum_{k=1}^{N} \frac{1}{2} \frac{\partial^2}{\partial \lambda^2} \left(\frac{\mathcal{O}_k(\lambda) - \mathcal{O}_k^{obs}}{\sigma_k} \right)^2 \\ &= \sum_{k=1}^{N} \frac{\partial}{\partial \lambda} \left(\frac{\mathcal{O}_k(\lambda) - \mathcal{O}_k^{obs}}{\sigma_k^2} \right) \frac{\partial \mathcal{O}_k(\lambda)}{\partial \lambda} \\ &= \sum_{k=1}^{N} \frac{1}{\sigma^2} \left(\frac{\partial \mathcal{O}_k(\lambda)}{\partial \lambda} \right)^2 + \left(\frac{\mathcal{O}_k(\lambda) - \mathcal{O}_k^{obs}}{\sigma_k^2} \right) \frac{\partial^2 \mathcal{O}_k(\lambda)}{\partial \lambda^2} \end{aligned} \tag{3.7}$$

Now, taking the expectation value of the above equation over the realizations of the observed data $\mathcal{O}_k^{obs}$, where the mean value of the observations is $\mathcal{O}_k(\lambda)$. We see that the second term in the Eq. (3.7) goes to zero. Thus resulting in to

$$\left\langle \frac{1}{2}\frac{\partial^2 \chi^2}{\partial \lambda^2} \right\rangle = \sum_{k=1}^{N} \left\langle \frac{1}{\sigma_k^2} \left(\frac{\partial \mathcal{O}_k(\lambda)}{\partial \lambda} \right)^2 \right\rangle \tag{3.8}$$

Now generalizing for a set of parameters $\lambda_\alpha = \{\lambda_1, \lambda_2 ... \lambda_r\}$, the FIM can now be written as,

$$\mathcal{F}_{ij} = \left\langle \frac{1}{2}\frac{\partial^2 \chi^2}{\partial \lambda_i \lambda_j} \right\rangle = \sum_{k=1}^{N} \left\langle \frac{1}{\sigma^2} \frac{\partial \mathcal{O}_k(\lambda)}{\partial \lambda_i} \frac{\partial \mathcal{O}_k(\lambda)}{\partial \lambda_j} \right\rangle \tag{3.9}$$

The summation here is over all the observables $\mathcal{O}_k$ ($\equiv \mathcal{O}_k(\lambda)$); for our case, this is the sum over all the pixels of the damping wings and scatter width signals (ΔSW_{68}). In Eq. (3.9), these observations are assumed to be uncorrelated, but in reality, the pixels of our signals can have a correlation. In that case, we can re-write the above equation in the more general form (Tegmark, 1997), by replacing the single uncertainties with a full covariance matrix $\mathbf{C}$ of our data and $(\mathcal{O}_1, \mathcal{O}_2, ..., \mathcal{O}_N)$ by the vector $\hat{\mathcal{O}}$. Thus re-writing the Eq. (3.9) as

$$\mathcal{F}_{ij} = \frac{1}{2}\mathrm{tr}\left[\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \lambda_i}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \lambda_j}\right] + \left\langle \frac{\partial \hat{\mathcal{O}}^T}{\partial \lambda_i}\mathbf{C}^{-1}\frac{\partial \hat{\mathcal{O}}}{\partial \lambda_j}\right\rangle \tag{3.10}$$

For our work, we assume that the covariance matrix $\mathbf{C}$ does not vary with parameters. Hence, we can ignore the first term in the above equation and write the general correlated FIM as:

$$\mathcal{F}_{ij} = \left\langle \frac{\partial \hat{\mathcal{O}}^T}{\partial \lambda_i}\mathbf{C}^{-1}\frac{\partial \hat{\mathcal{O}}}{\partial \lambda_j}\right\rangle. \tag{3.11}$$

This is the form of FIM that we will use in our calculations. We can then derive the uncertainties in our parameter estimation by taking its inverse.

3.2.0.3 Alternative Expression of Fisher Information Matrix

In this section, we will derive the form used in Eq. (3.5) from the very definition of the score function and the original definition of FIM, being the expectation value of the square of the score function[1].

The score function in statistics is defined as the derivative of the log-likelihood function with respect to the model parameters. Mathematically, it is written as,

$$s(\lambda; \mathcal{O}) \equiv \frac{\partial \ln \mathcal{L}(\lambda; \mathcal{O})}{\partial \lambda} \tag{3.12}$$

The score function represents the steepness of the log-likelihood function evaluated at the fiducial value of the parameter vector. Here $\mathcal{O}$ represents the observable vector described earlier, and λ is the parameter vector. As discussed in §3.2.0.1, the score function vanishes at the maximum (or minimum) value of the log-likelihood function. Now, let's calculate the expectation value of the score function,

$$\begin{aligned} E\left[\frac{\partial \ln \mathcal{L}(\lambda; \mathcal{O})}{\partial \lambda}\right] &= E\left[\frac{1}{\mathcal{L}(\lambda; \mathcal{O})}\frac{\partial \mathcal{L}(\lambda; \mathcal{O})}{\partial \lambda}\right] \\ &= \int_{\mathcal{O}} \left(\frac{1}{\mathcal{L}(\lambda; \mathcal{O})}\frac{\partial \mathcal{L}(\lambda; \mathcal{O})}{\partial \lambda}\right) \mathcal{L}(\lambda; \mathcal{O}) d\mathcal{O} \end{aligned} \tag{3.13}$$

A likelihood function is just another probability distribution function of the data points at a given parameter vector. When maximizing the likelihood function, we

[1] A part of this derivation follows from the lecture notes on Cramér-Rao Bound from Harvard Business School.

usually treat it as a function of the model parameter for fixed data points. However, thinking of likelihood as a function of the model parameter and the data points is more useful. Thus, we can calculate the expectation values using the likelihood function as the probability distribution function. The expectation value is calculated for the sample set of observables within a given model parameter vector. That being set, continuing the above derivation, we get,

$$E\left[\frac{\partial \ln \mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda}\right] = \int_{\mathcal{O}} \frac{\partial \mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda} d\mathcal{O} \tag{3.14}$$

If the likelihood function is well-behaved (i.e., it is continuous and differentiable within its domain, refer to Appendix E of Ly et al., 2017 for more rigorous treatment of the regularity conditions), we can interchange the order of integral and derivative (see Leibniz integral rule), giving us

$$\begin{aligned} E\left[\frac{\partial \ln \mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda}\right] &= \frac{\partial}{\partial\lambda}\int_{\mathcal{O}} \mathcal{L}(\lambda;\mathcal{O}) d\mathcal{O} \\ &= \frac{\partial}{\partial\lambda}(1) = 0 \end{aligned} \tag{3.15}$$

Thus, the expectation value of the score function is 0. Since the score function is the gradient of the log-likelihood function, the further the value of the log-likelihood function from the maximum value (which is the 0 of the score function), the higher the value of the score function, while the sign depends upon if the parameter value is smaller or bigger than the most likely value. If we draw a physical analogy where the negative log-likelihood function is some potential, where the parameter values tend to be closer to the minimum (maximum for log-likelihood), then the gradient of log-likelihood of it represents the force that the data exerts on the model parameters to bring it closer to the data itself. Since the expectation value of the score function is 0, it is more useful to look at its absolute magnitude or square. The expectation value of the square of the score function (with zero mean) is nothing but the variance. This expectation value is called the Fisher Information.

$$I(\lambda) = E[(s(\lambda;\mathcal{O}))^2] = E\left[\left(\frac{\partial \ln \mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda}\right)^2\right] = Var\left[\frac{\partial \ln \mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda}\right] \tag{3.16}$$

Using the above definition of Fisher Information, we want to get the relation we used in Eq. (3.5). To do so, let us look at the second derivative of the log-likelihood function,

$$\begin{aligned}\frac{\partial^2 \ln \mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda^2} &= \frac{\partial}{\partial\lambda}\left[\frac{\partial \ln \mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda}\right] = \frac{\partial}{\partial\lambda}\left[\frac{1}{\mathcal{L}(\lambda;\mathcal{O})}\frac{\partial\mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda}\right] \\ &= -\frac{1}{\mathcal{L}^2(\lambda;\mathcal{O})}\left[\frac{\partial\mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda}\right]^2 + \frac{1}{\mathcal{L}(\lambda;\mathcal{O})}\frac{\partial^2\mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda^2} \\ &= -\left[\frac{\partial \ln \mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda}\right]^2 + \frac{1}{\mathcal{L}(\lambda;\mathcal{O})}\frac{\partial^2\mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda^2}\end{aligned} \tag{3.17}$$

Now, taking the expectation value,

$$E\left[\frac{\partial^2 \ln \mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda^2}\right] = -E\left[\left[\frac{\partial \ln \mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda}\right]^2\right] + E\left[\frac{1}{\mathcal{L}(\lambda;\mathcal{O})}\frac{\partial^2\mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda^2}\right] \tag{3.18}$$

Again, if the log-likelihood is well-behaved, we can interchange the order of the derivative and the integral from the second term on the right-hand side of the above equation.

$$\begin{aligned}E\left[\frac{1}{\mathcal{L}(\lambda;\mathcal{O})}\frac{\partial^2\mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda^2}\right] &= \int_{\mathcal{O}}\frac{1}{\mathcal{L}(\lambda;\mathcal{O})}\frac{\partial^2\mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda^2}\mathcal{L}(\lambda;\mathcal{O})d\mathcal{O} \\ &= \int_{\mathcal{O}}\frac{\partial^2\mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda^2}d\mathcal{O} \\ &= \frac{\partial^2}{\partial\lambda^2}\int_{\mathcal{O}}\mathcal{L}(\lambda;\mathcal{O})d\mathcal{O} = 0\end{aligned}$$

Thus, the Eq. (3.18) can be finally written as,

$$E\left[\frac{\partial^2 \ln \mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda^2}\right] = -E\left[\left[\frac{\partial \ln \mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda}\right]^2\right] \tag{3.19}$$

Thus proving the alternate form of FIM used in the Eq. (3.5).

3.2.0.4 Cramér-Rao Bound

Before we prove the Cramér-Rao bound, let us recall some properties of the random variables;

- The Covariance matrix of two random variables, X and ,Y is given as: $Cov(X,Y) = E[XY] - E[X][Y]$, if Either of $E[X]$ or $E[Y]$ is 0, then $Cov(X,Y) = E[XY]$
- The correlation between two random variables is given by $\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}$, and $|\rho_{X,Y}| \leq 1$

Let's consider an unbiased estimator $U(\mathcal{O})$ of the parameter λ, which is the function of the data/ observable $\mathcal{O}$ only. Using the definition of the score function $s(\lambda;\mathcal{O})$ from Eq. (3.12), we first calculate the covariance function between the two.

$$\begin{aligned} Cov(U(\mathcal{O}), s(\lambda;\mathcal{O})) &= E[U(\mathcal{O})s(\lambda;\mathcal{O})] \\ &= E\left[U(\mathcal{O})\frac{\partial \ln \mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda}\right] \\ &= \int_{\mathcal{O}} \left(U(\mathcal{O})\frac{\partial \ln \mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda}\right)\mathcal{L}(\lambda;\mathcal{O})d\mathcal{O} \\ &= \int_{\mathcal{O}} \left(\frac{U(\mathcal{O})}{\mathcal{L}(\lambda;\mathcal{O})}\frac{\partial\mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda}\right)\mathcal{L}(\lambda;\mathcal{O})d\mathcal{O} \\ &= \int_{\mathcal{O}} U(\mathcal{O})\frac{\partial\mathcal{L}(\lambda;\mathcal{O})}{\partial\lambda}d\mathcal{O} \end{aligned} \tag{3.20}$$

Since $U(\mathcal{O})$ is the function of the data only, we can bring the derivative out of the integral,

$$\begin{aligned} Cov(U(\mathcal{O}), s(\lambda;\mathcal{O})) &= \frac{\partial}{\partial\lambda}\int_{\mathcal{O}} U(\mathcal{O})\mathcal{L}(\lambda;\mathcal{O})d\mathcal{O} \\ &= \frac{\partial}{\partial\lambda}E[U(\mathcal{O})] \end{aligned} \tag{3.21}$$

Since, $U(\mathcal{O})$ is an unbiased estimator of the parameter λ, the expectation value of $U(\mathcal{O})$ is nothing but λ $E[U(\mathcal{O})] = \lambda$. Thus, the covariance function becomes

$$Cov(U(\mathcal{O}), s(\lambda;\mathcal{O})) = \frac{\partial}{\partial\lambda}E[U(\mathcal{O})] = \frac{\partial\lambda}{\partial\lambda} = 1 \tag{3.22}$$

Now, let's look at the correlation function between $U(\mathcal{O})$ and $s(\lambda;\mathcal{O})$,

$$\begin{aligned} (\rho(U(\mathcal{O}), s(\lambda;\mathcal{O})))^2 &= \frac{(Cov(U(\mathcal{O}), s(\lambda;\mathcal{O}))^2)}{Var(U(\mathcal{O})Var(s(\lambda;\mathcal{O}))} \\ Var(U(\mathcal{O})) &= \frac{1}{(\rho(U(\mathcal{O}), s(\lambda;\mathcal{O})))^2 Var(s(\lambda;\mathcal{O}))} \\ Var(U(\mathcal{O})) &\geq \frac{1}{Var(s(\lambda;\mathcal{O}))} \\ Var(U(\mathcal{O})) &\geq \frac{1}{I(\lambda)} \end{aligned} \tag{3.23}$$

Recalling the fact that $|\rho_{X,Y}| \leq 1$ and the variance of score function is Fisher information $I(\lambda)$ (Eq. (3.16)), the above equation proves the fact that the inverse of the Fisher Information provides the lower bound on the uncertainties produced by any unbiased estimator.

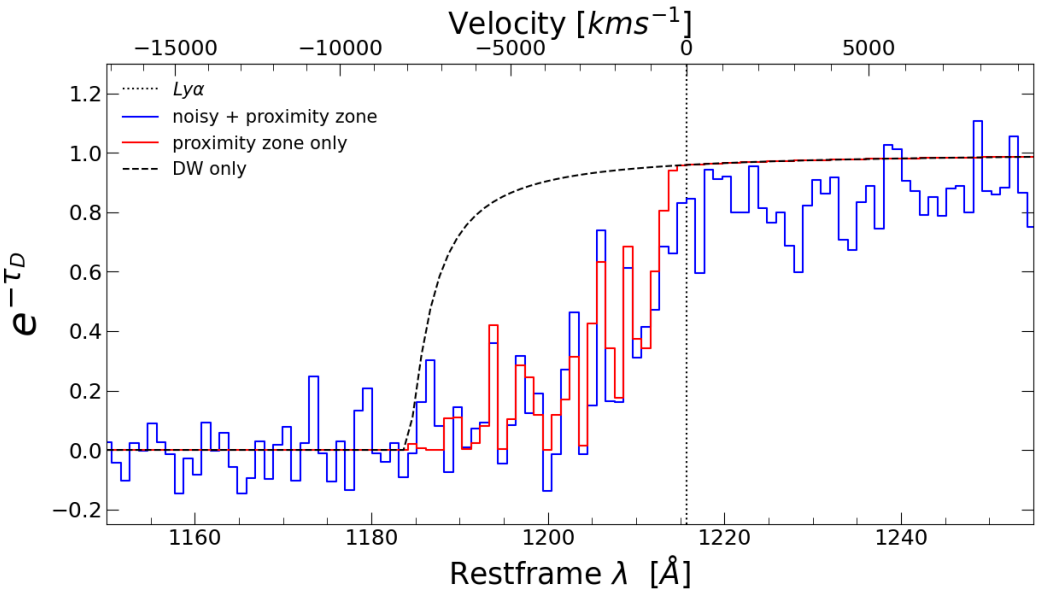


Figure 3.1. In this figure, we show how a typical damping wing signal (black dashed curve) is modified by the addition of a proximity zone (red curve) and by both a proximity zone and noise (blue curve)

3.3 Model Setup

This section will first define the parameter space we filter after studying the damping wings in Chapter 2. Then, we will set up the formalism for the Fisher Matrix or FIM calculations. We will derive the form of FIM that is convenient for our studies of the astrophysical parameters and look at the physical understanding of it. We will also look at an interpretation of the FIM that relates to the covariance matrix of the parameters.

3.3.1 Parameter Space

In this work, we will continue the results of Chapter 2 where we filtered the parameters based on their impact on the median damping wing signal and the sightline-to-sightline scatter. The final list of parameters is the global mean neutral fraction x_{HI}, the quasar activity lifetime t_q, the minimum mass of the halos that can support star formation M_{min} and the mass of the quasar's host halo M_{qso}. The fiducial values and ranges of these parameters are shown in Table 3.1.

1. **The Global Mean Neutral Fraction** — x_{HI}: This parameter represents the mean of the total amount of neutral hydrogen in our simulation boxes. In 21cmFAST, individual pixels are flagged as ionized if they meet the following condition

(Bradley Greig and Mesinger, 2018):

$$\zeta f_{\rm coll}(x, z, R, {\rm M_{min}}) \geq 1 \tag{3.24}$$

In this equation, R is the filter scale at which our excursion set formalism works, $\rm f_{coll}$ denotes the fraction of collapsed matter residing within halos in the region R with a mass exceeding $\rm M_{min}$ (Press and Schechter, 1974), (Bond et al., 1991), (Lacey and Cole, 1993), (Sheth and Tormen, 1999). Effectively, $\rm M_{min}$ establishes the minimum mass of halos capable of supporting star formation. The parameter ζ represents the UV ionizing efficiency of dark matter halos within the filter size, as described by (Bradley Greig and Mesinger, 2017):

$$\zeta({\rm M_h}) = 30 \left(\frac{\rm f_{esc}}{0.12}\right) \left(\frac{\rm f_{star}}{0.05}\right) \left(\frac{\rm N_{\gamma/b}}{4000}\right) \left(\frac{1.5}{1 + \rm n_{rec}}\right), \tag{3.25}$$

Where $\rm f_{esc}$ controls the fraction of ionizing photons escaping from star-forming galaxies into the intergalactic medium (IGM). $\rm f_{star}$ represents the fraction of galactic baryons that are converted into stars, $\rm N_{\gamma/b}$ is the number of ionizing photons produced per baryon in the stars, and $\rm n_{rec}$ indicates the number of times a hydrogen atom recombines. For this specific scenario regarding the IGM, we can neglect $\rm n_{rec}$ in the last term.

We can now calibrate the simulation boxes to achieve the desired $\rm x_{HI}$ by adjusting the value of ζ, by tuning the value of $\rm f_{esc}$, as noted in the Chapter 2. For our fiducial case at redshift $\rm z = 7$, we set $\rm x_{HI} = 0.5$. We also run models for different redshifts, namely, $\rm z = 6.5$ with $\rm x_{HI} = 0.35$ and $\rm z = 6$ with $\rm x_{HI} = 0.15$.

2. **The Quasar Activity Lifetime** — $\rm t_q$: Similar to the approach used in the Chapter 2, we represent the effects of $\rm t_q$ by carving out an ionized bubble around the halo, with the radius depending on the quasar's lifetime. Assuming our selected halo hosts a quasar with lifetime $\rm t_q$, it carves out an ionized bubble with a radius of $R(\rm t_q)$ around it. We assume a constant ionizing photon emission rate of $\dot{N}_{ph} \simeq 10^{57}\rm s^{-1}$ (Daniel J Mortlock et al., 2011). The radius of this bubble is given by (Paul R Shapiro and Mark L Giroux, 1987; Cen and Haiman, 2000):

$$R_{(\rm t_q)} = \left(\frac{3\dot{N}_{ph}\rm t_q}{4\pi\langle n_{\rm H}\rangle}\right)^{1/3}, \tag{3.26}$$

Where $\langle n_{\rm H} \rangle$ is the average number density of neutral hydrogen within the sphere. Although this equation assumes homogeneous reionization, it can still be used to calculate the expansion of the ionized bubble along the line of sight. Another consideration is that while this equation assumes a fixed value of t_q, in reality, quasar lifetimes can vary even at a fixed redshift. To account for this variability, we used a log-normal distribution of quasar lifetimes, with the mean value representing the required t_q for the model and a fixed standard deviation of 0.8 for all the values of t_q (Morey et al., 2021; Eilers et al., 2021; Khrykin et al., 2021).

3. **The Minimum Mass Of Halos That Support Star Formation** — M_{min}: M_{min} defines the minimum mass of halos capable of supporting star formation, acting as a threshold below which star formation becomes inefficient. In 21cmFAST simulations, this suppression is estimated using a redshift-independent duty cycle for a given halo mass M_h (Park et al., 2019):

$$f_{\rm duty}(M_h) = \exp\left(-\frac{M_{\rm min}}{M_h}\right). \tag{3.27}$$

Thus, for halos with mass approximately equal to M_{min}, only a fraction f_{duty} can form stars with an efficiency of f_{star}. M_{min} and x_{HI} together control the distribution of ionized bubble sizes within our simulation boxes.

4. **The Mass Of The Quasar's Host Halo** — M_{qso}: Once the density and density fields are established, we utilize the method from (Mesinger and S. Furlanetto, 2007) to identify dark matter halos using excursion set theory. The positions of these halos are then perturbed using velocity fields calculated via linear perturbation theory to obtain the corrected locations at the desired redshift. While the ionized field we derive is based solely on the density field, we use this halo catalog to place our quasars. This strategy allows us to include lower-mass halos in our damping wing analysis, which would otherwise require much higher particle resolution. Since quasars are likely to reside within massive halos, we select M_{qso} $\sim 3.3 \times 10^{11} M_\odot$ as our fiducial value to ensure a sufficient number of massive halos are present.

Parameter	Fiducial model	$\sigma(N_{quasars} = 64)$	Parameter type
x_{HI}	0.5	$0.5^{+0.02}_{-0.02}$	IGM parameter
$\log 10M_{min}/M_\odot$	8.78	$8.78^{+0.49}_{-0.49}$	Source parameter
$\log t_q/yr$	1 Myr	$6.0^{+0.10}_{-0.09}$	Quasar parameter
$\log M_{qso}/M_\odot$	11.52	$11.52^{+0.24}_{-0.24}$	Quasar parameter

Table 3.1. Astrophysical Parameters Space: This table lists all the parameters used in our study, along with their fiducial values, the constraints obtained after FIM analysis for $N_{quasars}$ = 64, and at redshift z = 7, and parameter types.

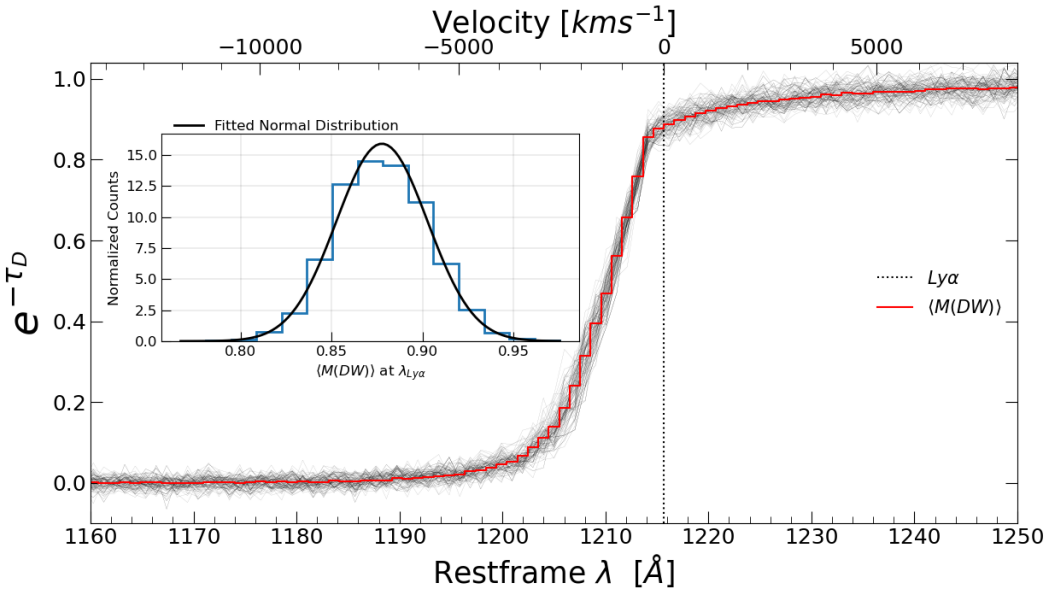


Figure 3.2. This figure show the median damping wing signal (M(DW)) from each sample and its distribution. The red curve in the main plot shows the expectation value of M(DW) profiles and the thin black lines show the distribution of the M(DW) for 100 random samples. In the mini-plot we show the distribution of the M(DW) at a specific pixel (at Lyα). The solid black line on the mini-plot is the fitted normal distribution over the distribution of M(DW).

3.3.2 Proximity Zone

The quasar proximity zone is the region surrounding the quasar where the intense radiation from the quasar ionizes the IGM, influencing the nearby topology. To extract constraints from the damping wing signal blueward of rest-frame Lyα, we must also model the proximity zone Lyα transmission, which was originally neglected in Chapter 2. As our goal is to construct an efficient forecasting framework, we opt for a simplified approach compared to the radiative transfer simulations employed by recent works (e.g. (Frederick B. Davies et al., 2018; Ďurovčíková et al., 2024; Kist et al., 2024)). We instead use a modified version of the hydrodynamical simulation-calibrated approach

from the appendix of Frederick B. Davies et al., 2020, shown to reproduce the statistics of proximity zone sizes at $z \sim 6$.

In this method, the mean profile of the Lyα transmission is computed assuming a relationship between the effective optical depth $\tau_{\rm eff} \equiv -\ln \bar{F}$, where $\bar{F}$ is the mean Lyα transmission, and the strength of the ionizing flux from the quasar. We recalibrate the relationship from Frederick B. Davies et al., 2020 using skewers from the same hydrodynamical simulation employed in their analysis, but now including snapshots at $z = 7$ and $z = 6.5$ as well as at $z = 6$; we direct the reader to Frederick B. Davies et al., 2020 for details. We find that the relationship between the Lyα optical depth and the photoionization rate $\Gamma_{\rm HI}$ closely follows a power-law relation,

$$\tau_{\rm Ly\alpha} = \tau_0 \times \left(\frac{\Gamma_{\rm HI}}{2.5 \times 10^{-13}\,{\rm s}^{-1}} \right)^{-0.55}, \tag{3.28}$$

where the best-fit normalization factor τ_0 is 9.3, 7.5, and 5.7 at $z = 7.0$, $z = 6.5$, $z = 6.0$, respectively. For our fiducial ionizing quasar luminosity and the average quasar spectral energy distribution from (Lusso et al., 2015), the photoionization rate scales with comoving distance R from the quasar as

$$\Gamma_{\rm HI} = 1.85 \times 10^{-11}\,{\rm s}^{-1} \left(\frac{\dot{N}_{\rm ion}}{10^{57}\,{\rm s}} \right) \left(\frac{R}{1+z} \right)^{-2}, \tag{3.29}$$

where, for simplicity, we neglect the contribution from the UV background as the quasar's ionization radiation always dominates inside the proximity zone.

The stochasticity of Lyα absorption is then accounted for by drawing τ_i from a random distribution centered on $\tau_{\rm Ly\alpha}(R)$. Similar to Frederick B. Davies et al., 2020, from the hydrodynamical simulation skewers we find that a lognormal distribution with $\sigma_{\ln \tau} = 0.7$ provides a good fit to the distribution of $\tau_{\rm Ly\alpha}$ on our 1 Mpc pixel scale (see also George D. Becker et al., 2007).

3.4 Data Sampling

As established, the FIM can be utilized to derive the lowest possible estimates for the uncertainties of our parameters based on a given set of observations. This section will discuss how we derive the set of observables required for FIM calculations from the models previously presented in §3.3.

To summarize §3.3, we generate a suite of astrophysical parameters dependent suite of patchy reionization models using 21cmFAST (Mesinger et al., 2011; Murray et al., 2020;

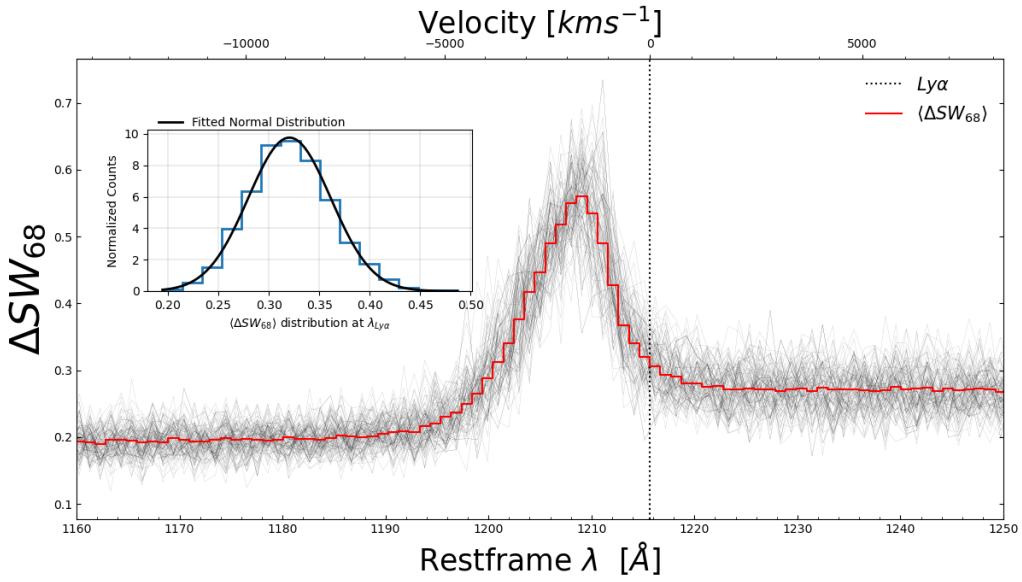


Figure 3.3. In this figure, we show the scatter width signal (ΔSW$_{68}$) from each sample and its distribution. The red curve in the main plot shows the expectation value of ΔSW$_{68}$ profiles and the thin black lines show the distribution of the ΔSW$_{68}$ for 100 random samples. In the mini-plot we show the distribution of the ΔSW$_{68}$ at a specific pixel (at Lyα). The solid black line on the mini-plot is the fitted normal distribution over the distribution of ΔSW$_{68}$.

Muñoz et al., 2022; Qin et al., 2020; Park et al., 2019) over large cosmological volumes (512 cMpc^3) at z = 7. We identify halos within the desired mass bin in the simulation box and randomly select a sample of 10,000 halo-sightline pairs. The damping wing optical depth along each sightline is then calculated to obtain our damping wing signal. This process yields an ensemble of 10,000 damping wing profiles across the entire box, serving as a probe of reionization topology over the whole box. We then add the proximity zone, followed by spectral and continuum noises to our dataset. For our analysis, we define the median signal of this ensemble of damping wings of a specific mass (M(DW)) and the 68-percentile scatter about the median signal, referred to as the scatter width (ΔSW$_{68}$), as our observables for the FIM.

In this work, we aim to assess the constraining power of M(DW) and ΔSW$_{68}$ on the parameters summarized in the Table 3.1 using the FIM, and to forecast the minimum number of quasars required for a robust analysis. To achieve this, we first modify our damping wing signals to create a more realistic mock spectrum.

We begin by incorporating the proximity zone into all damping wing signals, as discussed in sub §3.3.2. After this, we introduce noise into our dataset of damping wings by first adding normally distributed spectral noise to each sightline, with a mean of 0

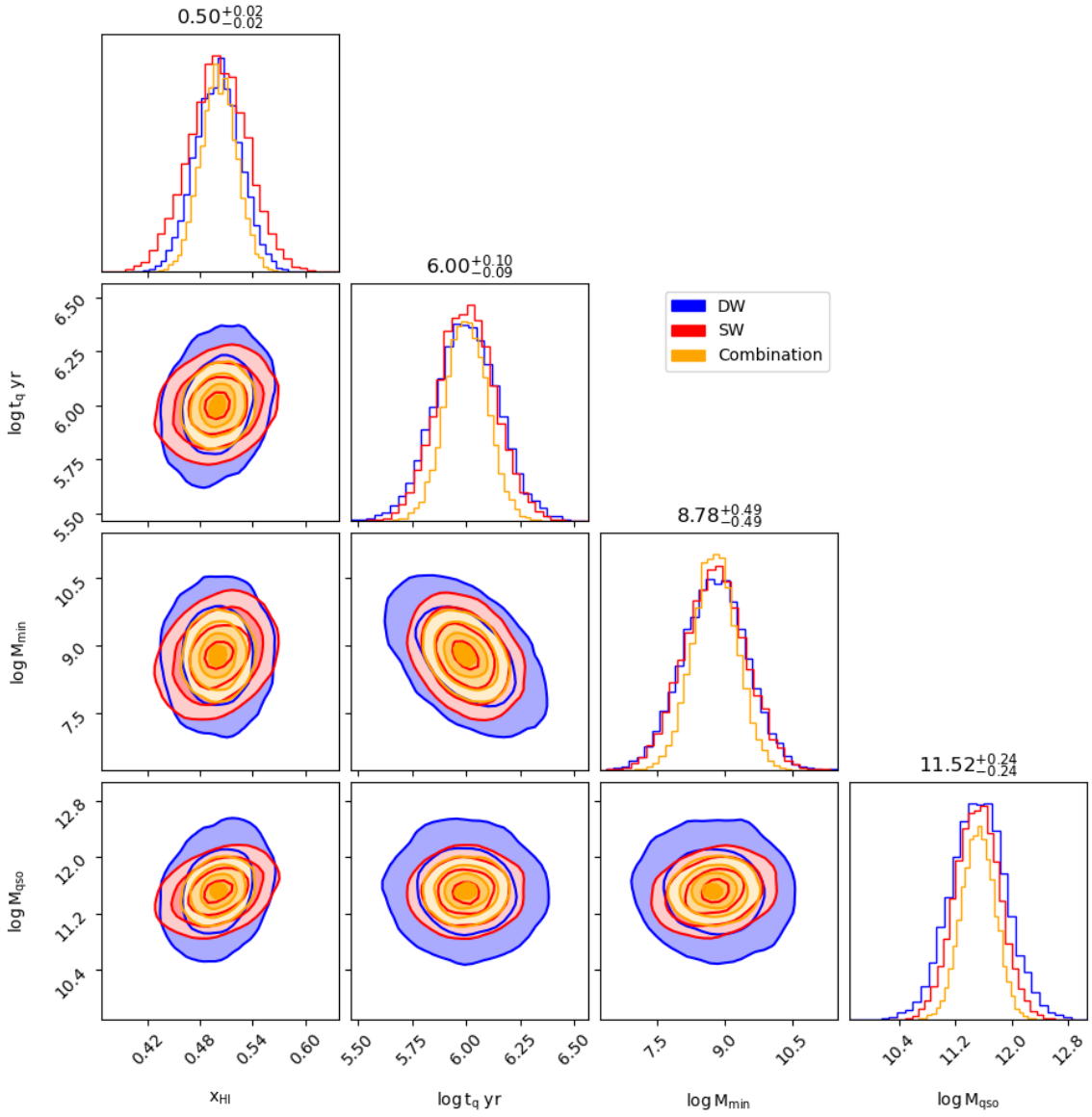


Figure 3.4. In this figure, we plot the overlaying corner plot for three different observables, with the parametric constraints stated in the Table 3.1. From the top left/ along the diagonal plots show the 1D marginalized distributions, which show the uncertainties on the individual parameter. The off-diagonal plots show the 2D marginalized distributions, showing the correlation between a given combination of parameters. The contours show $1, 2\sigma$ confidence intervals. The blue contours represent 1D, 2D distributions with M(DW) as the only observable; the red contours represent the same for ΔSW_{68} as the only observable, while the orange contours represent the combination of both the M(DW) and ΔSW_{68}. The numbers on certainties are for the combined signal only. These calculations are for z = 7, with $N_{quasars}$ = 64, SNR − A = 10 and SNR − M = 10.

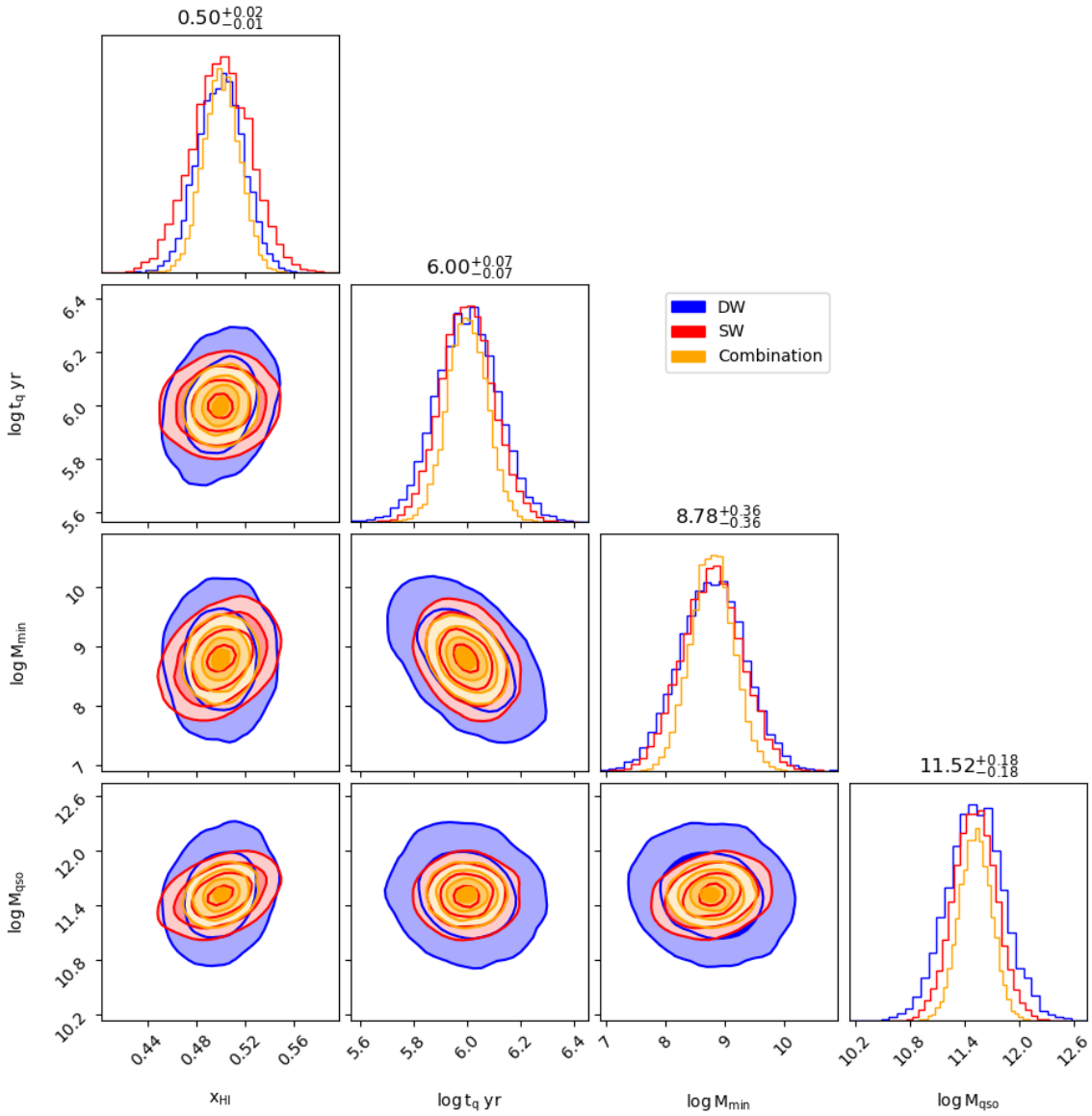


Figure 3.5. Similar to the Fig. 3.4, in this figure we show the 1D marginalized distribution along the diagonal representing the uncertainties on each parameter and 2D marginalized contour plots of each pair of parameters. Here we plot the contours for $N_{quasars} = 128$, SNR − A = 10 and SNR − M = 10 at z = 7.

and a standard deviation of 1/(additive-signal-to-noise ratio (SNR − A)). Subsequently, we multiply each sightline by normally distributed continuum noise, with a mean of 1 and a standard deviation of 1/(continuum-signal-to-noise ratio (SNR − M)). Unlike the spectral noise, for continuum noise, we draw a single value for the noise per sightline. This conservatively treats the continuum error as fully covariant. We add the continuum noise because the complete picture of the intrinsic quasar spectra is still not perfectly understood. However, the current models can predict the quasar spectra roughly within

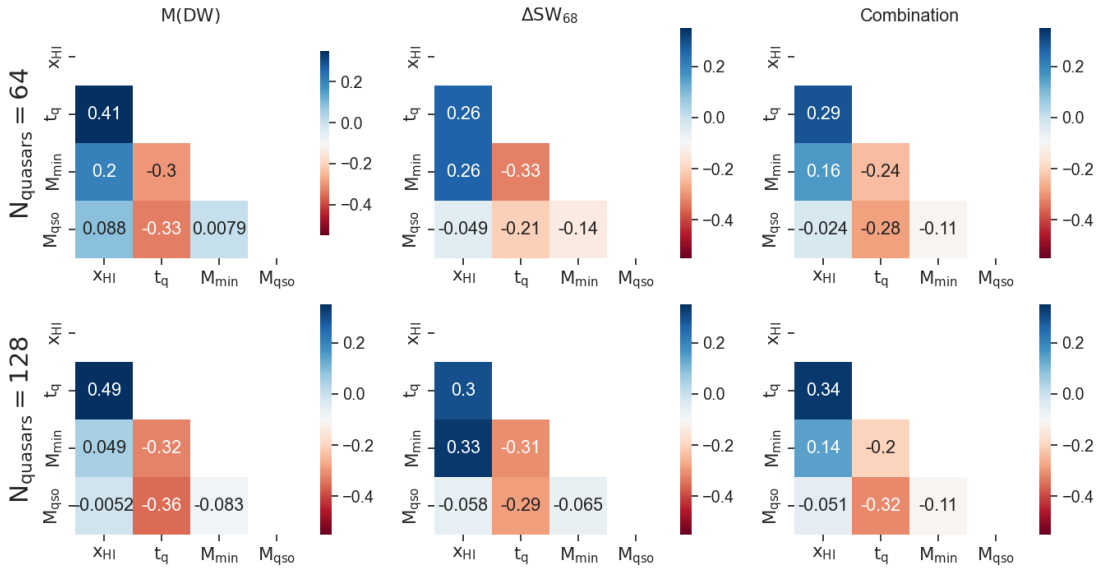


Figure 3.6. In this figure we plot the correlation matrices of all the parameters from M(DW), ΔSW_{68} and their combination, arranged column-wise respectively, for $N_{quasars} = 64$ (first row) and $N_{quasars}$ 128 (second row).

the 10% error limit. To account for this uncertainty, we include some continuum noise in our damping wing spectra. Fig. 3.1 visualizes a mock damping wing signal after incorporating both the proximity zone and the added noises, revealing a suppression of the signal on the bluer side of the spectra due to the presence of the proximity zone. This forward-modeled dataset constitutes our new mock data, which we will use to sample M(DW) and ΔSW_{68} for FIM calculations.

Once we have our mock dataset, we prepare N_{sample} (= 6000) samples, each containing $N_{quasars}$ random sightlines selected from the more extensive set of mock data. We vary $N_{quasars}$ in powers of 2 to examine how constraints change as a function of $N_{quasars}$. For each sample, we calculate the median value of the damping wing profile (M(DW)) and the scatter width (ΔSW_{68}), thus generating a set of realizations for both M(DW) and ΔSW_{68}. Subsequently, we compute the expectation of M(DW) and ΔSW_{68} values defined in Eq. (3.5) over the whole sample.

The mean values of these realizations (N_{sample}) for each pixel of M(DW) and ΔSW_{68} provide the expectation values of our vector $\hat{\mathcal{O}}$. In conjunction with varying $N_{quasars}$, we explore different values of SNR − A and SNR − M for all FIM calculations. We use the following fiducial values: $N_{quasars} = 64$, SNR − A = 10, and SNR − M = 10.

One of the primary advantages of employing FIM is that it requires the calculation

of only the first-order derivative with respect to the model parameters. This process can be efficiently performed by running only three simulation boxes per parameter, that is: one for the fiducial box and two for the parameter values slightly perturbed around the fiducial value, i.e., $\lambda_i(1 \pm \delta)$, where δ represents a small change around the fiducial values. Consequently, for a set of P parameters, we need to run only $2P + 1$ simulations to determine the FIM (see Eq. (3.11)) and ultimately derive the uncertainties in the parameters. We performed the convergence tests on the derivative and subsequently the FIM, we find for x_{HI}, $\delta = 2\%$; $\log t_q/\mathrm{yr}$, $\delta = 0.7\%$; and $\log M_{min}/M_\odot$, $\delta = 2.5\%$, the derivatives and FIM converge. For $\log M_{qso}/M_\odot$, the δ depends on our simulation mass bins, which depend on our pixel scale. For our case, the $\delta(\log M_{qso}/M_\odot) = 1.4\%$

3.5 Results

In this section, we present the results of the FIM calculations from our study. We begin by demonstrating a typical damping wing signal from our simulation in the Fig. 3.1, where the black dashed curve shows the damping wing profile generated by our simulation box. Once we add the effects of the proximity zone, we observe the suppression on the bluer side of the spectrum, which is shown by the red curve. The blue curve shows the effect of the proximity zone, spectrum noise, and the continuum noise. Next, we demonstrate the mock data set after sampling used in the FIM calculations. In Fig. 3.2, we show the distribution of the median damping wing signal (M(DW)) represented by thin black lines alongside the expectation value depicted as a red curve. The distribution of M(DW) at the Lyman-alpha (Lyα) pixel is also included. The pixel distribution of M(DW) is approximately Gaussian, consistent with our assumption for the likelihood function discussed in §3.2. Similarly, in Fig. 3.3, we present the distribution and expectation value of the scatter width (ΔSW_{68}), along with its pixel distribution. We note that the pixel variation of ΔSW_{68} is also roughly Gaussian, further validating our likelihood assumption.

In Fig. 3.4, we present the corner plots for M(DW) only, ΔSW_{68} only, and their combination, highlighting the constraints on each parameter derived from the FIM calculations. The diagonal elements represent the one-dimensional marginalized distributions of uncertainties for the individual parameters, while the off-diagonal elements show the two-dimensional marginalized contours, illustrating the correlation between each pair of parameters. The constraint values in this plot are derived from the combination

of M(DW) and ΔSW_{68}, which we refer to as the combined signal. Notably, with just 64 quasars, we achieve strong constraints on x_{HI} and t_q from the combined signal. The constraints on M_{min} (approximately 0.53 dex) and M_{qso} (approximately 0.32 dex) are also reasonable.

In Fig. 3.6, we display the correlation matrices for all parameters about each observable. The columns represent correlation matrices calculated using M(DW), ΔSW_{68}, and their combination, with the first row corresponding to $N_{quasars} = 64$, while the second row corresponds to $N_{quasars} = 128$. The inverse of the FIM provides the covariance matrix of the parameter estimates $\hat{\lambda}$ (where $\hat{\lambda}$ serves as an estimator of a parameter $\lambda \in \lambda_\alpha$ from the dataset). The off-diagonal elements of each plot represent the correlation between the estimates of two parameters, λ_i and λ_j.

Fig. 3.6 shows that x_{HI} positively correlates with t_q and M_{min}. Referring to Fig. 2.11, increasing the estimated value of x_{HI} for a given dataset causes more damping, shifting the M(DW) to the longer wavelength. But to observe the given dataset, this shift must be balanced by a counter shift from t_q or M_{min}. Thus, we must also increase the estimates of t_q and M_{min}. This same reasoning applies to ΔSW_{68}; increasing x_{HI} reduces the overall ΔSW_{68} while raising t_q, especially M_{min}, increases the ΔSW_{68}. Conversely, we can use similar arguments to explain the negative correlation for M(DW) between t_q and M_{min}. As increasing t_q increases the ionized bubble around the quasar, which can be balanced by decreasing the estimate for M_{min}. We also note that the correlation between x_{HI} and M_{min} for M(DW) is relatively weaker than for ΔSW_{68} since the M(DW) of M_{min} does not exhibit significant changes. Whereas the correlation between x_{HI} and t_q for M(DW) is relatively stronger than for ΔSW_{68}, since the ΔSW_{68} of t_q does not vary much over the extensive range of t_q (refer to Fig. 2.12). Similarly, in the second row of Fig. 3.6, we present the correlation matrices for $N_{quasars} = 128$. The correlation behavior remains consistent with $N_{quasars}$.

In Fig. 3.5, we see that with just 128 quasars, the constraints on M_{min} are already ≈ 0.36 dex. The constraints on M_{min} are close to those obtained in Park et al., 2019; Charlotte A Mason et al., 2018. This tells us that with a sufficient number of quasar spectra, with a significantly smaller number of simulations, we can achieve as reasonable constraints as obtained from the 21cm signal in Park et al., 2019; Charlotte A Mason et al., 2018. These calculations so far are done at redshift 7. The forecasts for Euclid (Schindler et al., 2023) predict a reasonable number of quasars at the redshift ≥ 7.

In Fig. A.8 and Fig. A.9 (the plots are in the Appendix A.2), we present the contour

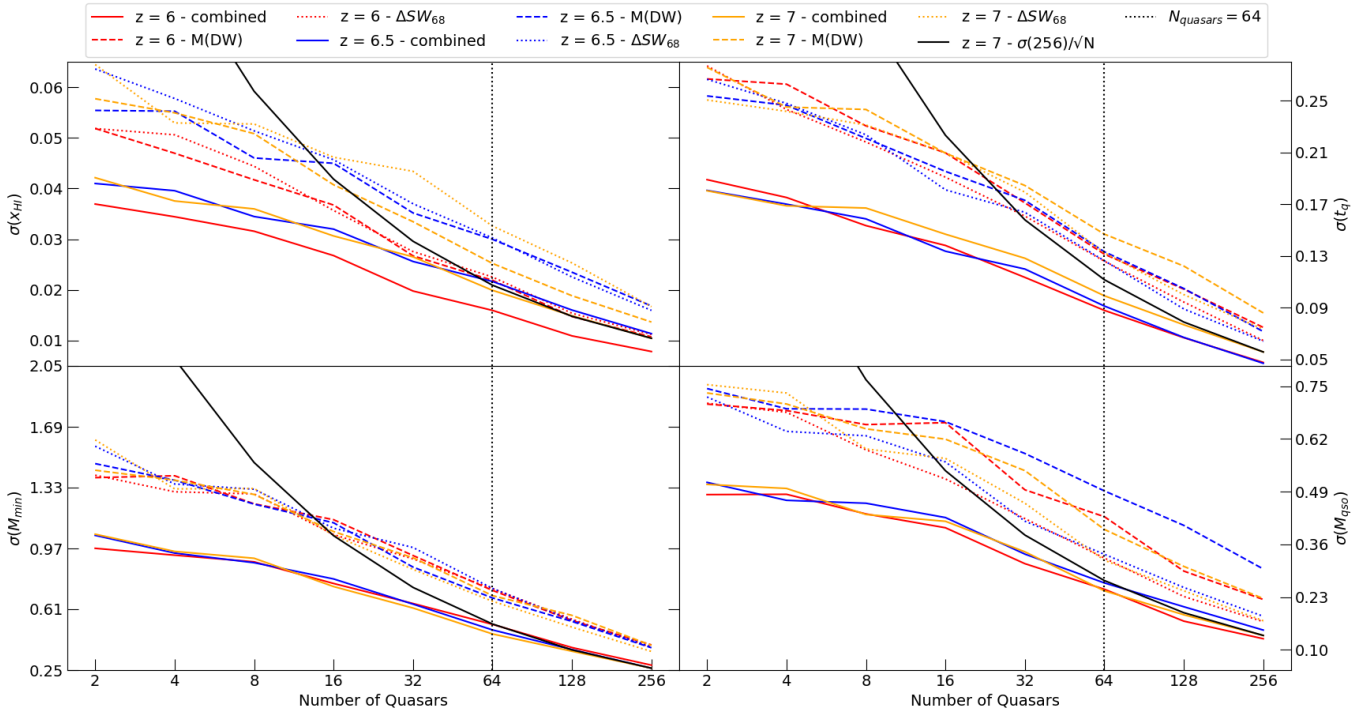


Figure 3.7. In this Figure, we show the redshift dependency of the constraints on each individual parameter as a function of $N_{quasars}$. In each plot, the different line styles show the constraints obtained from: the solid lines are for the combined (M(DW) +ΔSW_{68}) signal, the dashed lines are for the M(DW) signal, and the dotted lines are for the ΔSW_{68} signal. Whereas the color-scheme represents the redshifts, red is for $z = 6$, blue is for $z = 6.5$, and orange is for $z = 7$. The upper-left plot shows the variation of constraints of x_{HI} as a function of $N_{quasars}$, the upper-right plot shows the variation of constraints of t_q as a function of $N_{quasars}$, the lower-left plot shows the variation of constraints of M_{min} as a function of $N_{quasars}$, and the lower-right plot shows the variation of constraints of M_{qso} as a function of $N_{quasars}$.

plots for redshifts 6 and 6.5, respectively. The fiducial values of x_{HI} for these respective cases are 0.15 and 0.35. According to the findings of the Chapter 2, reducing the values of x_{HI} results in a significant increase in the width of the ΔSW_{68}, amplifying the uncertainty in the ΔSW_{68} measurement. Intuitively, this should degrade the constraints on M_{min}. However, as illustrated in Fig. 3.1, the proximity zone also suppresses the spectrum, further diminishing the constraints derived from M(DW). As we progress to lower redshifts, the effects of the proximity zone also decrease, thereby increasing the constraints obtained from M(DW) and ΔSW_{68}. Consequently, we observe that the constraining power on M_{min} remains relatively consistent across all redshifts examined. We specifically analyze M_{min}, as it presents the least stringent constraints among all the parameters considered.

In the Fig. 3.7, we explore the redshift dependency of the constraints on each individ-

ual study parameter as a function of $N_{quasars}$. On each panel in the Fig. 3.7, we also plot a solid black curve which shows the scaling relation of the uncertainties with $N_{quasars}$ at $z = 7$. One should expect the uncertainties to scale as $\sigma(N_{quasars} = 256) \times \sqrt{256}/\sqrt{N_{quasars}}$ for large $N_{quasars}$, assuming the likelihood has a Gaussian distribution. We see that for $N_{quasars} \geq 32$, the black curve is very close to the orange curve (combined signal at $z = 7$). Whereas, it deviates very rapidly for low $N_{quasars}$, signaling that our Gaussian likelihood approximation fails for very low $N_{quasars}$. Our model overestimates the constraints for low $N_{quasars}$; this is because the effective ΔSW_{68} and the range of M(DW) profiles get very narrow for low $N_{quasars}$, hence making the constraints stricter. For high enough $N_{quasars}$ we match the $1/\sqrt{N_{quasars}}$ scaling, which confirms that our Gaussian likelihood approximation is valid.

The constraints on the parameters suffer from a trade-off between the effects of the proximity zone and the ΔSW_{68} as we vary the redshift. Fig. 3.7 captures this effect with all the other parameters. We notice that at redshifts 6.5 and 7, the constraints are more or less the same for all the parameters except t_q. Whereas, the constraints are always better at redshift 6. Let's look at individual parameters; for x_{HI} there is suppression on the M(DW) from the proximity zone, and as we lower the redshift, the ΔSW_{68} spreads out, making the constraints worse. If we look at the redshift 7 (orange curve), the constraints from ΔSW_{68} are relatively better than M(DW). As we go lower in redshift, the suppression from proximity zone on M(DW) decreases much rapidly, causing the constraints from M(DW) and ΔSW_{68} (dashed and dotted lines respectively) to be much closer for redshift 6 and 6.5 (red and blue curves respectively). The suppression from the proximity zone for redshift 6 is so low that the overall constraints are much better than the redshifts 6.5 and 7. For M_{min} and M_{qso} where the constraints from ΔSW_{68} play a major role, we see the constraints overall do not change much with redshift. For t_q, we see another factor playing in, which is the effective bubble size around the halo. As the redshift decreases, the effective ionized bubble sizes increase due to the decrease in x_{HI}. This means that the overall ionized region around a halo once the quasar is switched on increases. Since the quasar erases any neutral topology around it, it effectively reduces the difference between individual sightlines for a larger length. Thus reducing the spread of ΔSW_{68} as we go to lower redshift (keeping in mind that there is intrinsic scatter on t_q since it is drawn from a lognormal distribution with some variance, hence, the ΔSW_{68} cannot get arbitrarily low).

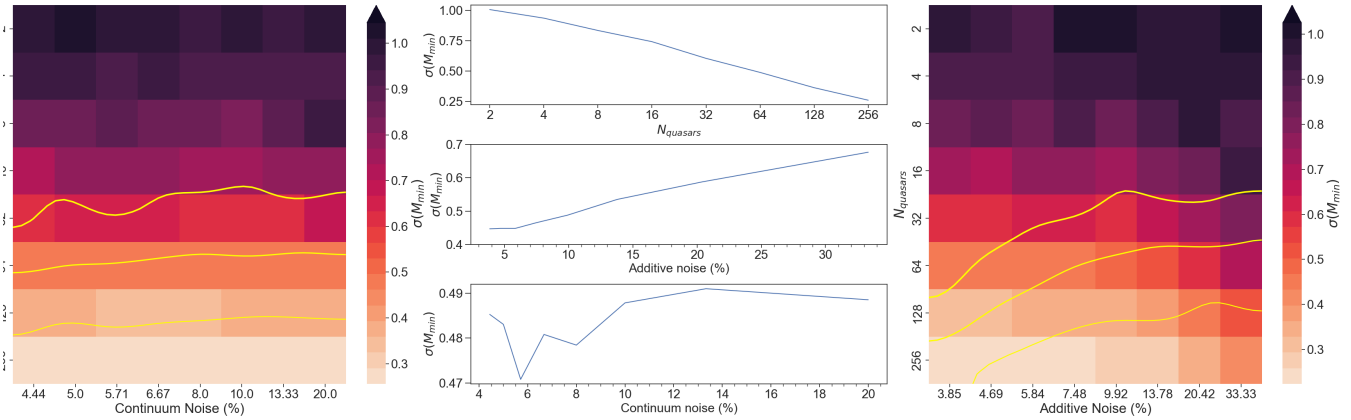


Figure 3.8. In this Figure, we show the dependency of $\sigma(M_{min})$ on the survey parameters ($\mathrm{N_{quasars}}$, SNR − A and SNR − M). on the leftmost panel we plotted the heatmap of $\mathrm{N_{quasars}}$ vs SNR − M with the colorbar representing $\sigma(M_{min})$. The yellow curves show the contours of $\sigma(M_{min}) = [0.35, 0.5, 0.65]$. Similarly, the rightmost panel is the heatmap for $\mathrm{N_{quasars}}$ vs SNR − A with the same colorbar and contour scheme. The middle panel has 3 line plots of $\sigma(M_{min})$ vs each of the survey parameters. The topmost line-plot show the variation of $\sigma(M_{min})$ vs $\mathrm{N_{quasars}}$ with SNR − A = 10 and SNR − M = 10. The middle line-plot show the variation of $\sigma(M_{min})$ with SNR − A with $\mathrm{N_{quasars}} = 64$ and SNR − M = 10. While, the bottom-most line-plot show the variation of $\sigma(M_{min})$ vs SNR − M with $\mathrm{N_{quasars}} = 64$ and SNR − A = 10.

3.6 Dependence on Survey Parameters

So far, we have seen how the constraints and correlations between various parameters vary as we vary their fiducial values. We also looked at the relationship between the shape of the 2D marginalized contours and all the parameter combinations. In this section, we would like to see how much these constraints are affected if we change the survey parameters, namely SNR − A, SNR − M or $\mathrm{N_{quasars}}$ in our sampling. The value of $\mathrm{N_{quasars}}$ is the number of quasars we need to observe to sufficiently use both the M(DW) and the $\Delta\mathrm{SW_{68}}$ statistics to justify the results obtained in the Chapter 2. We see that for $\mathrm{x_{HI}}$ and $\mathrm{t_q}$, the constraints are already very small for just 64 quasars. Hence, it is rather more useful to look at the convergence for the uncertainties in $\mathrm{M_{min}}$ and $\mathrm{M_{qso}}$. We show the calculations for redshift 7, unless stated otherwise.

In the Fig. 3.8, we show the variation on the constraints on $\mathrm{M_{min}}$ ($\sigma(\mathrm{M_{min}})$) with respect to the survey parameters. The left-most and the right-most figures show the heatmap plot of $\mathrm{N_{quasars}}$ vs SNR − M and SNR − A with the colorbar representing the $\sigma(\mathrm{M_{min}})$. The yellow curves on each heatmap show the contours for $\sigma(\mathrm{M_{min}}) = [0.35, 0.5, 0.65]$. We see that the contours, hence $\sigma(\mathrm{M_{min}})$, do not change much with the changes in the continuum noise. The middle three plots in the same figure show the

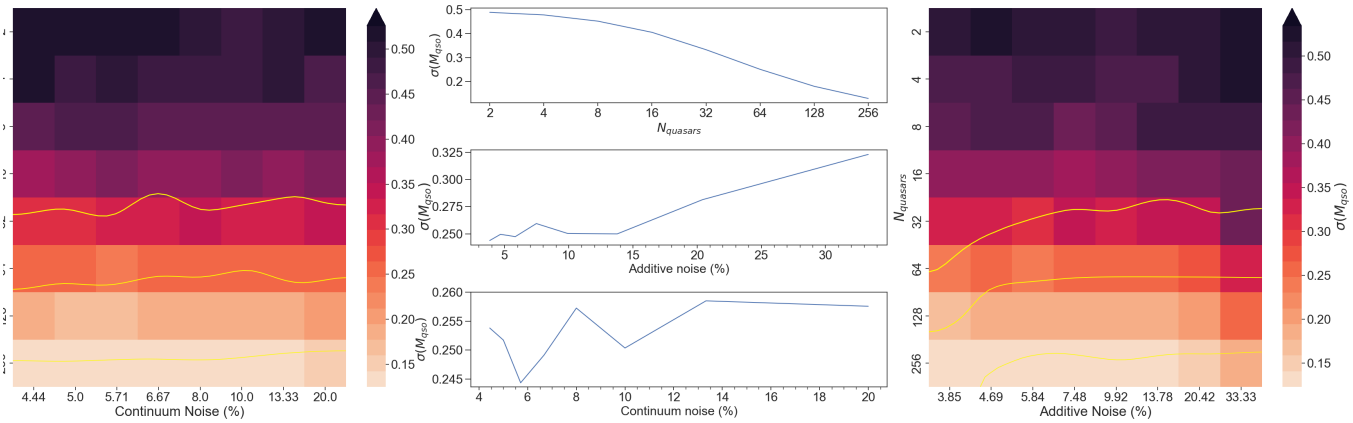


Figure 3.9. Similar to the Fig. 3.8, here we show the dependency of $\sigma(M_{qso})$ on the survey parameters, through the heatmaps and line plots. The leftmost and rightmost heatmaps show the distribution of $\sigma(M_{qso})$ over the range of $\mathrm{N_{quasars}}$ and SNR − M and SNR − A respectively. The yellow curves are the contours for $\sigma(M_{qso}) = [0.15, 0.25, 0.35]$. While the middle panel contains the line-plots of $\sigma(M_{qso})$ as a function of $\mathrm{N_{quasars}}$, SNR − A and SNR − M from top to bottom respectively.

line-plot variation of $\sigma(\mathrm{M_{min}})$ vs $\mathrm{N_{quasars}}$, SNR − A, and SNR − M from top to bottom, respectively. For each line plot of the survey parameter, we assume the fiducial values of the other two parameters. From these line plots, it is evident that $\mathrm{N_{quasars}}$ has the most significant impact on the constraints of $\mathrm{M_{min}}$. The additive/ spectral noise does not show a significant drop beyond 10%, which is also our fiducial value.

Similarly, in the Fig. 3.9, we show the similar heatmap with contours $\sigma(M_{qso}) = [0.15, 0.25, 0.35]$, and line-plot figures for $\mathrm{M_{qso}}$. Since the constraints on $\mathrm{M_{qso}}$ were a bit already stricter, the dynamic range of $\sigma(\mathrm{M_{qso}})$ is even smaller. The observations remain the same: that $\mathrm{N_{quasars}}$ is the most dominant survey parameter, and signal-to-noise ratio (SNR) = 10 is a reasonable amount of noise for both spectral and continuum noises. This suggests that we would need to observe as many quasars as possible to increase the robustness of our analysis. Furthermore, the measurements of $\mathrm{N_{quasars}}$ increase roughly linearly with time. In contrast, the observation time increases quadratically to improve the SNR for the spectral noise, thus making our analysis even more useful and time-conserving.

3.7 Discussion and Conclusion

In this work, we studied the constraining prowess of an ensemble of Lyα damping wings using the FIM analysis. We began by generating patchy reionization simulation boxes

with the specific values of the astrophysical parameters described in the Table 3.1. Using each box, for the desired mass bins, we generated a set of 10,000 randomly generated halo-sightline pairs. We then calculated our Lyα damping wing profiles along these sightlines to get our mock dataset. We then forward-modeled each sightline with the proximity zone, spectral noise, and continuum noise to generate a more realistic dataset. Finally we calculated the median (M(DW)) and the 68 percentile sightline-to-sightline scatter width around the median (ΔSW_{68}) for a batch of $N_{quasars}$ sightlines and used them along with their combination as the observables for the calculation of FIM as described in the Eq. (3.11). From the inverse of the FIM of the observables described above, we got the lower bound on the constraints on our chosen set of astrophysical parameters. We also looked at the redshift dependency of our study, by repeating the same analysis for $z = 6, 6.5$ and 7, where $z = 7$ is our fiducial case.

	$\dot{N}_{ph} = 10^{57} s^{-1}$		$\dot{N}_{ph} = 10^{56} s^{-1}$	
constraints (σ)	$N_{quasars} = 64$	$N_{quasars} = 128$	$N_{quasars} = 64$	$N_{quasars} = 128$
$\sigma(t_q)$	± 0.12	± 0.08	± 0.13	$+0.11, -0.10$
$\sigma(M_{min})$	± 0.53	$+0.37, -0.38$	± 0.49	± 0.39
$\sigma(M_{qso})$	$+0.32, -0.31$	$+0.24, -0.25$	$+0.33, -0.32$	± 0.24

Table 3.2. This table lists the change in constraints of t_q, M_{min} and M_{qso} for $N_{quasars}$ 64 and 128, as we change the luminosity of quasar, i.e., for $\dot{N}_{ph} = 10^{57} s^{-1}$ and $\dot{N}_{ph} = 10^{56} s^{-1}$

.

The results from our study suggest that for a big enough set of quasars ($N_{quasars} \geq 32$), the Gaussian likelihood assumption for the FIM is a valid approximation. We also observe that for ($N_{quasars} \approx 128$) our constraints on M_{min} are as good as the constraints obtained from 21 cm studies (Park et al., 2019; Charlotte A. Mason et al., 2023). But unlike 21 cm, we have treated each redshift independently, which is a conservative choice. Furthermore, if we extrapolate the constraints at $N_{quasars} = 256$, assuming the Gaussian likelihood, with $\sigma(N_{quasars} = 256) \times \sqrt{256}/\sqrt{N_{quasars}}$ curve for x_{HI} we get for $N_{quasars} = 1$ $\sigma(x_{HI}) = 0.20$ and for t_q we get $N_{quasars} = 1$ $\sigma(t_q) = 0.95$, which are close to the values obtained in (Kist et al., 2024) with full-continuum and full-simulation model.

We should note that our analysis has some caveats. We have assumed a simple model for the proximity zone and neglected the non-equilibrium effects, the effects of flickering (Frederick B. Davies et al., 2020; Satyavolu et al., 2023; Zhou et al., 2024), and the heating from Helium reionization. We leave these complications to future work.

Another caveat that we addressed was that we assumed all the quasars have the same luminosity, i.e., $\dot{N}_{ph} = 10^{57} s^{-1}$, but the quasars at high redshift could be fainter. The

luminosity of the quasars affects our analysis in two ways: first, by determining the size of the ionized bubble around the quasar, and then by calculating the proximity zone. As we have seen, a decrease in either of these factors changes the constraining power of our analysis. To test this we run another set of models with $\dot{N}_{ph} = 10^{56} s^{-1}$. From Eq. (3.29) and Eq. (3.28) we can see that as we decrease the $\dot{N}_{ph}$, the Γ_{HI} decreases but the Lyα optical depth, τ_{Lya} increases, which increases the suppression of the damping wing signal. Thus reducing the constraining power of M(DW). But the decrease of $\dot{N}_{ph}$ also decreases the ionized bubble around the quasar, due to which the variation from sightline-to-sightline and hence ΔSW_{68} increases for t_q, as noted from the Fig. 3.7. We know from the Fig. 3.6 that the t_q and M_{min} are negatively correlated, hence the reduction of the bubble size increases the constraining power of M_{min} instead. The amalgamation of these processes reduces the overall constraining power for t_q as we look at the fainter quasars, and changes the constraints on M_{min} and M_{qso} as shown in the Table 3.2. The changes to the constraints of x_{HI} are negligible.

With our study, we have shown the constraining abilities of M(DW) and ΔSW_{68} given a high enough number of quasars ($N_{quasars} \geq 32$), and the constraints obtained from our calculations are on par with the studies using 21 cm signal (Park et al., 2019; Charlotte A. Mason et al., 2023). The number of quasars predicted is well within the forecast range shown in (Schindler et al., 2023). Furthermore, there will imminently be dozens (of the order of 30) of JWST quasars spectra for $z \geq 7$, with the signal-to-noise ratios comparable to or even better than our fiducial values (GO 9180; PI Hennawi).

Take away points

- The FIM constraints follow the Gaussian likelihood approximation for $N_{quasars} \geq 32$
- For $N_{quasars} = 64$ quasars at redshift 7, $x_{HI} = 0.5^{+0.02}_{-0.02}$, $M_{min} = 8.78^{+0.53}_{-0.53}$, $\log t_q/yr = 6.0^{+0.12}_{-0.12}$, and $\log M_{qso}/M_\odot = 11.52^{+0.32}_{-0.31}$
- The constraints improve as go lower in redshift due to decreasing quasar proximity zone effects
- for quasars with $\dot{N}_{ph} = 10^{57}$, and 10^{56}, the constraints remain more or less same
- The constraints are most dependent on $N_{quasars}$, followed by spectral noise and least dependent on continuum noise

4

Understanding the short mean free path at $z = 6$

The work presented in this chapter is being prepared to be submitted for publication. My contribution to this work as a primary author has been to develop the necessary reionization simulation boxes, calculate the bubble size distribution and MFP from these boxes as a function of the set of astrophysical parameters, and analyze the results from the MFP calculations. As a co-author, Fred Davies provided guidance and supervision during this work. Yongda Zhu provided us with the Lyman series attenuation model and normalized quasar spectra to compare our models with the data.

Abstract

The Lyman continuum (Lyman Continuum (LyC)) mean free path (MFP) for low redshifts has been shown to follow a power-law relation. But at redshift, $z = 6$, the MFP is extremely short. The source of the short MFP is debatable. Hence, in this work, we explore the astrophysical parameter dependency of LyC MFP. We aim to fit the quasar spectra with the transmission profiles generated by our extensive suite of astrophysical parameters, to look for the set of parameter values that best describe the data. We find that M_{min} and M_{qso} have negligible effect on the final transmitted flux. While x_{HI} and t_q affect the signal the most. Our results implied that to explain the short MFP from the late reionization models, we would need $x_{HI} \geq 0.35$ and a very short t_q $(= 10^4 \text{yrs})$ or $x_{HI} \approx 0.5$ for $t_q = 10^6 \text{yrs}$. Thus, indicating that the late reionization models are insufficient to explain the short MFP of ionizing photons, we need some additional Lyman limit systems within the quasar proximity zone to explain the short MFP of the ionizing photons at $z = 6$.

Contents

4.1 Introduction

So far, we have studied how to probe the reionization topology using Lyα damping wings. In the Chapter 2, we saw that these damping wings are not only sensitive to the global topology but also to the local topology. We also saw how changing M_{min} changes the distribution of the bubble sizes across the whole box, and by changing the t_q, which changes the ionized bubble, locally shifts the whole damping wing profile. In the Appendix A.1, we studied how the distribution of the ionized bubbles in a simulation box changes as we change the box size, while keeping the resolution and initial conditions the same. In the Fig. A.7, we plotted the distribution of the ionized bubble around the 100 most massive halos as a function of the box size. However, the study of the ionized bubbles has far more implications than just testing the convergence of the boxes. Studying the ionized bubble distribution is crucial in understanding the evolution of neutral gas during reionization. In a very simplistic case, the boundary of these bubbles acts as the length scale at which the radiation from the luminous objects within these bubbles starts to suffer absorption or scattering from the IGM. Thus, on first order, the distance traveled by photons before suffering from any attenuation is at least as large as the size of the ionized bubble. This distance is nothing but the mean free path (MFP) of these photons. Of course, it is not the complete picture, but it is a good start to derive the mean free path of photons, whether they are absorbed or scattered.

Using a similar methodology as in Appendix A.1, we select a set of 10,000 random halos of a given mass from our simulation boxes. We then draw a random sightline from

these halos, assuming they host a quasar. We then look at the neutral gas density along this sightline. We define the boundary of the ionized bubble as the length from the center of the halo to the first pixel at which the HI fraction is greater than 0 ($x_{HI}(r) > 0$). Averaging over all these sightlines from all the halos of a given mass gives us the average ionized bubble radius for that mass. Repeating the same exercise for different x_{HI}, t_q, M_{min}, and M_{qso} provides us with the variation of the bubble size distribution as a function of our astrophysical parameters. Now, assuming the flux to be 1 in the ionized region and 0 once the sightline hits the neutral pixel, we then stack all these step-wise fluxes to calculate the flux as a function of distance. We define the MFP for such flux as the distance at which the flux drops to 1/e. This is a very generic way of calculating MFP from simulations. We do need to take into account the fact that the absorption cross-section of different radiation could be different, and hence, the MFP or the change in flux is a function of wavelength/ frequency of the incoming photon. We will discuss the methods to include the wavelength/ frequency dependencies later in this chapter.

In observations, especially at high redshift ($z > 5$), where the signal-to-noise ratio of the quasar spectra is not good enough, we usually stack a bunch of quasars at specific redshift bins to average out the noise and get average absorption features. This averaged flux is then compared with the intrinsic quasar flux to get the MFP. The intrinsic flux gets attenuated due to the absorption while traveling through the IGM. This attenuation is given by $e^{-\tau(r)}$, where $\tau(r)$ is the effective optical depth due to the absorption of the IGM, which is a function of the distance from the quasar. For the ionizing photons with wavelengths less than 912 Å, we can compare the observed flux to the intrinsic flux blueward of 912 Å. The ratio of the observed flux to the intrinsic flux gives us the relation for the optical depth ($e^{-\tau(r)}$). We define MFP (λ_{MFP} as the distance at which the observed flux drops to 1/e from the intrinsic value, i.e., the distance at which the optical depth reaches unity.

The MFP of the Lyman system can be broadly categorized into two categories: the LyC region and the Lyman Series (LyS) region. The LyC photons are the photons with the energy greater than the ionization potential of HI , $E_{ion} = 13.6$ eV. This corresponds to the wavelength shorter than the Lyman limit, $\lambda_{LL} = 911.76$ Å (for the sake of simplicity, we will use $\lambda_{LL} = 912$Å). In this process, the photon is completely absorbed, and the energy is used to eject the electron out of the HI atom, thus ionizing it in the process. In the case of LyS, the photons interact with the HI via resonant scattering (or resonant absorption, which is followed by re-emission). In this case, the

incoming photon excites the electron into some excited state, re-emitting the photon in a random direction and de-exciting back to the ground state. The LyS refers to the transition of the electron from the ground state to any excited state or vice versa. We showed some basic properties of Lyman absorption in §1.5.4.

In this chapter, we will focus on the MFP of LyC photons, i.e., the ionizing photons, as it provides a clear indication of the progress of reionization. In observations, the MFP of ionizing photons is measured by stacking the spectra of quasars at similar redshifts and observing the average drop of their flux past the Lyman limit (Stengler-Larrea et al., 1995; Prochaska et al., 2010; Songaila and Cowie, 2010). At $z \sim 3-4$, this MFP comes from optically-thick HI gas residing in the halos of galaxies (Lyman limit systems), i.e., small dense clouds with HI column density high enough to be optically thick to the LyC photons at the Lyman limit ($N_{HI} \geq 1.6 \times 10^{17} cm^{-2}$), within the fully-reionized IGM at those redshifts. Initially, the MFP for LyC photons was believed to increase rapidly as the ionized bubbles overlap. However, this did not fit the observational picture; either the Lyman limit systems were insufficient, or the bubbles were too big.

At $z = 6$, this MFP is very short (R. H. Becker et al., 2001) relative to the extrapolation from lower z, and this has been interpreted as a sign of incomplete reionization, i.e., there is some additional effective opacity from the presence of IGM-scale neutral islands (edges of ionized bubbles). Another competing interpretation is that the opacity primarily comes from increased HI within the large-scale ionized regions. In this work, we will assess this reionization connection directly, by estimating what combinations of source parameters and IGM neutral fractions would be required to reproduce the signal in the former case. Before we proceed with the modeling and results, we must first set the stage with a theoretical framework.

4.2 The Lyman Absorption Series

As discussed in the §1.5.4, the Lyman series refers to the transition of an electron from the ground state (principal quantum number $n = 1$) to any excited state ($n \geq 2$ or vice-versa. The energy of this transition is given by the Rydberg formula,

$$E_{n_2 \to n_1} = R_H \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right) \tag{4.1}$$

where $R_H \approx 13.6\ eV$ is the Rydberg constant for hydrogen and $n_1 = 1$ is the ground state for the Lyman series. The excited states $n_2 = 2, 3, 4, ...$ represent Lyα ($\lambda_\alpha \approx$

1215.67Å), Lyβ ($\lambda_\beta \approx 1025.72$Å), Ly$\gamma$ ($\lambda_\gamma \approx 972.53$Å), etc., series at the rest frame. For $n_2 = \infty$ or photons with energy greater than the ionization potential of hydrogen, $E_{ph} \geq 13.6\ eV$, we enter into the LyC regime. This corresponds to the wavelengths shorter than the Lyman limit ($\lambda_{\rm LyC} = 912$ Å).

4.2.1 Lyman Series Cross-Section

Any photon with energy $E_{ph} < 13.6\ eV$ can be scattered away from the HI by the absorption and subsequent emission from the electron within HI. The probability of any photon being absorbed/ scattered is described by the LyS line absorption cross-section, which is given by,

$$\sigma_{\rm j}(\nu) = \frac{\pi e^2}{m_e c} f_{\rm j} \phi(\nu - \nu_{\rm j}) \tag{4.2}$$

where $\phi(\nu - \nu_{\rm j})$ is the Voigt line profile, with $\nu_{\rm j}$ as the frequency of the jth Lyman transition, and $f_{\rm j}$ is the oscillator strength associated with the jth Lyman transition and is given by,

$$f_{\rm j} = \frac{256\, j^5 (j-1)^{2n-4}}{3(j+1)^{2j+4}}. \tag{4.3}$$

4.2.2 Lyman Continuum Cross-Section

On the other hand, any photon with the energy $E_{ph} \geq 13.6\ eV$ will suffer from the photoionization absorption with the cross-section given as

$$\sigma_{HI}(\nu) = \sigma_{\rm LyC} \left(\frac{\nu}{\nu_{\rm LyC}} \right)^{-2.75} \tag{4.4}$$

where, $\sigma_{\rm LyC} \approx 6.30 \times 10^{-18} cm^2$ (Kirkman and Tytler, 2008) is the photoionization cross-section at the Lyman limit.

4.2.3 Opacity and Optical Depth

The opacity of a medium is the measure of the medium's ability to absorb or scatter the incoming photons. For photons with frequency ν, passing through a medium with $\rm n_{HI}$ (in our case, the absorbers are HI) density of absorbers, with $\sigma(\nu)$ (either LyS or LyC) being the absorption cross-section, the opacity can be written as,

$$\kappa_\nu = \sigma(\nu) n_{HI} \tag{4.5}$$

The opacity is the inverse of the MFP of the photons with the frequency ν, passing through the same medium. It also relates to the optical depth. The optical depth, $\tau(\nu)$, as a function of the frequency of the incoming photons, helps quantify the amount of absorption these photons suffer. For the photons with frequency ν, traveling through the same medium and the absorption cross-section given by σ_ν, the optical depth along this infinitesimal distance dr is defined as,

$$d\tau(\nu) = \mathrm{n_{HI}(r)}\sigma(\nu)(\mathrm{r})\mathrm{dr} = \kappa(r,\nu)\mathrm{dr} \tag{4.6}$$

If we assume the IGM to be uniform, then we can integrate the above equation to get the optical depth for LyS and LyC transitions by using the definition of the cross-sections defined in the Eq. (4.2) and Eq. (4.4). Let's look at them one by one.

4.2.3.1 LyC optical Depth

We will first look at the LyC optical depth. After substituting the form of the LyC cross-section from the Eq. (4.4) into the §4.2.3, and expressing it in terms of redshift, we get the following equation:

$$\kappa_{\mathrm{LyC}}(z) = \kappa_{912}(z)\left(\frac{1+z}{1+z_{912}}\right)^{-2.75} \tag{4.7}$$

where, $\kappa_{912}(z) = \sigma_{\mathrm{LyC}} \times n_{HI}$ is the opacity at 912 Å. Now converting dr into dz from FLRW cosmology,

$$\frac{dr}{dz} \equiv \frac{c}{H(z)(1+z)} = \frac{c/H_0}{(1+z)\sqrt{\Omega_{\mathrm{m}}(1+z)^3+\Omega_\Lambda}} \tag{4.8}$$

substituting the Eq. (4.7) and Eq. (4.8) in Eq. (4.6) and assuming the matter dominated universe approximation $dr/dz \approx c/(H_0\Omega_m^{1/2})(1+z)^{-5/2}$, we can integrate Eq. (4.6) for the redshift range $\mathrm{z_{qso}}$ (redshift of the quasar source) to the redshift at which the wavelength of the emitted photon redshifts to 912 Å, to get the relation in (Prochaska et al., 2009),

$$\tau_{\mathrm{eff,LyC}}(z_{912}, z_q) = \frac{c}{H_0\Omega_{\mathrm{m}}^{1/2}}(1+z_{912})^{2.75} \times \int_{z_{912}}^{z_q} \kappa_{912}(z')\,(1+z')^{-5.25}\,dz' \tag{4.9}$$

4.2.3.2 LyS optical Depth

For the LyS optical depth, we just need to substitute the frequency, ν, to the observed frequency $\nu_{\rm obs} = \nu(1 + {\rm z})$ and use the Eq. (4.8) for dr/dz. Now substituting Eq. (4.2) in the Eq. (4.6) and using the identity

$$\int_{-\infty}^{\infty} \phi(x)dx = 1,$$

we can integrate the Eq. (4.6) to get the following result,

$$\tau_j^{\rm LyS} = \frac{\pi e^2}{m_e c} f_{\rm j} \frac{n_{\rm HI}}{\nu_{\rm j} H(z_{\rm j})} \tag{4.10}$$

where $z_{\rm j}$ is the redshift at which the $\nu(1 + z) = \nu_{\rm j}$.

4.3 Quasar Photoionization Rate

The quasars' high luminosity makes their ionizing flux dominate locally over the ultraviolet background radiation (UVB). Due to which the transmission near the quasar increases compared to the background IGM transition. This increased transmission can be used to measure the background photoionization rate of UVB, $\Gamma_{\rm bkg}$, by measuring the reduction in the number of strong absorption troughs (Bajtlik et al., 1988) or using the flux statistics.

We can characterize the transmission flux using the concept of optical depth. For any flux F, we can also define optical depth as τ as ${\rm F} = {\rm N_0 e^{-\tau}}$, where $\rm N_0$ is unattenuated flux. Now, ignoring the gas motion or the temperature gradient in the vicinity of the quasar, we can define the optical depth in the vicinity of the quasar proximity zone using a very simple model (Calverley et al., 2011)

$$\tau = \tau_{\rm bkg}[1 + \omega(r)]^{-1} \tag{4.11}$$

where $\tau_{\rm bkg}$ is the background optical depth in the absence of the quasar, and $\omega(r)$ can be defined as,

$$\omega(r) = \frac{\Gamma_{\rm qso}(r)}{\Gamma_{\rm bkg}} \ . \tag{4.12}$$

The Γ terms here are the HI photoionization rates, where $\Gamma_{\rm qso}(r)$ is the quasar photoionization rate and $\Gamma_{\rm bkg}$ is the background photoionization rate from UVB. We

assume $\Gamma_{\rm qso}(r)$ to be spherically symmetric and only depend on the distance from the quasar, r, which can be approximated as,

$$r \simeq \frac{c}{H(z)} \frac{\Delta z}{1+z} \ . \tag{4.13}$$

Now we define the characteristic length $\mathrm{R_{eq}}$, which is a representative of the quasar proximity zone. It is the distance from the quasar where the photoionization rate from the quasar is equal to that of the UVB, i.e., $\Gamma_{\rm qso}(\mathrm{R_{eq}}) = \Gamma_{\rm bkg}$. Using this definition, we can rewrite the Eq. (4.11) as

$$\tau = \tau_{\rm forest} \left[1 + \left(\frac{r}{R_{\rm eq}} \right)^{-2} \right]^{-1} \ . \tag{4.14}$$

Although we use $\mathrm{R_{eq}}$ as the proxy of quasar proximity zone, it must be noted that observationally, the proximity zone is defined as, "the maximum extent of the enhanced transmission flux, i.e., the distance at which the transmission flux drops to 0.1 in the spectrum, when the spectrum is smoothed with 20 Å filter" (Xiaohui Fan et al., 2006; Carilli et al., 2010). Since this is the observational definition and not the physical, we choose to stick to the definition of (Calverley et al., 2011), which is, it is the characteristic length at which $\Gamma_{\rm qso}(\mathrm{R_{eq}}) = \Gamma_{\rm bkg}$ or $\omega = 1$.

For our simplistic case, we can assume that the flux drops as $1/r^2$. We can further consider the frequency dependence of the intrinsic flux (or the spectral energy density) to follow a power law relation (in George D. Becker et al., 2021, the absolute magnitude of the quasars was measured for the mean luminosity at restframe 1450 Å, but the choice of this wavelength hence does not impact the understanding of the theoretical framework),

$$f_\nu(\nu) \propto \begin{cases} \nu^{-\alpha_\nu^{\rm UV}}, & 912 \text{ Å} < \lambda < 1450 \text{ Å} \\ \nu^{-\alpha_\nu^{\rm ion}}, & \lambda < 912 \text{ Å} \end{cases} \tag{4.15}$$

We can define the flux or the luminosity at the Lyman limit 912 Å by some normalization, which will be $f_{912} = f_{1450}(\nu_{912}/\nu_{1450})^{-\alpha_\nu^{\rm UV}}$. Where $\alpha_\nu^{\rm UV} = 0.6 \pm 0.1$ taken from Lusso et al., 2015 and $\alpha_\nu^{\rm ion} = 1.5 \pm 0.3$ taken from George D. Becker et al., 2021. Thus, we can now define the luminosity at the Lyman limit as,

$$L_{912} = 4\pi d_L^2 \frac{f_{912}}{(1+z_{\rm q})} \tag{4.16}$$

where d_L is the luminosity distance of the quasar. Using this, we define the Lyman limit flux density at any distance r from the quasar,

$$F_{912}^{\rm Q}(r) = \frac{L_{912}}{4\pi r^2} \ . \tag{4.17}$$

The general form of the photoionization rate of HI, Γ (in s^{-1}) due to a source of UV flux is given by (see Eq. (1.120)),

$$\Gamma = \int_{c/912}^{\infty} \frac{4\pi J(\nu)\sigma_{HI}(\nu)}{h\nu} d\nu \tag{4.18}$$

where $J(\nu)$ is the source intensity, $\sigma_{HI}(\nu)$ is the ionization cross-section of HI as a function of ionizing photon frequency ν, and h is Planck's constant. For an isotropic source, $F_{912}^{Q} = \pi J(\nu)$ and integrating the Eq. (4.18) keeping in mind $\Gamma_{\rm qso}(\mathrm{R_{eq}}) = \Gamma_{\rm bkg}$, we get

$$\Gamma_{\rm bkg} = \frac{\sigma_{\rm LyC} F_{912}^{\rm Q}(R_{\rm eq})}{(\alpha_{\nu}^{\rm ion} + 2.75)} \tag{4.19}$$

where $\sigma_{\rm LyC}$ is the HI ionization cross-section at 912 Å, The above equation can be rewritten in terms of luminosity,

$$\Gamma_{\rm bkg} = \frac{\sigma_{\rm LyC} L_{912}}{(\alpha_{\nu}^{\rm ion} + 2.75) 4\pi h R_{\rm eq}^2} \tag{4.20}$$

here Γ is in the units of s^{-1}, L_{912} is in erg s^{-1} Hz^{-1}, and $\mathrm{R_{eq}}$ is in cm. We can rearrange the above equation to get the definition of $\mathrm{R_{eq}}$

$$R_{\rm eq} = \left[\frac{L_{912}\sigma_{\rm LyC}}{4\pi h \Gamma_{\rm bg}(\alpha_{\nu}^{\rm ion} + 2.75)} \right]^{1/2} \tag{4.21}$$

4.4 Model setup

Similar to the model setup in Chapter 2 and Chapter 3, we begin by generating a suite of astrophysical parameters dependent suite of patchy reionization models using 21cmFAST over large cosmological volumes (512 cMpc^3) at $z = 6$. To study the parameter dependency of the MFP, we select a much larger range of $\mathrm{x_{HI}}$, $\mathrm{M_{min}}$, and $\mathrm{t_q}$. The fiducial values and the corresponding ranges of our astrophysical parameters are shown in the Table 4.1. Again, we define an ionized box with a unique combination of $\mathrm{x_{HI}}$, $\mathrm{M_{min}}$, and $\mathrm{t_q}$ as one model. Thus, we generated a suite of 375 models. For each model, we selected 12 mass bins for $\mathrm{M_{qso}}$ to effectively cover the entire mass range, with a larger focus on heavier halos. Using the same halo finder algorithm as discussed in §2.2, we

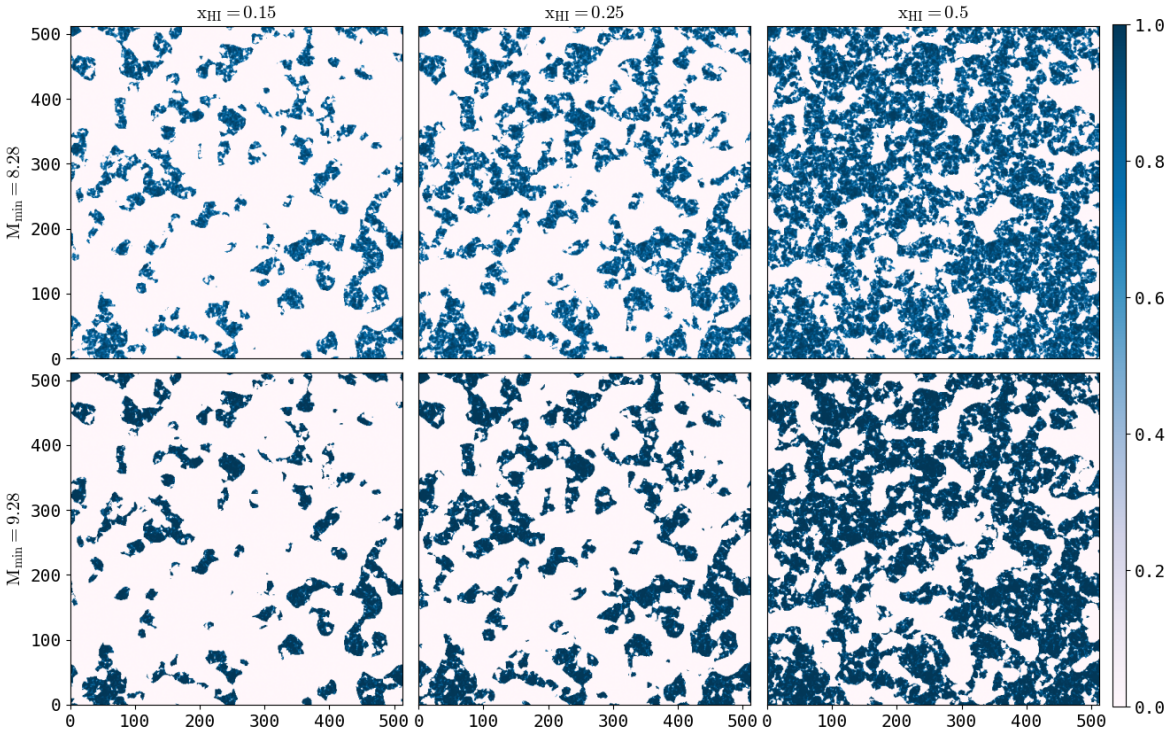


Figure 4.1. In this Figure, we show the changes in the reionization topology and its effects on the ionized bubble as a function of x_{HI} and M_{min}. The rows show different values of x_{HI} while the columns show different values of M_{min}. The changes in x_{HI} change the overall HI density within the box, whereas the changes in M_{min} redistribute these HI densities.

randomly select a sample of 10,000 halo-sightline pairs for each mass bin. Along each sightline, we calculate the HI fraction and trim the length of the sightline as soon as it hits a neutral pixel. This trimmed sightline shows the shortest distance to the neutral patch from the center of the halo, representing the size of the ionized bubble around the halo. Since our reionization model is patchy, the ionized bubbles around any halo are not spherical. In this case, we take the average of all the trimmed sightlines to get the average bubble size around the halo of a given mass. The size of these bubbles depends not only on the halo mass but also on the topology of reionization. As shown in the Chapter 2, x_{HI} and M_{min} play a crucial role in governing the size and the distribution of these ionized bubbles, while t_q modifies the bubble size around the halo. In the Fig. 4.1 we show the reionization topology as a function of x_{HI} and M_{min}. We see that as we change the M_{min}, the distribution of bubble size changes. For low M_{min} we see a lot more smaller and finer bubbles, while for large M_{min} we see bigger and coarser bubbles. This is already explored in §2.4.

Our models for the t_q are the same as used in the Chapter 2 and Chapter 3 where we carve out a bubble of radius R_{t_q}, given by the Eq. (3.26) along the direction of sightline

Parameter	**Fiducial model**	Range of values	**Parameter type**
x_{HI}	0.15	$[0.05, 0.10, ..., 0.70, 0.75]$	IGM parameter
$\log M_{min}/M_{\odot}$	8.78	$[8.28, 8.61, 8.78, 8.95, 9.28]$	Source parameter
t_q yrs	6	$[0, 10^4, 10^5, 10^6, 10^7]$	Quasar parameter

Table 4.1. Astrophysical Parameters Space: This table lists all the parameters used in our study, along with their fiducial values, the constraints obtained after FIM analysis for $N_{quasars}$ = 64 at redshift z = 7, and parameter types.

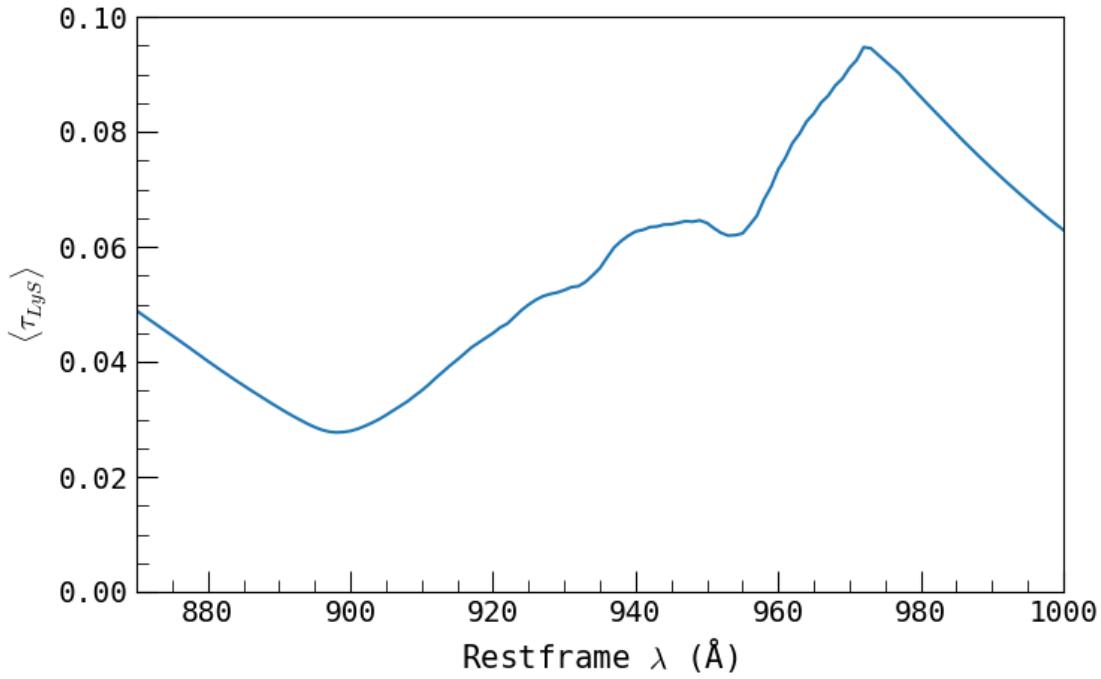


Figure 4.2. In this Figure, we plot the transmission flux attenuation factor for the Lyman series as a function of the restframe wavelength.

(assumed as our line of sight) in the post processing. Similar to the Chapter 3, we take into account the variability of t_q from quasar to quasar by drawing the values of t_q from a log-normal distribution with the mean value representing the required t_q and the standard deviation of 0.8. After generating the trimmed sightlines, we stack them together to get the distribution of the bubble radii. To convert the bubble size distribution to the transmission flux from the quasar, we assume the flux to be 1 within the bubble and 0 outside of it. We then stack all the fluxes and normalize them. Thus generating the general transmission flux as a function of distance. We then convert the distance scale to the rest-frame wavelength scale of LyC, which peaks at 912 Å.

To derive the observed flux, we need to take into account the IGM absorption of the flux. These absorptions will attenuate the overall signal. Since we want to look at the MFP of the ionizing photons, we need to attenuate the general flux with the effective Lyman series opacity due to the foreground IGM absorption, κ^{LyS}, and the

attenuation due to the effective Lyman continuum absorption, $\kappa^{\rm LyC}$. Thus, we can write the final observed flux as,

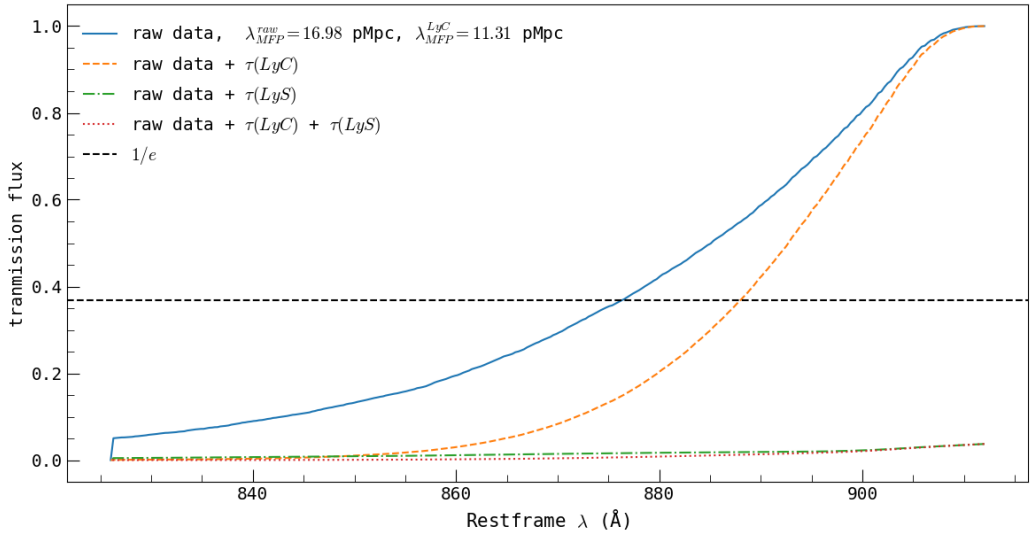


Figure 4.3. In this Figure, we plot the raw transmission flux (blue curve), transmission flux modified by the LyC attenuation only(orange dashed curve), transmission flux modified by the LyS attenuation only(green dotted and dashed curve), and the final observed transmission flux (red dotted curve). We also show the effective mean free path after the LyC absorption as well.

$$f_\lambda^{obs} = N_0 f^{gen} \exp\left(-\tau^{\rm LyC}\right) \exp\left(-\tau^{\rm LyS}\right) \tag{4.22}$$

Where N_0 is the normalization free parameter, f^{gen} is the general transmission flux we get after stacking the sightlines, $\tau^{\rm LyC}$ is the LyC optical depth and $\tau^{\rm LyS}$ is the foreground Lyman series absorption optical depth. This equation is analogous to equation 1 in (George D. Becker et al., 2021).

4.4.1 LyC Absorption

We have already derived the LyC optical in the Eq. (4.9), where the optical depth depends upon the opacity of the medium. Since our medium is now affected by the presence of a quasar, we need to include the contribution from the quasar proximity zone on the LyC opacity/ optical depth. We can calculate the effective LyC optical depth as a function of distance r from the quasar, as:

$$\tau^{\rm LyC} = \int_0^R \kappa^{\rm LyC} dr \tag{4.23}$$

where $\kappa^{\rm LyC}$ is the LyC opacity and can be written as a function of the quasar photoionization rate,

$$\kappa^{\rm LyC}(\rm r) = \kappa^{\rm LyC}_{\rm bg}\left[1 + \frac{\Gamma_{\rm qso}(\rm r)}{\Gamma_{\rm bg}}\right]^{-\xi}. \tag{4.24}$$

Where $\kappa^{\rm LyC}_{bg}$ is the background opacity which we equate to $1/42$ $\rm h^{-1}_{70}$ (Worseck et al., 2014; Qin et al., 2025). The power law index $\xi = 2/3$ is a fiducial number derived from the radiative transfer simulations through IGM (S. R. Furlanetto and Oh, 2005; McQuinn et al., 2011). $\Gamma_{\rm qso}(\rm r)$ is the quasar photoionization rate at some distance r away from the quasar, and $\Gamma_{\rm bg}$ is the background photoionization rate. The photoionization rate at any distance r away from the quasar is then calculated using the cumulative optical depth $\tau^{\rm LyC}$ up to r. The quasar photoionization rate is further characterized by $\rm R_{eq}$, which represents the characteristic distance at which the $\Gamma_{\rm qso}(\rm r)$ is equal to the $\Gamma_{\rm bg}$ in the absence of any absorption/ attenuation of the ionizing photons from the quasar. The $\Gamma_{\rm qso}(\rm r)$ can thus be written as,

$$\Gamma_{\rm qso}(\rm r) = \Gamma_{\rm bg}(\rm r)\left(\frac{r}{R_{eq}}\right)^{-2} e^{-\tau^{\rm LyC}(<\rm r)} \tag{4.25}$$

4.4.2 LyS Absorption

The LyC spectrum of light, i.e., for the wavelength ($\lambda \leq 912$Å), will also suffer from the redshift effects while traveling through IGM and may enter into the Lyman series regime. Thus, we need to correct for the Lyman series foreground opacity attenuation at the observed wavelength ($\lambda_{\rm obs} < 912(1 + z_{\rm qso})$Å). The attenuation factor can be derived in the same way we derived the LyC optical depth, but this time we have to sum over the whole Lyman resonant series.

$$\tau^{\rm LyS}(\lambda_{\rm obs}) = \sum_{j={\rm Ly}\alpha,{\rm Ly}\beta,\ldots,} \tau^{j}(z_j) \tag{4.26}$$

The $\tau^{\rm j}$ is the optical depth of transition j at the redshift $\rm z_j$ such that $\lambda_{\rm j}$ is the restframe wavelength of the transition j, and $\lambda_{\rm j} \times (1 + \rm z_j) = \lambda_{\rm obs}$. The quasar proximity zone modifications on the LyS opacity/ optical depth are similar to the equations described for LyC opacity. The only differences being, here when calculating $\Gamma_{\rm qso}$, we do not use

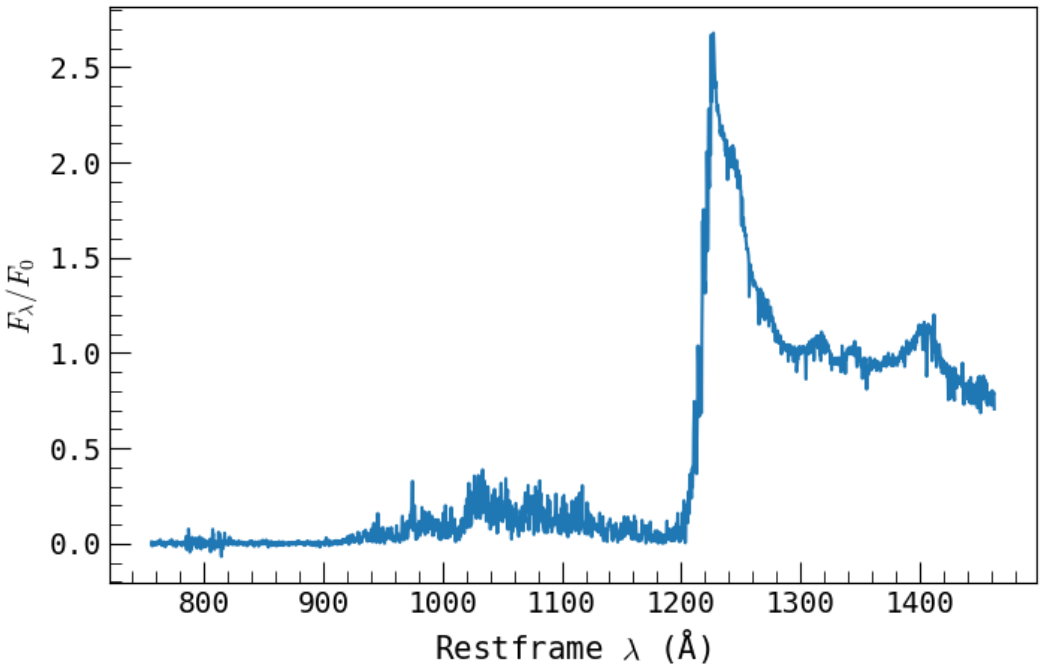


Figure 4.4. In this Figure, we show the composite quasar mean quasar spectra after stacking the quasars from the Table 4.2. The flux has been normalized by the median of the continuum flux over the range of 1270 – 1380Å.

the cumulative LyC optical depth ($\tau^{\rm LyC}(<$ r), since it is a line transition and not the continuum transition, we replace it with $\tau^{j}(r)$. In Becker21, Zhu23, they calculate it using a suite of Sherwood simulations (James S. Bolton et al., 2017) across a range of redshifts and Γ values, for the first 39 Lyman series. They then include the effects of the proximity zone by calibrating the values of Γ for each series such that the optical depth $\tau_{\rm j}(\Gamma)$, matches to that of $\Gamma_{\rm qso} + \Gamma_{\rm bg}$ as a function of distance from the quasars. We skip the details of these simulations as they are beyond the scope of this chapter; the curious reader can refer to James S. Bolton et al., 2017. We received the final Lyman series attenuation factor model from Yongda Zhu and a stack of 9 normalized quasar spectra, which they also used in their calculations (Zhu et al., 2023). In the Fig. 4.2 we plot the mean attenuation factor ($\langle \tau_{LyS} \rangle = \exp\left(-\tau^{\rm LyS}\right)$), as a function of restframe wavelength.

For our work, we adopt the $\rm R_{eq}$ values from George D. Becker et al., 2021; Zhu et al., 2023, where they estimated the $\rm R_{eq}$ from observational constraints on the metagalactic hydrogen ionization rate and the mean spectral energy distribution of high-redshift quasars. We received the stack of pre-normalized quasar spectra from Yongda Zhu and their respective values of $R_{\rm eq}$. The average value of $\langle \rm R_{eq} \rangle$ = 17.96 pMpc is calculated from the $\rm R_{eq}$ values for the set of quasar spectra shown in Table 4.2. We calculated LyC

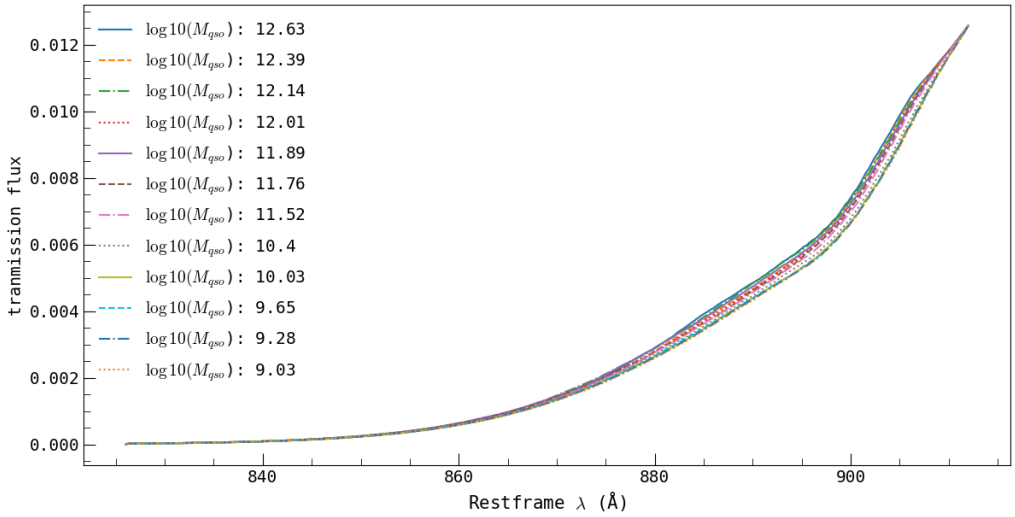


Figure 4.5. In this Figure, we plot the observed flux as a function of the restframe LyC wavelength for a range of mass bins. We plot it for the fiducial values of x_{HI}, M_{min}, and t_q. We see that the choice of the halo mass does not change the overall flux much.

Quasar Name	**R.A. (J2000)**	Dec (J2000)	z_{qso}	R_{eq} (pMpc)	Instrument
SDSSJ0836+0054	08:36:43.85	+00:54:53.3	5.805	21.41234719084271	ESI
SDSSJ0002+2550	00:02:39.40	+25:50:34.8	5.824	17.26091101003634	ESI
PSOJ242-12	16:09:45.53	−12:58:54.1	5.8468	14.757086164439244	X-Shooter
SDSSJ0840+5624	08:40:35.09	+56:24:19.8	5.8441	16.75118105781647	ESI
PSOJ183-12	12:13:11.81	−12:46:03.5	5.899	20.060635831726962	X-Shooter
SDSSJ0818+1722	08:18:27.40	+17:22:52.0	6.001	20.58434709674912	X-Shooter
SDSSJ1137+3549	11:37:17.72	+35:49:56.9	6.03	19.657898865487297	ESI
SDSSJ0842+1218	08:42:29.43	+12:18:50.6	6.0763	15.614824104210964	X-Shooter
SDSSJ1602+4228	16:02:53.98	+42:28:24.9	6.084	15.614824104210964	ESI

Table 4.2. Quasars spectra list: Column 1 shows the name of the quasar. Columns 2 and 3 show the coordinates of the quasar. Column 4 shows the quasar redshift. Column 5 shows the R_{eq} values of these quasars, and column 6 shows the instrument used for λ_{mfp} measurements.

optical depth using the mean R_{eq}. We assume the value of $\Gamma_{bg} = 1.5 \times 10^{-13} s^{-1}$ based on the (Gaikwad et al., 2023). The value of $\kappa_{bg}^{LyC} = 1/42$ (Worseck et al., 2014), and the power law index $\xi = 2/3$ (S. R. Furlanetto and Oh, 2005; McQuinn et al., 2011), as defined above. This leaves only N_0 as the free parameter. We fix the value of N_0 by normalizing our observed flux (as shown in the Eq. (4.22)) such that the transmission from our models at 912Å is equal to the best fit model from (Zhu et al., 2023). In the Fig. 4.3 we show the effective transmission flux for the raw profile, and profiles attenuated by LyC, LyS, and by both LyC and LyS. It shows us how the shape of the transmission flux is also affected by the different attenuation terms. In §1.5.4, we read

that the absorption cross section of any LyS is much higher than the LyC absorption, because of which, and also the fact that we added the contribution from multiple LyS lines for LyS optical depth, we see much higher damping effects from the LyS attenuation term as compared to the LyC attenuation term. In this plot, we use the fiducial values of our parameter set and $M_{qso} = 10^{11.52} M_{\odot}$.

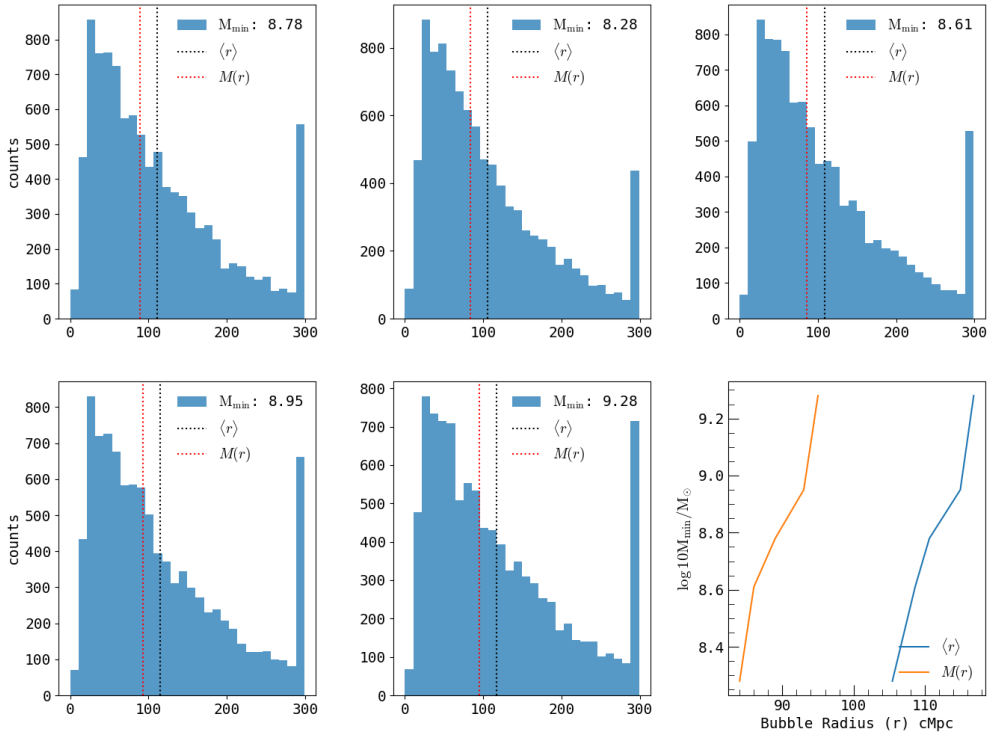


Figure 4.6. In this Figure, we show the distribution of the ionized bubbles as a function of M_{min}. The first five plots show the histogram of the ionized bubbles radii with their mean radius $\langle r \rangle$, and the median radius M(r). On the sixth plot, we plot the $\langle r \rangle$ (orange curve) and M(r) (blue curve) radii of ionized bubbles for all the values of M_{min}. As expected, we see that both the $\langle r \rangle$ and M(r) grow as M_{min} is increased. But this growth is not very significant.

In the Fig. 4.4, we plot the stacked flux across all quasars mentioned in the Table 4.2 normalized with the median of the continuum flux over the range of $1270 - 1380$Å, with the average redshift $\langle z \rangle = 5.93$.

4.5 Results

Once our models are ready, we can now begin to analyze the parameter dependency of the final observed flux after taking into account the attenuation from the LyC and

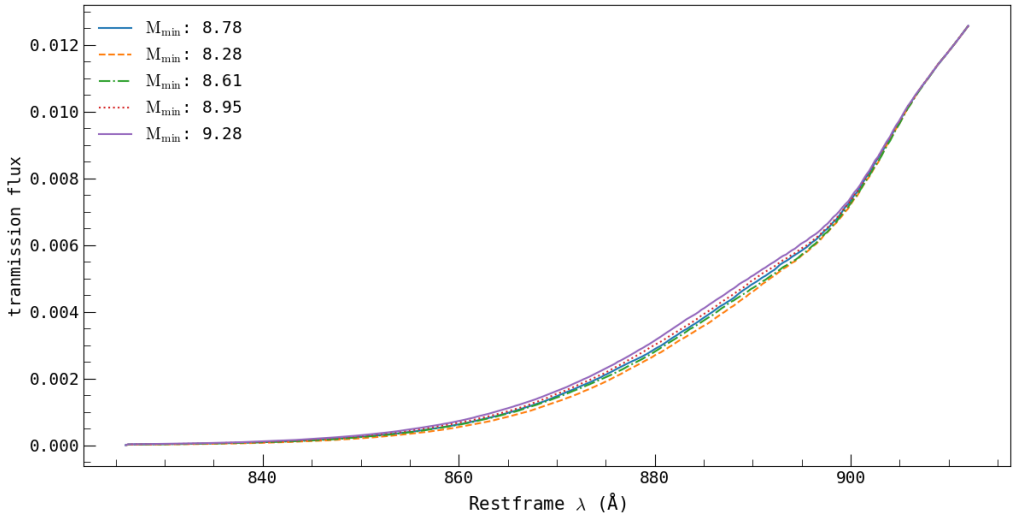


Figure 4.7. Similar to the Fig. 4.5, in this Figure we plot the observed flux as a function of the restframe LyC wavelength for a range of M_{min}. There is a negligible effect of M_{min} on the final observed flux.

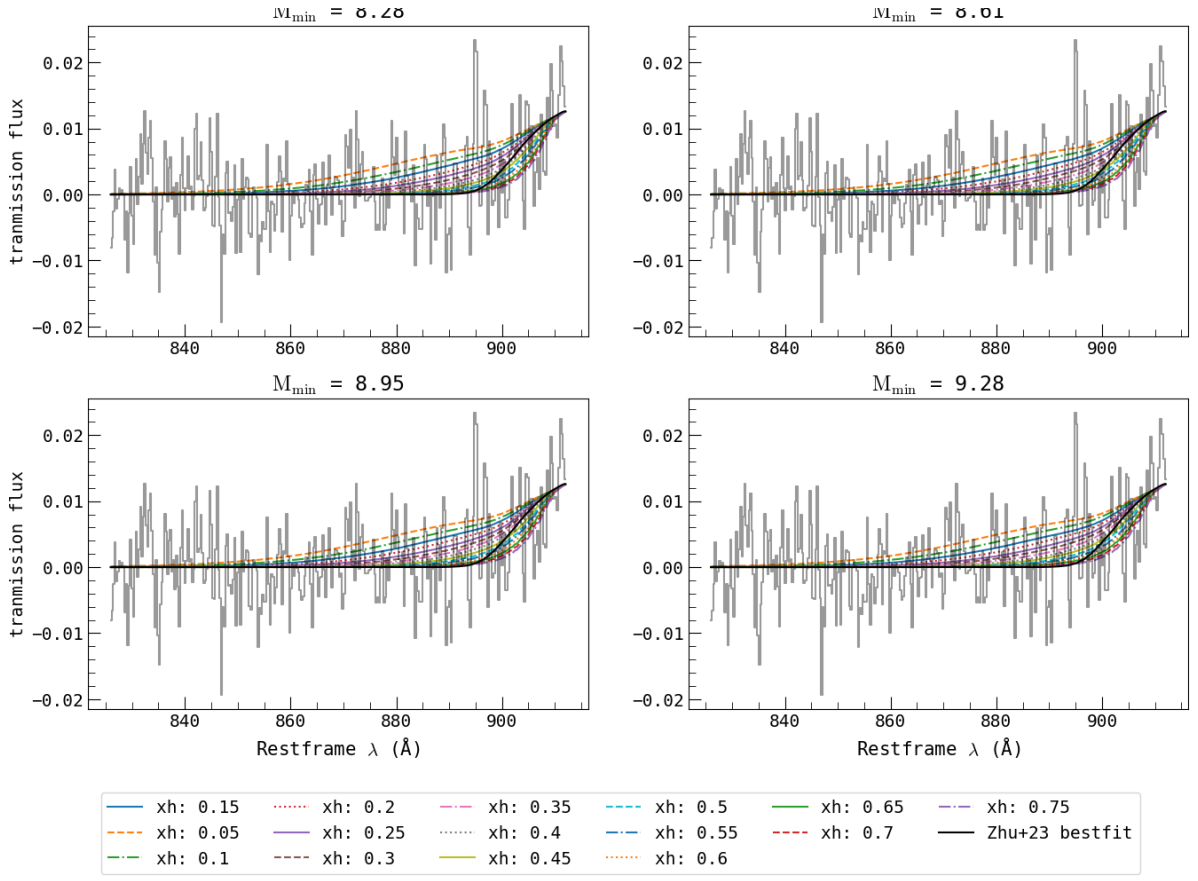


Figure 4.8. In this Figure, we plot the transmission/ final observed flux over the range of M_{min} for all x_{HI} over the quasar spectra, along with the best fit model from (Zhu et al., 2023) (solid black line). The distribution of x_{HI} curves from one plot to another (or for different M_{min}) remains more or less the same.

LyS absorptions. We then calculate the LyC MFP ($\lambda_{\rm MFP}^{\rm LyC}$) by calculating the distance at which the general flux, when attenuated with the LyC absorption, drops to 1/e of its maximum value. We will then fit our models to the quasar spectra as shown in the Fig. 4.4 along with the best fit model from (Zhu et al., 2023) to figure out the range of parameters which could explain the short MFP at redshift z = 6. Unless stated, we will use the fiducial values of the astrophysical parameters as shown in the Table 4.1 for the rest of our plotting.

We begin by looking at the $M_{\rm qso}$ dependency of the observed flux as a function of the rest-frame LyC wavelength regime. In the Fig. 4.5, we see that the overall flux does not vary much with the choice of the $M_{\rm qso}$. It can also be seen from Fig. A.6, where the ionized bubble around the halo rises slowly with the halo mass. Thus, we can assume the $M_{\rm qso} = 10^{11.52} M_\odot$, to be constant for the rest of our study.

We will now look at the parameters that can alter the reionization topology. Namely, $x_{\rm HI}$, $M_{\rm min}$, and $t_{\rm q}$. Where $x_{\rm HI}$ and $M_{\rm min}$ change the global topology, $t_{\rm q}$ dominates the local topology. As we have described already, $x_{\rm HI}$ changes the overall HI fraction, $M_{\rm min}$ redistributes the ionized bubbles over the permitted halo-mass ranges. Since the effect of $M_{\rm qso}$ is negligible on the final observed flux, we expect $M_{\rm min}$ also to have significantly less impact on the final observed flux.

In Fig. 4.6, we show the distribution of the ionized bubbles around the quasar for different values of $M_{\rm min}$. The first five plots show the histogram of the ionized bubbles with mean radius $\langle r \rangle$, and the median radius $M(r)$. On the sixth plot, we plot the $\langle r \rangle$ (orange curve) and $M(r)$ (blue curve) radii of ionized bubbles for all the values of $M_{\rm min}$. As expected, we see that both the $\langle r \rangle$ and $M(r)$ grow as $M_{\rm min}$ is increased. This is because as we increase the $M_{\rm min}$ we allow only the massive halos, which are exponentially less in number, to form the ionized bubbles, while keeping the $x_{\rm HI}$ the same. This results in the formation of coarser and larger bubbles. Whereas, when we allow even the less massive halos to participate in the reionization process, we redistribute the ionized bubbles to an exponentially larger set of halos. This results in the formation of smaller and finer bubbles. As shown in the sixth panel of the Fig. 4.6, this redistribution of the bubbles over the $M_{\rm qso}$ increases the effective bubble size slowly. Thus, the effect of $M_{\rm min}$ can also be neglected and can be kept constant for the rest of the analysis. In the Fig. 4.7 we see that the variation of the final observed flux over the range of $M_{\rm min}$ is negligible, as expected. Finally, in the Fig. 4.8 we plot the distribution of final observed flux for all the values of $x_{\rm HI}$ over the range of $M_{\rm min}$. We also plot the stacked quasar spectra

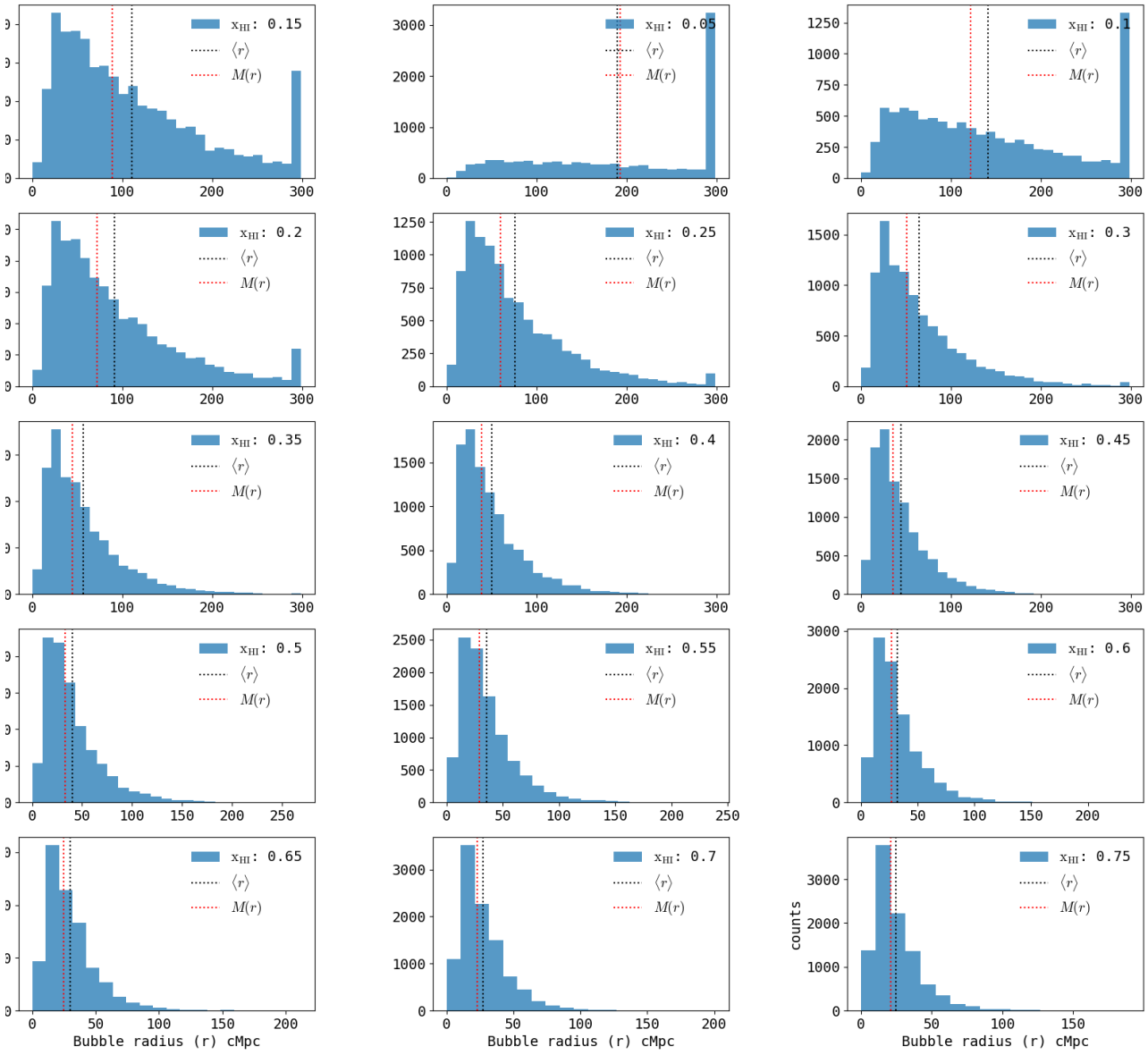


Figure 4.9. In this Figure, we show the distribution of the ionized bubbles as a function of x_{HI}. Each plot shows the histogram of the ionized bubbles radii with their mean radius $\langle r \rangle$, and the median radius M(r). The distribution shifts more and more towards the left (smaller bubble size) as we increase the x_{HI}. The upper limit of the bubble size is 300 cMpc because the maximum distance traveled by our quasar sightline is 300 cMpc.

and best fit model from (Zhu et al., 2023) (solid black line) along with our models. The distribution of x_{HI} curves from one plot to another (or for different M_{min}) remains more or less the same. Even if we change the value of M_{min} drastically (e.g., $M_{min} = 10^{11} M_{\odot}$), we see minimal changes in the final observed flux.

Now we look at the x_{HI}, another parameter affecting the global reionization topology as shown in the Fig. 4.1. In the Fig. 4.9, we plot the histogram of ionized bubble

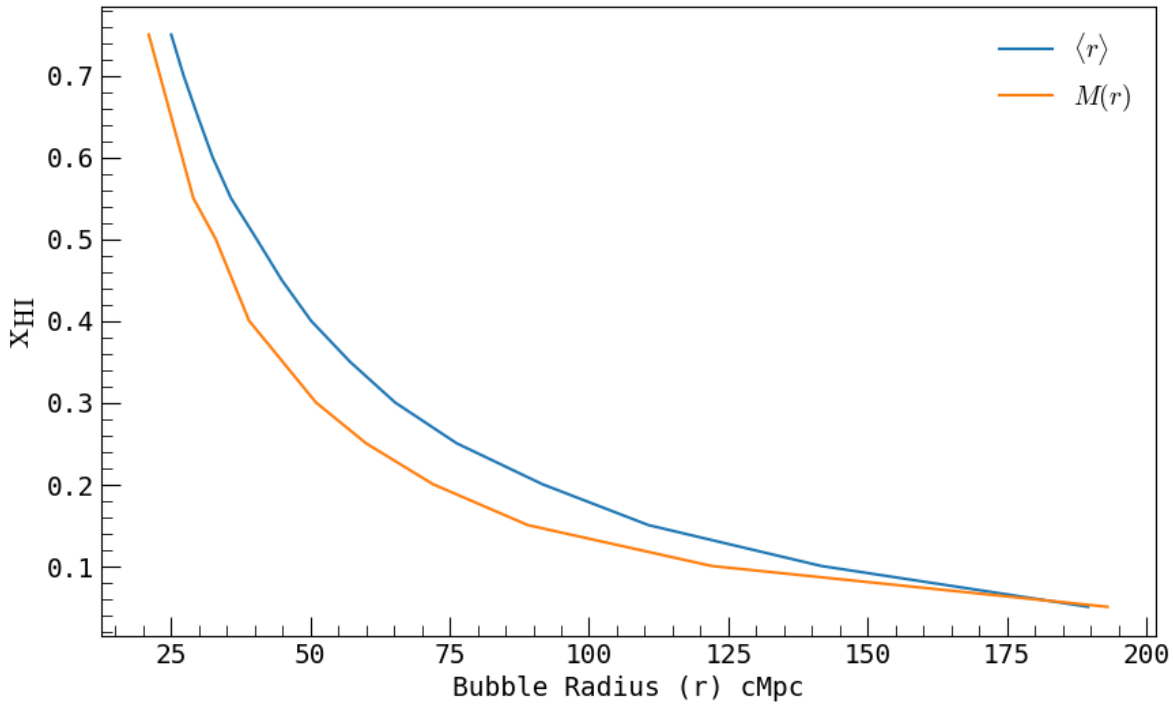


Figure 4.10. In this Figure, we plot the mean bubble radius $\langle r \rangle$ (blue curve) and the median bubble radius M(r) (orange curve) radii of ionized bubbles as a function of x_{HI}.

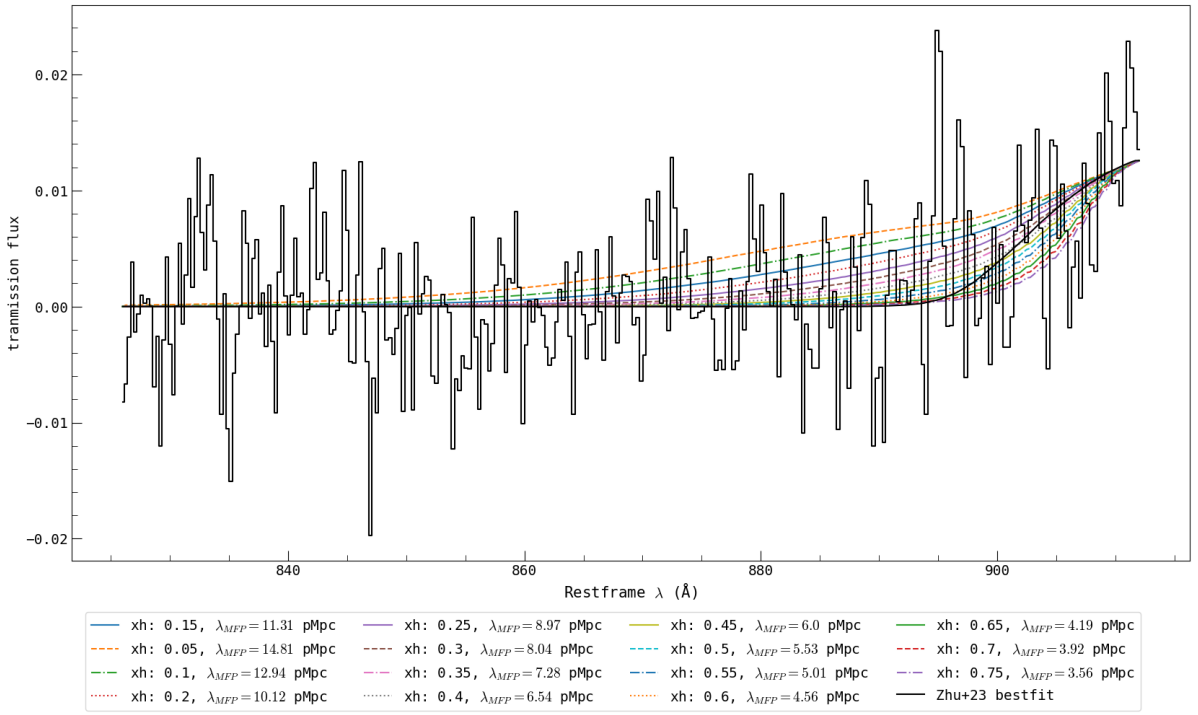


Figure 4.11. In this Figure, we plot the final observed flux for all x_{HI} over the quasar spectra, along with the best fit model from (Zhu et al., 2023) (solid black line). As we increase x_{HI} and hence the number of HI atoms, the damping of the transmission flux increases rapidly. The models with $x_{HI} \geq 0.35$ provide a much better fit to the quasar spectra.

radii distribution for all the values of x_{HI}. While in the Fig. 4.10 we plot the mean bubble radius $\langle r \rangle$ (blue curve) and the median bubble radius $M(r)$ (orange curve) radii of ionized bubbles as a function of x_{HI}. From these two figures, we can see that the x_{HI} significantly affects the ionized bubble distribution. In the Fig. 4.11, we plot the final observed flux for all x_{HI} over the quasar spectra, along with the best fit model from (Zhu et al., 2023) (solid black line). Since increasing the x_{HI} increases the total number of HI atoms, increasing the attenuation. We can produce shorter λ_{MFP}^{LyC} with higher x_{HI}. Our models with $x_{HI} \geq 0.35$ provide a much better fit to the quasar spectra, while the models with $x_{HI} \in [0.35, 0.6]$ are very close to the best-fit model from (Zhu et al., 2023).

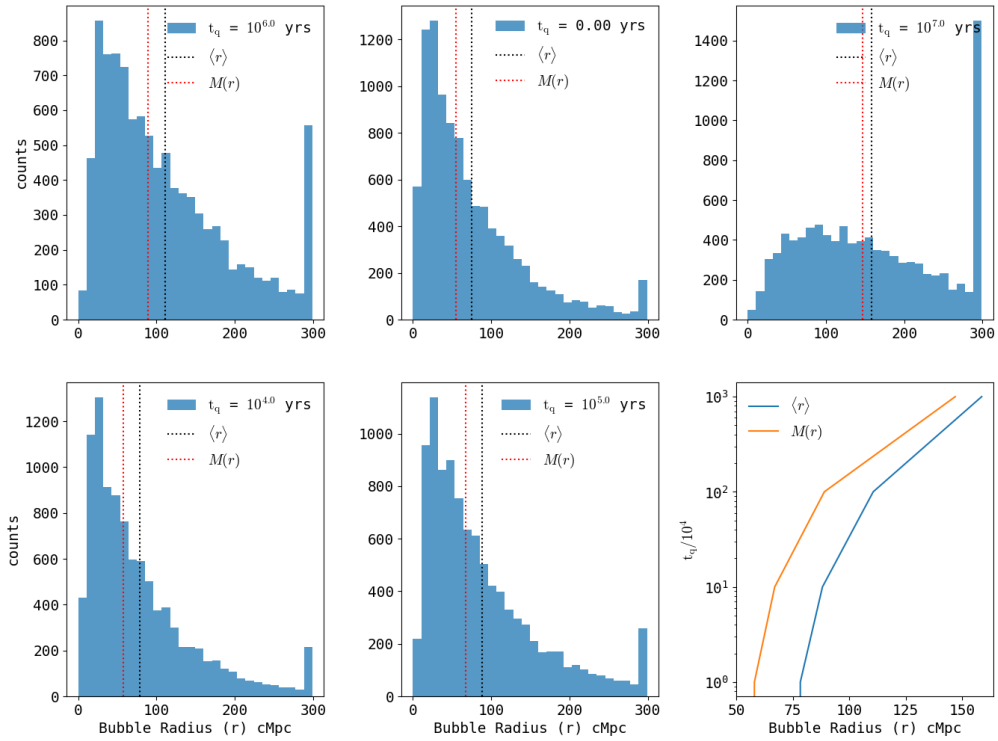


Figure 4.12. In this Figure, we show the distribution of the ionized bubbles as a function of t_q. The first five plots show the histogram of the ionized bubbles radii with their mean radius $\langle r \rangle$, and the median radius M(r). On the sixth plot, we plot the $\langle r \rangle$ (blue curve) and M(r) (orange curve) radii of ionized bubbles for all the values of t_q. As expected, we see that both the $\langle r \rangle$ and M(r) grow rapidly as t_q is increased.

Now let's look at the effect of t_q, a parameter that affects the reionization topology locally. From our modeling, t_q carves an additional bubble proportional to the quasar luminosity (see Eq. (3.26)). This changes the size of the ionized bubbles drastically. In

the Fig. 4.12, we plot the distribution of the ionized bubbles around the quasar for various t_q values. In the sixth plot of the same figure, we can see that the $\langle r \rangle$ (blue curve) and M(r) (orange curve) radii of ionized bubbles grow rapidly as t_q is increased, as expected. In the Fig. 4.13 we show the distribution of the final observed flux with t_q for a range of x_{HI}. As noted earlier for high x_{HI} and the fiducial t_q, we can match the best fit model from (Zhu et al., 2023). As we have seen, t_q affects the reionization topology by erasing the HI density within the bubble radius carved by the quasar. If we decrease t_q (or as we bring back the local HI gas), we can get a better match to Zhu et al., 2023 even at lower x_{HI}. Although, even with $t_q = 0$, we would still need $x_{HI} \geq 0.35$ for a good match. This strongly indicates that we need some dense neutral patches or Lyman limit systems within the quasar proximity zones to explain the shorter MFP. Finally, in the Fig. 4.14 we plot the observed flux for the non-fiducial values of t_q, for all x_{HI} over the stacked quasar spectra, along with the best fit model from Zhu et al., 2023. We see that t_q strongly affects the shape of the flux. From $t_q = 0$ yrs to $t_q = 10^4$ yrs, the curves remain more or less the same, hence we can rather keep $t_q = 10^4$ yrs as our minimum quasar lifetime.

4.6 Discussion and Conclusion

We use 21cmFAST to generate patchy reionization boxes at $z = 6$ for a wide range of astrophysical parameters defined in the Table 4.1. We describe a model as a unique combination of these parameters. Within each model, we locate halos for a set of mass bins ranging from $M_{qso} = 10^{10} - 10^{12} M_\odot$. For each mass bin, we create a set of 10,000 randomly directed halo-sightline pairs. We then look at the distance to the first neutral pixel on every sightline and define the transmission flux as 1 until the first neutral pixel and 0 afterwards. We then stack and normalize all these fluxes for a given mass bin to generate the general transmission flux as a function of M_{qso} and the other astrophysical parameters. The wavelengths shorter than the $\lambda_{LL} = 912$Å will suffer the LyC absorption from HI atoms present in the IGM. To account for this, we modify this general flux with the LyC attenuation by modeling it using the equations from Eq. (4.23) to §4.4.1.

Some of these LyC photons could travel far enough to be redshifted into the LyS regime and will suffer from the LyS absorption attenuation. Thus, we need to account for this foreground LyS absorption, and we do so by using the effective LyS attenuation optical depth as shown in the §4.4.2. The combination of the LyC and LyS attenuation along with the general transmission flux gives us our final observed flux as shown in

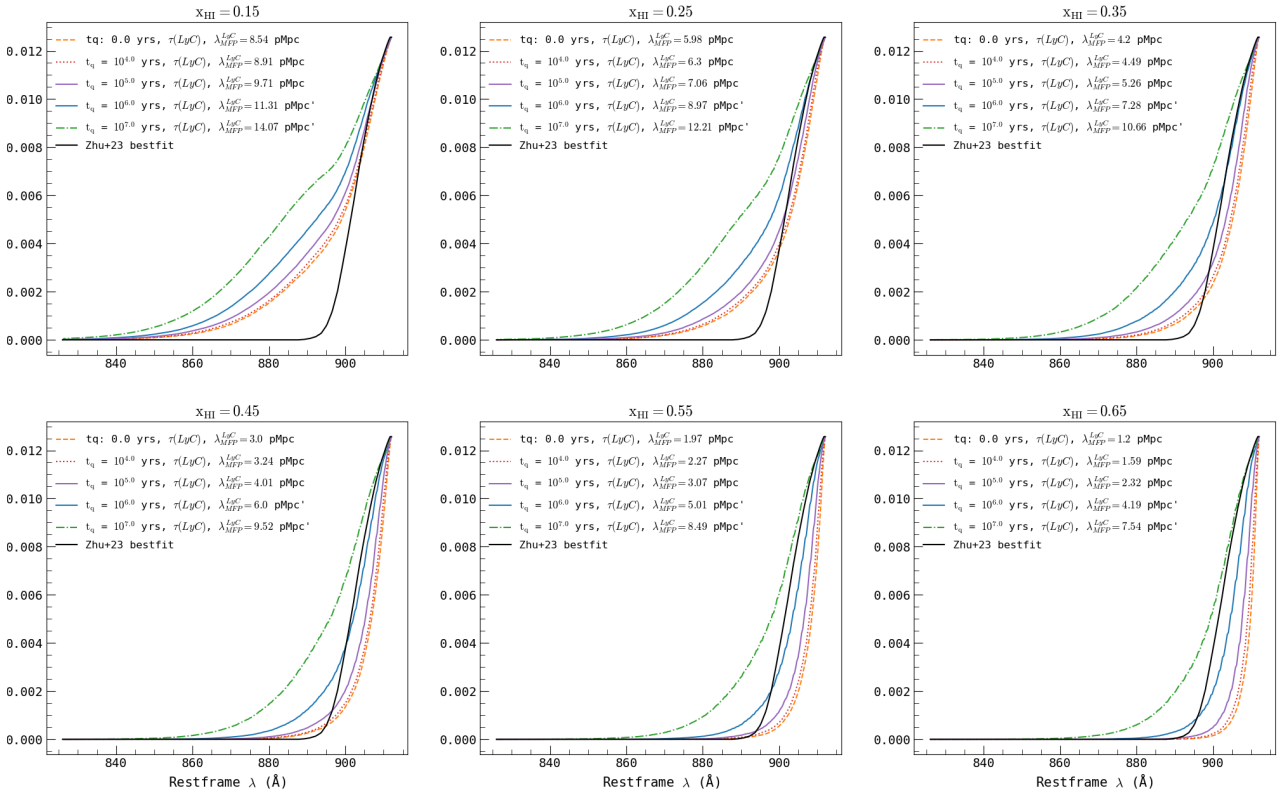


Figure 4.13. In this Figure, we draw a set of plots showing the distribution of the transmission/ final observed flux over the whole range of t_q for a set of x_{HI} values, along with the best fit model from (Zhu et al., 2023) (solid black line). As we decrease t_q, i.e., as we introduce back the local HI gas, we get good fits to (Zhu et al., 2023) at lower x_{HI}. But, even with $t_q = 0$, we would still need $x_{HI} \geq 0.35$ for a good match. This indicates that we need some dense HI clouds or Lyman limit systems within the proximity zone of the quasars to explain the short MFP for LyC photons.

the Eq. (4.22), with some normalization constant as a free parameter. We fix the normalization constant by normalizing our final observed flux to match the flux of the best fit model from (Zhu et al., 2023) at $\lambda_{obs} = 912$Å. We then plot all our final flux models as a function of restframe wavelength for all the astrophysical models against the quasar spectra for the set of quasars shown in the Table 4.2 and the best fit model from (Zhu et al., 2023). For this project, we did a qualitative comparison to get a rough estimate for the ranges of the models we would need to fit the quasar data.

While we compare our results with (Zhu et al., 2023), one significant distinction exists between our models. In (Zhu et al., 2023), they assumed a very short IGM MFP (i.e., large κ_{bg}^{LyC}), which causes their model to transition from low to high transmission flux rapidly. Whereas in our case, we assumed a much longer MFP (short κ_{bg}^{LyC}) (Worseck

et al., 2014; Qin et al., 2025), due to which our models tend to have a long tail behaviour. In the Appendix A.4 we demonstrate the same idea by varying $\kappa_{\rm bg}^{\rm LyC}$.

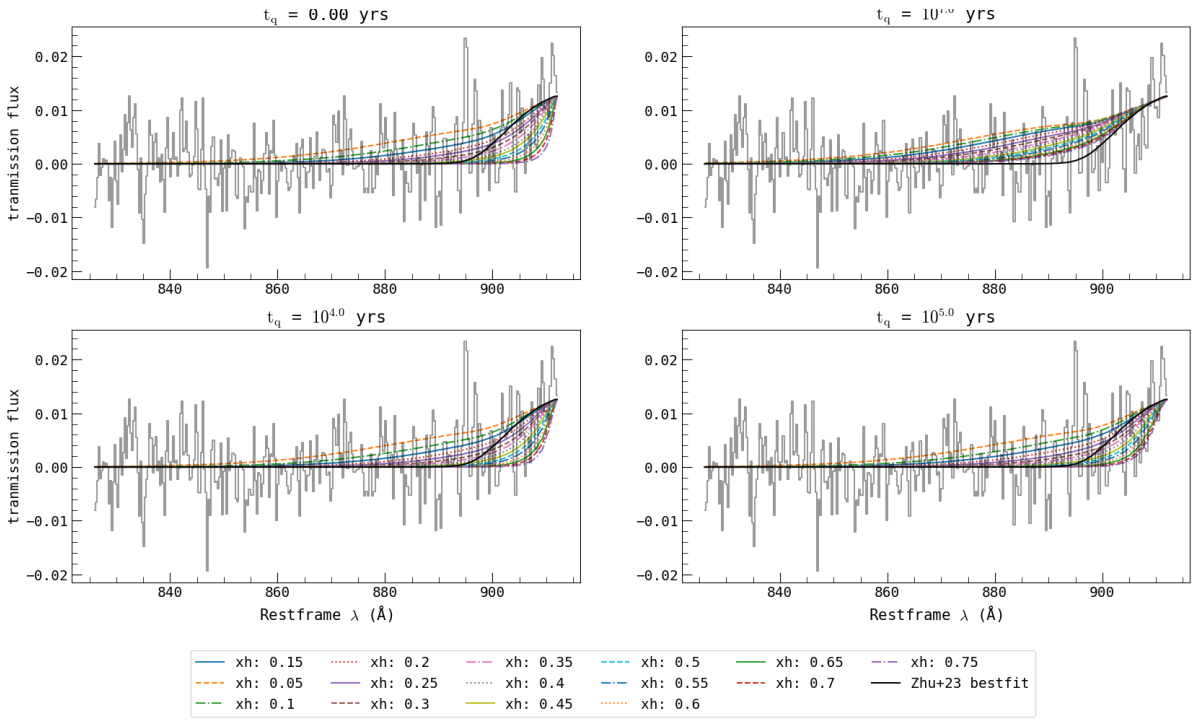


Figure 4.14. In this figure, we plot the transmission/ final observed flux over the range of t_q for all x_{HI} over the quasar spectra, along with the best fit model from (Zhu et al., 2023) (solid black line). The distribution of x_{HI} curves from one plot to another varies a lot with t_q.

From our analysis, we find that M_{qso} and M_{min} have negligible impact on the final observed flux, i.e., our flux is independent of the source models. x_{HI} and t_q, on the other hand, affect both the shape and MFP of the observed flux significantly. Even if we reduce t_q to 0, we still need $x_{HI} \geq 0.35$ to get a good match to both the data and the best fit from (Zhu et al., 2023). The idea of reducing t_q is the same as adding more neutral pixels within the quasar proximity zone, since we define t_q by erasing all the neutral topology in the vicinity of the quasar until a characteristic bubble radius R_{t_q} (see the Eq. (3.26)). Thus, if we assume just late reionization models (i.e. there is some additional effective opacity from the presence of IGM-scale neutral islands around the edges of the ionized bubbles), we would still need a very high x_{HI}, much higher than the currently accepted range of x_{HI} to justify the short MFP at $z = 6$ (Gaikwad et al., 2023; Qin et al., 2025; Cain et al., 2024). Whereas we know it has to end around $z \approx 5.3$ (Bosman et al., 2022). This indicates that we will need some dense HI gas or the Lyman limit systems within the proximity zone of the quasars to explain such a short MFP of the LyC photons.

There are still some caveats to our analysis; we use the opacity $\kappa_{\rm bg}^{\rm LyC}$ from (Worseck et

al., 2014), which was derived by extrapolating their MFP to z = 6. We can, in principle, treat $\kappa_{\rm bg}^{\rm LyC}$ as a free parameter and jointly vary with $x_{\rm HI}$, and $t_{\rm q}$, then calculate its value by fitting to the quasar data. In the future, we plan to quantitatively fit all our models to the individual bootstrapped sample of all the quasars from (Zhu et al., 2023) to calculate the distribution of models that would fit the data the best. Nonetheless, with our study, we have shown that the current status of the short MFP of LyC photons cannot be explained by late reionization models alone, and one needs to incorporate Lyman limit systems within the quasar proximity zone to account for such strong absorption. Similar to the (Roth et al., 2024), we plan to include the contribution to $\kappa^{\rm LyC}$ from the distribution of neutral islands near the quasar. Including neutral islands' contribution along with variable $\kappa_{\rm bg}^{\rm LyC}$ will give us more accurate constraints on the parameters and the LyC MFP.

Take away points

- $M_{\rm min}$ and $M_{\rm qso}$ have negligible effect on the final observed flux after the LyC and LyS attenuation
- The mean of bubble size distribution increases rapidly after $t_{\rm q} = 10^5$ yrs.
- We need $x_{\rm HI}$ as large as 0.5 with $t_{\rm q} = 10^6$ yrs to match the best fit model from (Zhu et al., 2023)
- Even with $t_{\rm q} = 0$, we still need $x_{\rm HI} \approx 0.35$. Thus indicating the need of Lyman limit systems within the proximity zone.
- Late reionization models alone are insufficient to explain the short photoionization MFP at $z = 6$.

5 Conclusion

Throughout the thesis, we emphasised the strength of damping wings as an observational tool to decode the reionization topology at high redshifts. Their collective spectra can constrain the source parameters and help relax the degeneracy between all the astrophysical parameters. In this chapter, we will briefly summarize the results of Chapter 2, Chapter 3, and Chapter 4 and discuss the future prospects based on the results of these studies.

5.1 Summary

In Chapter 2, we studied how the imprints of astrophysical parameters on the reionization topology can be seen in the damping wings spectra of quasars. To study the effects, we first need to generate an extensive suite of reionization boxes as a function of these parameters, subjected to the same initial and boundary conditions. The parameters we focused on, in this chapter, are described in the Table 2.1. We define a model as a box generated by a unique combination of astrophysical parameters. We used 21cmFAST to generate this suite of patchy reionization models and located the halos of desired mass within these models. From each mass bin, we drew a set of 10,000 halo-sightline pairs. Modeling up to this point is common in all projects (chapters).

In Chapter 2, using these halo-sightline pairs, we calculated the Lyα damping wing optical depth along the sightline, treating it as the line of sight skewer. We then plot the median (M(DW)) and 68-percentile sightline to sightline scatter around the median (ΔSW_{68}) for each combination of the astrophysical parameters. By looking at the impact of the parameters on both the shape and strength of M(DW) and ΔSW_{68}, we selected x_{HI}, t_q, M_{min}, and M_{qso} as the final set of parameters for further study. Although

M(DW) damping wing signal correlates strongly to the changes in both x_{HI} and t_q , it also creates the issue of degeneracy. To counter this, we used ΔSW_{68} to qualitatively differentiate between the signatures of x_{HI} and t_q on the ensemble of damping wing spectra. Another significant result of this study was that we observed noticeable effects on the M(DW) and ΔSW_{68} from the source parameters, namely, M_{min} and M_{qso}. We established that by using an ensemble of damping wings, we cannot only see the signature of the multiple astrophysical parameters but also look at the scatter of this ensemble to help relax the degeneracy between all our selected parameters.

Continuing with the models developed in Chapter 2, we described how an ensemble of damping wings can provide much more relevant information than just a single spectrum. But the set of 10,000 seems far-fetched at present. We want to give a more reasonable number for the size of this ensemble. We began Chapter 3 in the search for the same. In this Chapter, we quantitatively explored the constraining prowess of an ensemble of Lyα damping wings using the FIM. We used the models established in Chapter 2 for the filtered set of parameters shown in the Table 3.1 to derive our observables (M(DW), ΔSW_{68}, and their combination) for the FIM analysis. We also modified the damping wing profiles with the effects of the proximity zone, spectral noise, and continuum noise for a more realistic analysis. We selected a batch of $N_{quasars}$ number of quasars from the initial ensemble of 10,000 to calculate the FIM constraints. We found that for $N_{quasars} \geq 32$, the Gaussian likelihood assumption for the FIM is a valid approximation. We also observed that for $N_{quasars} \approx 128$ our constraints on M_{min} are as good as those obtained from 21 cm studies (Park et al., 2019; Charlotte A. Mason et al., 2023). Furthermore, by extrapolating the constraints at $N_{quasars} = 256$, assuming the Gaussian likelihood, with $\sigma(N_{quasars} = 256) \times \sqrt{256}/\sqrt{N_{quasars}}$ curve for x_{HI} we get for $N_{quasars} = 1$ $\sigma(x_{HI}) = 0.20$ and for t_q we get $N_{quasars} = 1$ $\sigma(t_q) = 0.95$, which are close to the values obtained in Kist et al., 2024 with full-continuum and full-simulation model for a single quasar spectra. For $N_{quasars} = 64$ quasars at redshift 7, we got the following constrains; $x_{HI} = 0.5^{+0.02}_{-0.02}$, $M_{min} = 8.78^{+0.53}_{-0.53}$, $\log t_q/\mathrm{yr} = 6.0^{+0.12}_{-0.12}$, and $\log M_{qso}/M_\odot = 11.52^{+0.32}_{-0.31}$.

In Chapter 4, we take a slightly different route from calculating Lyα damping wings. The strength of the model routine isn't just producing an ensemble of damping wings, but also making an extensive suite of models for a wide range of astrophysical parameters. This chapter aims to figure out the range of parameters that could explain the short MFP of the ionizing photons at $z = 6$. We began by calculating the LyC transmission flux across a wide range of models at $z = 6$ using the ensemble of halo-sightline pairs

derived from the methodology of Chapter 2. We then stacked the step-wise transmission fluxes from all the sightlines within a given model and , where the step occurs at the position of the first neutral pixel. We multiplied this stack with the attenuation from LyC and foreground LyS absorptions from the IGM, blueward of λ_{LyC}= 912Å. We called this flux the "final observed flux". We compared our final observed flux against the best-fit model from Zhu et al., 2023 and the stack of normalized quasar spectra. With this analysis, we showed that among our filtered parameters x_{HI} and t_q, both affect the shape and MFP of the observed flux significantly. We found out that to match the observed quasar spectra, we need very high x_{HI}, much higher than the currently predicted values of x_{HI} at $z = 6$ Gaikwad et al., 2023; Cain et al., 2024; Qin et al., 2025. Even if we reduce t_q, which increases the neutral fraction in the vicinity of the quasar, the best fit values of $x_{HI} \gtrapprox 0.35$. Our study indicates that the current measurements of short MFP photons cannot be explained by late reionization models alone; one needs to include the contribution from the dense Lyman limit systems within the proximity zone.

5.2 Future Prospects

In the future, we will expand upon the applications of the suite of models that we developed in this thesis. Although we developed models with $t_q = 0$, we would like to compare the spectra of halos without quasar lifetime to the galaxy damping wing spectra. Since the abundance of galaxy damping wing spectra is much larger than current quasar spectra, this case study will be instrumental in constraining the astrophysical parameters. We will also work on developing an emulator based on our collection of models to emulate the effective M(DW) and ΔSW_{68} profiles for any given combination of the selected parameters. This Emulator will be helpful when we want to do the parametric fit (finding the range of all the parameters that best fit the data) of the quasar damping wing spectra data with our models.

While the results from Chapter 2 and Chapter 3 have compelling results, some caveats remain to be addressed. In our proximity zone model of calculations, we neglected the non-equilibrium effects, flickering effects Frederick B. Davies et al., 2020; Satyavolu et al., 2023; Zhou et al., 2024, and heating from Helium reionization. We will address these additional effects in the future extensions of this project. In Chapter 4 we used the long IGM MFP for the ionizing photons ($\kappa_{bg}^{LyC} = (1/42)h_{70}$), which was an extrapolation from Worseck et al., 2014. For future work, we will treat κ_{bg}^{LyC} as a fitting parameter

and instead derive its value by fitting our model to the quasar spectra. We will also incorporate the effects of the neutral islands, as shown in Roth et al., 2024, for the calculations of both LyC and foreground LyS optical depths. We expect that including the contribution from neutral islands will give us more reasonable ranges for x_{HI}. Finally, we will quantitatively fit our models to the quasar spectra to derive more accurate constraints on the parameters and the MFP of the ionizing photons.

Our work on damping wings has laid the groundwork for ambitious parametric exploration of reionization physics using an ensemble of quasars. It motivates the search for high redshift quasars with the upcoming surveys such as Euclid, Roman, LSST, and in particular, motivates extensive follow-up spectroscopic campaigns with JWST/ELT. Our simulation framework is independent of the choice of the observables and can be easily modified to suit other light sources, such as galaxies. It is also independent of the particular choice of observables. We demonstrated how, with the same set of sightlines, we can constrain both Lyα damping wings and LyC photons MFP. Thus, enabling more comprehensive analysis of other quasar-focused probes and other light sources.

Appendices

A Appendix

A.1 Convergence Test

Throughout this work, we have emphasized the crucial role that large-scale structures play in the analysis. It is then natural to consider what size of the box is required for the strength and variance of the damping wing signal to converge. A small box may fail to capture high-mass halos and large structures essential for analyzing sightline-to-sightline scatter. On the other hand, larger boxes are more computationally expensive, inhibiting exploration of the full parameter space. To identify an optimal box size, we perform a convergence test of box size at fixed spatial resolution.

We begin by comparing the *mean optical-depth* of damping wing profiles for halos at the highest and lowest common mass from each box. By analyzing the mean optical depth, we aim to understand the average reionization topology surrounding a given halo. In Fig. A.1, we observe that for low halo masses ($M_{qso} \approx 4 \times 10^9 M_\odot$), the optical depth damping wing profiles converge within a 10% error limit. This indicates that for the range of box sizes considered here, the topology and corresponding damping profiles for low-mass halos remain consistent and stable.

However in Fig. A.2, the convergence test for high-mass halos ($M_{qso} \approx 1 \times 10^{12} M_\odot$) shows an unexpected trend. We observe that the mean optical depth damping profiles increase as we move from our fiducial model with box length $(512\text{Mpc})^3$ up to $(768\text{Mpc})^3$. Beyond this size, however, the profiles start to decrease.

To ensure that this behavior is not a result of selecting a non-representative halo (i.e., not the most massive), but instead stems from selecting the common most massive halo across boxes, we replicate the analysis using the 100 most massive halos within each box. This procedure mimics that used by past works on quasar damping wings

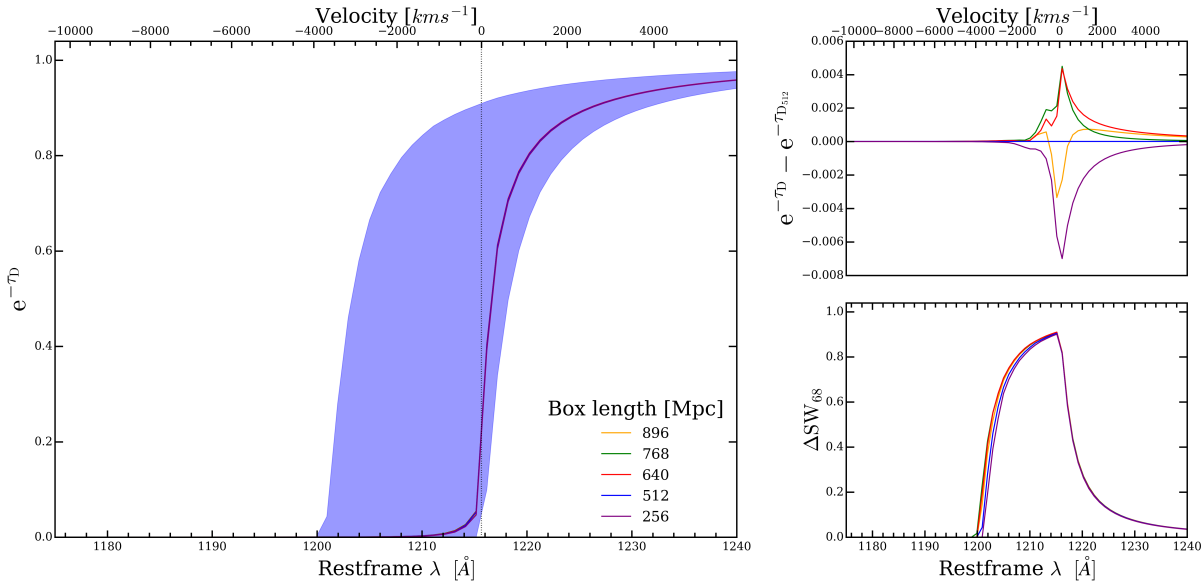


Figure A.1. The convergence test results for the low-mass halos, $M_{qso} \approx 4 \times 10^9 M_\odot$. The upper panel compares the difference between each box's mean optical depth damping wing relative to the fiducial box. The middle panel presents the mean optical depth for each box, with the shaded region representing the scatter for the fiducial box ($L_{box} = 512$ Mpc). The bottom panel shows the ΔSW_{68} of all the boxes, color-coded as in the middle panel. We see that the damping wings in this case seem to converge.

(e.g. Frederick B. Davies et al., 2018). Additionally, to account for uncertainties arising from the randomness of the initial conditions, we generate six different initial condition boxes for each box size and average the results. This analysis is presented in Fig. A.3, where we observe a similar trend. The consistent behavior across different initial conditions and multiple halos confirms that this phenomenon is not due to the choice of halos or initial conditions.

We originally expected that, as box size increases, the average damping effect decreases. This is because high-mass halos are typically located in ionized regions, and larger boxes, containing larger halos, would naturally have bigger ionized bubbles surrounding them. Consequently, damping should continue to decrease, and the deviation from the base model should increase monotonically. However, our simulation results consistently indicate the opposite once the box size exceeds $(768\text{Mpc})^3$.

One possible explanation for this unexpected behavior could be related to the halo mass function for high-mass halos, which decreases exponentially at the highest masses. The increase in the total number of particles (and thus the volume of the box) may not be enough to generate a sufficient number of high-mass halos. As a result, the reionization topology around this exponential tail of massive halos in larger boxes might still be dominated by a mix of smaller and larger ionized bubbles, rather than exclusively by

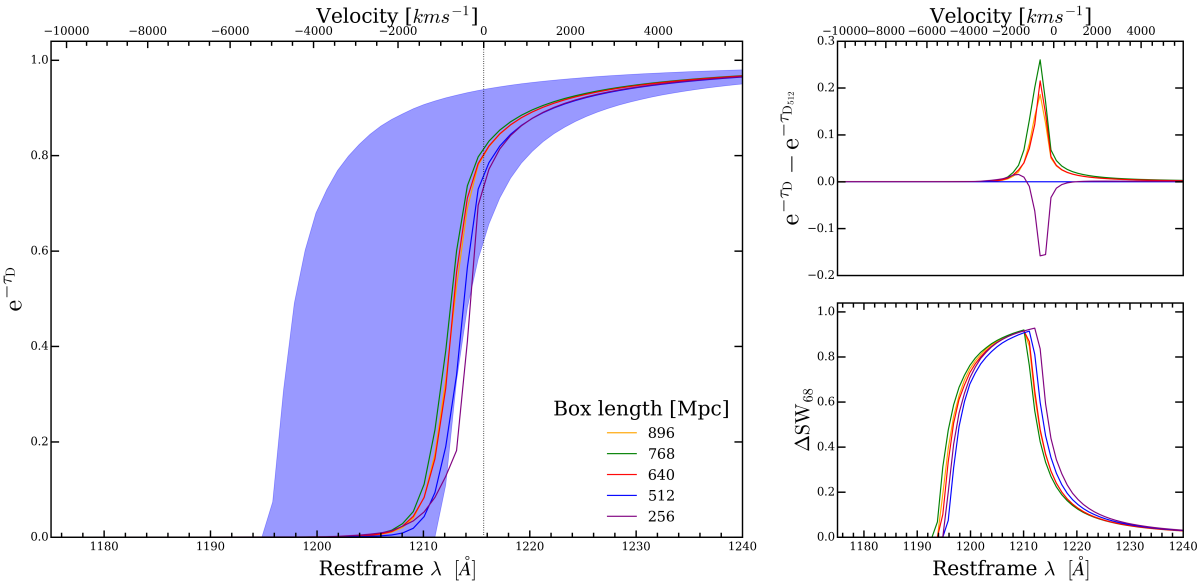


Figure A.2. Similar to Fig. Fig. A.1, we show the convergence test results for the high-mass halos, $M_{qso} \approx 1 \times 10^{12} M_{\odot}$. We see that the damping wings in this case do not seem to converge.

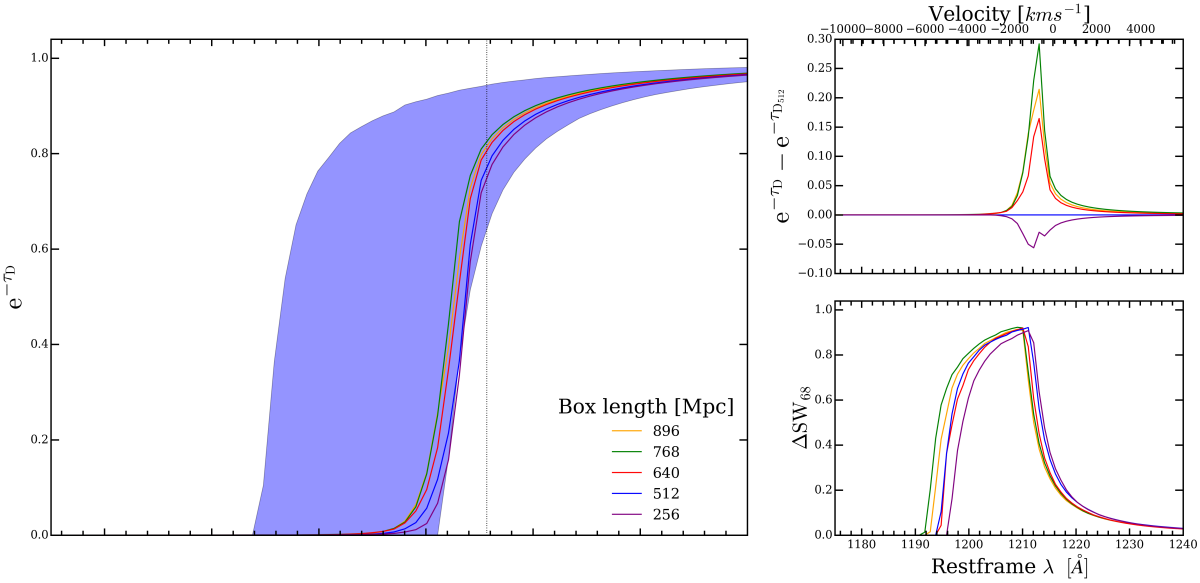


Figure A.3. The convergence test results for the 100 most massive halos from each box. The middle panel shows the mean optical depth damping wing for 100 most massive halos from each box. The upper and lower panels represent the difference between each box's mean optical depth damping wing relative to the fiducial box and the scatter ΔSW_{68}, similar to the Fig. A.1. The damping wings do not seem to converge even in this case either.

large-scale bubbles surrounding massive halos.

In all cases, we maintained a constant mean global neutral fraction ($x_{HI} = 0.5$), which should ensure consistency in the overall ionization state of the simulation. To further test this hypothesis, we eliminate smaller bubbles from forming by setting the minimum halo mass $M_{min} = 10^{11} M_{\odot}$. This will exclude low-mass halos, effectively removing the

contribution of smaller ionized bubbles from the topology.

In Fig. A.4, we observe that when we eliminate the smaller bubbles, the damping wing profiles converge monotonically from $(768\text{Mpc})^3$ to $(896\text{Mpc})^3$ as expected. This finding supports the idea that the issue of convergence likely arises from the interplay between small and large ionized bubbles. When both types of bubbles are present, as in the case of the larger $(896\text{Mpc})^3$ box, this interplay reduces the effective size of the ionized bubbles surrounding high-mass halos, leading to a decrease in the damping signal transmission instead of the anticipated increase.

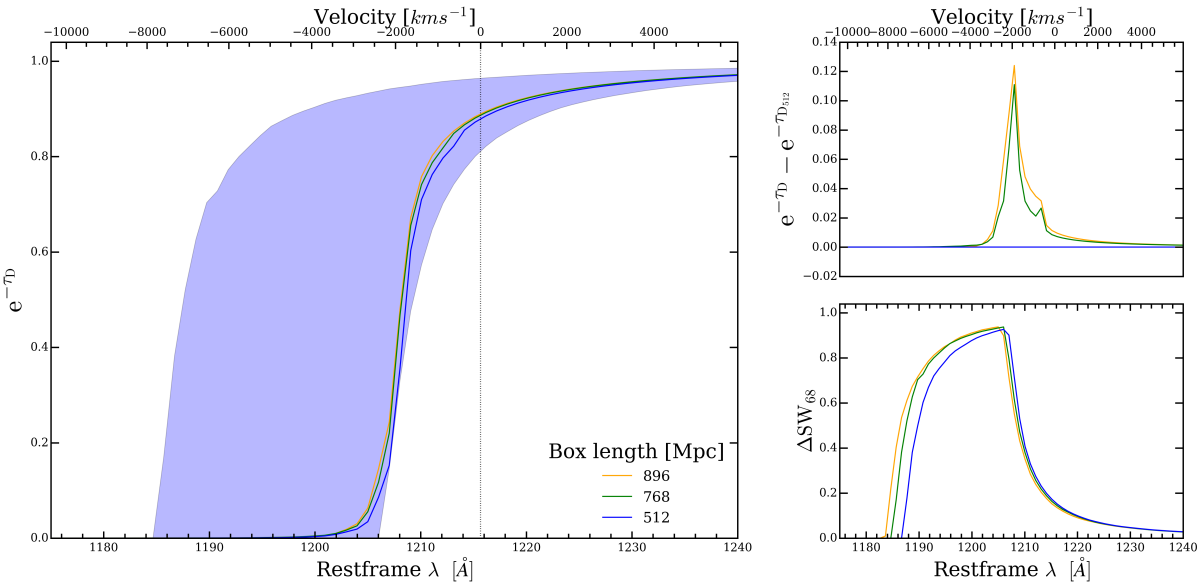


Figure A.4. Similar to the Fig. A.3, we show the convergence test results for the 100 most massive halos from each box with $M_{min} = 10^{11} M_\odot$. We see that the damping wings have started to converge as expected.

To further validate our hypothesis, we can take two approaches: (1) artificially increase the size of the ionized bubbles by turning on the quasar ($t_q > 0$) or (2) calculate the bubble size distribution across different mass bins and box sizes.

In the first case, we repeat the convergence test for the 100 most massive halos, but this time we set $t_q = 1\,\text{Myr}$, effectively increasing the ionized region surrounding each halo. By activating the quasar, we artificially boost the size of the bubbles, thereby mitigating the effect of small ionized bubbles. In Fig. A.5, we show the results of this convergence test. As expected, the damping wing profiles begin to converge within the 10% error limit, which supports the idea that the interplay between small and large bubbles was indeed responsible for the earlier discrepancies in convergence. This indicates that the effective size of the ionized bubbles from high-mass halos had been reduced when small bubbles were present in the simulation, and this effect was mitigated

when the quasar was turned on. Thus, we confirm that for larger boxes, the choice of box size becomes crucial. Care must be taken to ensure that the size of the simulation box adequately captures both large-scale structures and high-mass halos. *For all of our models, the fiducial box already assumes* $t_q = 1\,Myr$*, hence the box length of* 512 Mpc *provides an optimum compromise between large and fast box volumes.*

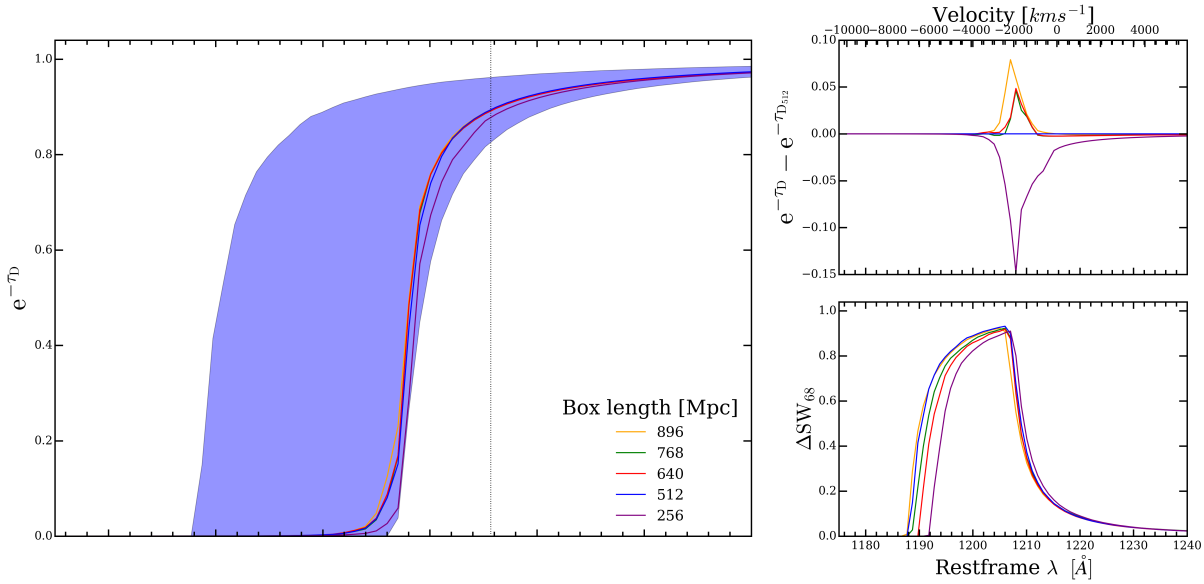


Figure A.5. Similar to the Fig. A.3, we show the convergence test results for the 100 most massive halos from each box with $t_q = 1$Myr. We see that the damping wings here converge within the 10% error limit.

In the latter approach, we calculate the bubble size distribution. As before, we use six different realizations of the initial conditions for each box size, referring to each as a batch for the respective box size. Within each batch, we select 100 halos from each mass bin. For each halo, we generate 10 sightlines in random directions, then measure the neutral fraction x_{HI} along the length of each sightline. If the number of halos in a mass bin is less than 100, we compensate by generating additional random sightlines to ensure that the total number of halo-sightline combinations remains consistent across all mass bins. For each sightline, we clip the length as soon as we encounter a neutral voxel ($x_{HI} = 1.0$), representing the boundary of the ionized region. We then average over all these sightlines for each halo to calculate the spherical average radius of the ionized bubble around the halo. By repeating this process for all halos in the mass bin, we obtain a distribution of bubble sizes for halos of a particular mass. We repeat this procedure for all mass bins across all boxes in each batch, which results in the curves shown in Fig. A.6. As we have seen, the average bubble size increases with the halo mass while the scatter decreases Fig. 2.6.

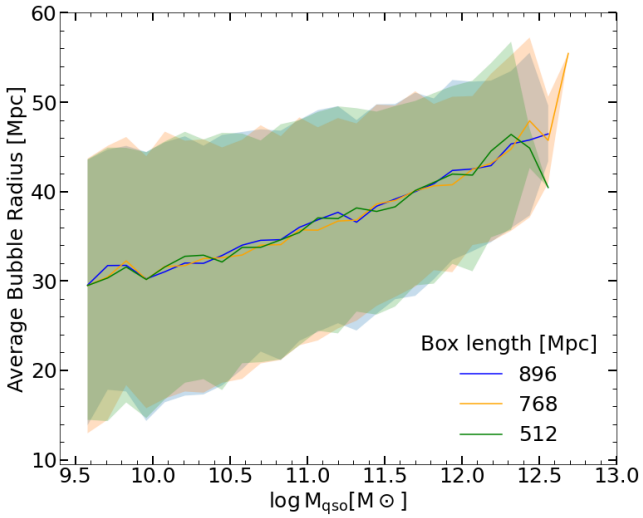


Figure A.6. The distribution of the average bubble radius vs the mass bin for the respective boxes. The shaded region second panel shows the 68 percentile scatter of the radii for each mass bin.

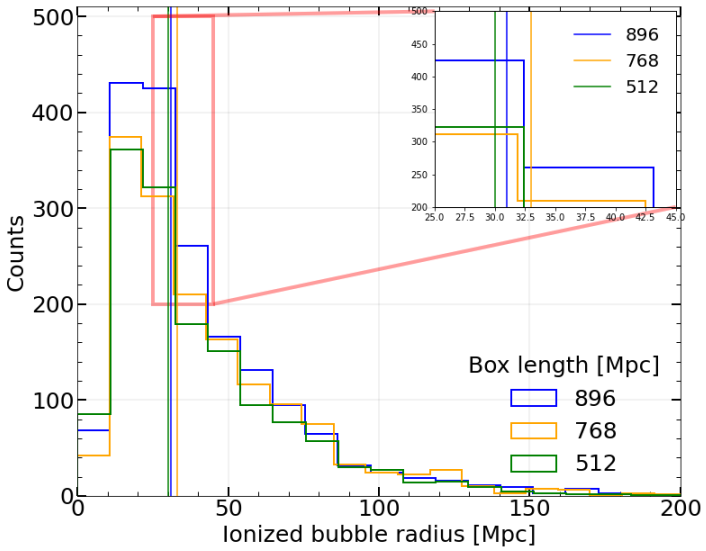


Figure A.7. Histogram of the ionized bubble radius distribution around the most massive halos within each box. The mini panel on this plot shows the zoomed-in picture of the peak of the distribution, where the dashed lines represent the median value of the distribution (color-coded respectively). We see from the zoomed-in section that the box with length 768^3 peaks at a bigger radius than the other two boxes.

We then select the mass bins corresponding to the 100 most massive halos from each simulation box in a batch, concatenate them, and repeat the above calculation to obtain the distribution of ionized region sizes. In Fig. A.7, we plot the aforementioned distribution of ionized radii/sizes. From the inset panel of the same Figure, we observe that the mean of the distribution for the $(768\ \mathrm{Mpc})^3$ box is higher than that of the $(896\ \mathrm{Mpc})^3$ box, thus confirming our hypothesis.

Finally, to conclude, for our fiducial model, we assume $t_q = 1$ Myr. Hence, for all our analyses, the 512 Mpc box converges within the 10% error limit and is sufficient for our study.

A.2 Additional Contour Plots

In this section, we plot the corner plots obtained through the inverse of the FIM for redshifts z = 6 and 7.

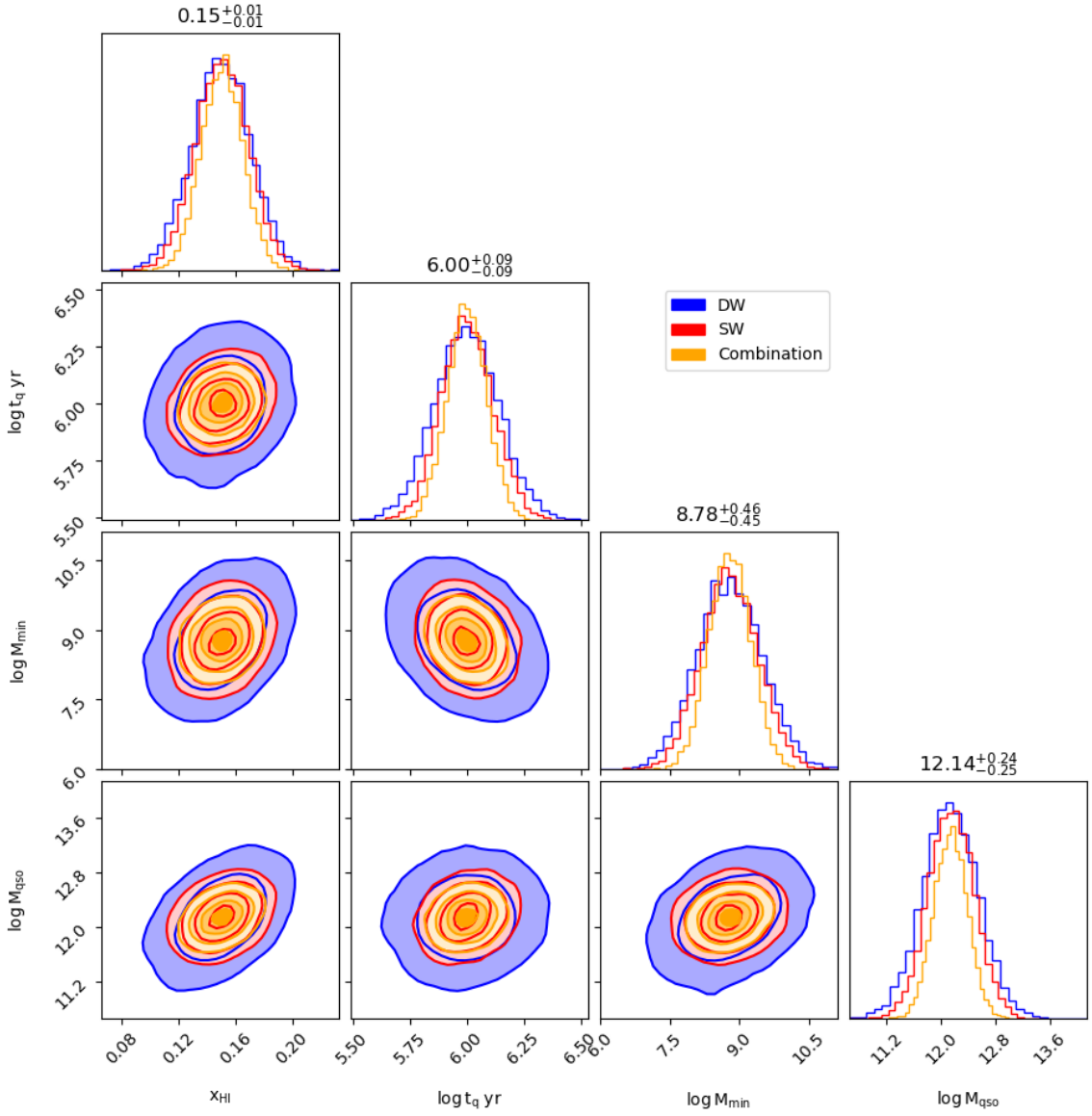


Figure A.8. Similar to the Fig. 3.4, here we show the 1D and 2D marginalized distributions, representing the uncertainties and the covariances of all the parameters, with the contours showing $1, 2\sigma$ confidence intervals, for z = 6, with $N_{quasars} = 64$, SNR − A = 10 and SNR − M = 10.

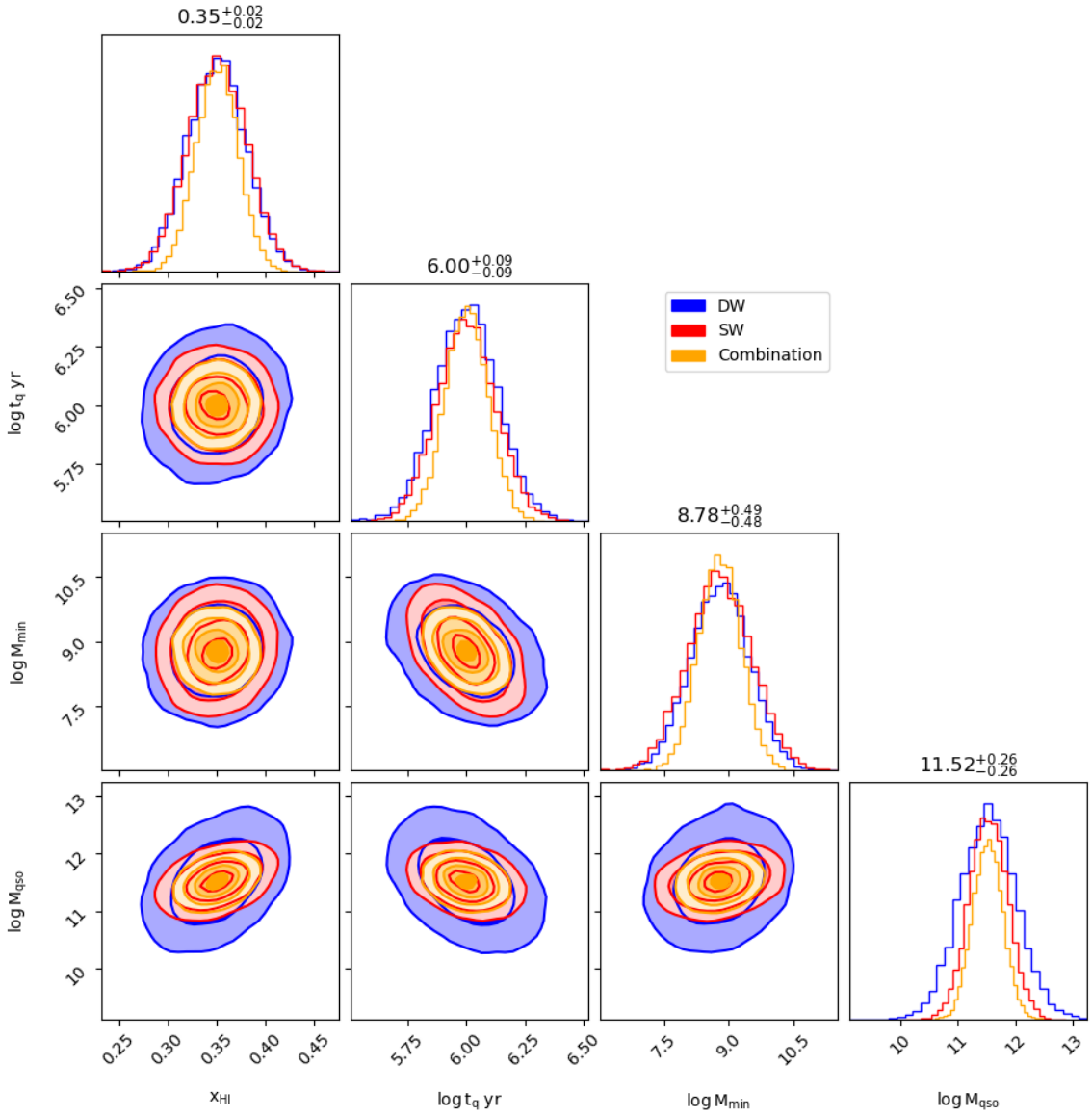


Figure A.9. In this figure, we plot the overlaying corner plot for three different observables, with the parametric constraints stated in the Table Table 3.1. From the top left/ along the diagonal plots show the 1D marginalized distributions, which show the uncertainties on the individual parameter. The off-diagonal plots show the 2D marginalized distributions, showing the correlation between a given combination of parameters. The contours show $1, 2\sigma$ confidence intervals. The blue contours represent 1D, 2D distributions with M(DW) as the only observable; the red contours represent the same for ΔSW_{68} as the only observable, while the orange contours represent the combination of both the M(DW) and ΔSW_{68}. The numbers on certainties are for the combined signal only. These calculations are for z = 6.5, with $N_{quasars} = 64$, SNR − A = 10 and SNR − M = 10.

A.3 High Quasar Host Halo Mass

In this section, we explore the redshift dependency of the constraints on each study parameter as a function of $N_{quasars}$, for higher M_{qso}. This is to study the bias introduced by selecting more and more massive halos and how their location affects the estimation of other parameters. In the figure Fig. A.10, we plot the combined signal for each parameter at three different redshifts $z = [6, 6.5, 7]$, as a function of $N_{quasars}$. As shown in (Sharma et al., 2025), the massive halos live deep in the ionized regions, reducing the overall damping effect and causing less suppression within the proximity zone compared to the low M_{qso}. Due to this, the constraints on M_{qso} at $z = 7$ get slightly better.

While the constraints at $z = 7$ get a little worse for all the cases except for M_{qso}. The constraints get better for M_{qso} due to massive halos only being localised to deeper ionized regions.

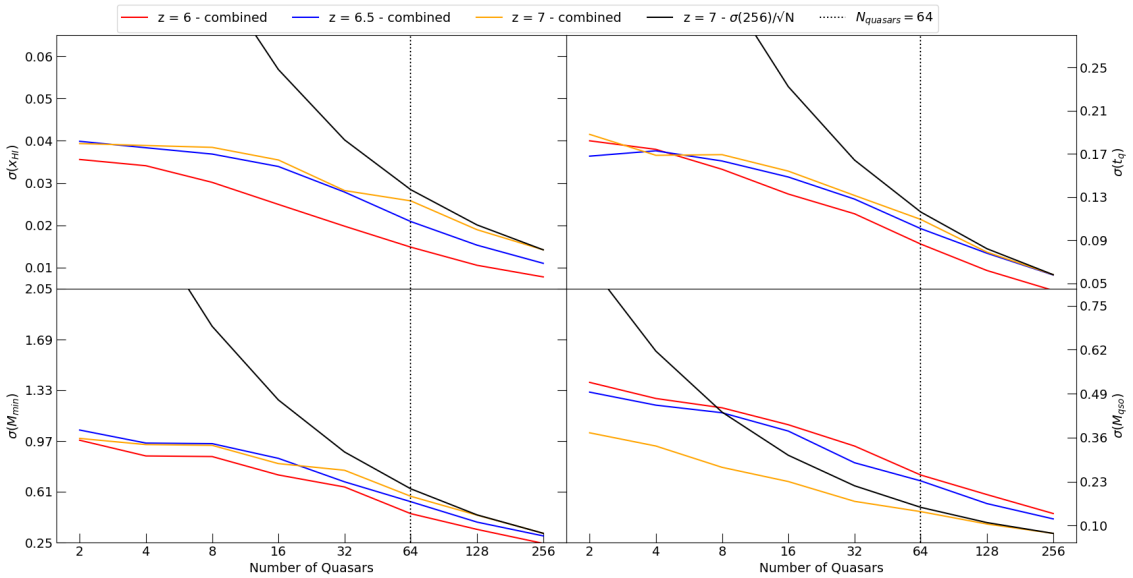


Figure A.10. In this Figure we show the redshift dependency of the constraints on each individual parameter as a function of $N_{quasars}$ for larger $M_{qso} \approx 10^{12} M_{\odot}$. In each plot, the different line styles show the constraints obtained from: the solid lines are for the combined (M(DW) $+\Delta SW_{68}$) signal, the dashed lines are for the M(DW) signal, and the dotted lines are for the ΔSW_{68} signal. Whereas the color-scheme represents the redshifts, red is for $z = 6$, blue is for $z = 6.5$, and orange is for $z = 7$. The upper-left plot shows the variation of constraints on x_{HI} as a function of $N_{quasars}$, the upper-right plot shows the variation of constraints on t_q as a function of $N_{quasars}$, the lower-left plot shows the variation of constraints on M_{min} as a function of $N_{quasars}$, and the lower-right plot shows the variation of constraints on M_{qso} as a function of $N_{quasars}$.

A.4 Plots with short IGM MFP

In this section, we explore how our final normalized transmission flux varies as a function of $\kappa_{\rm bg}^{\rm LyC}$. In the Fig. A.11, we plot our normalized flux with the fiducial model parameters as a function of $\kappa_{\rm bg}^{\rm LyC}$. As expected, we see stronger attenuation for larger $\kappa_{\rm bg}^{\rm LyC}$ and the rise of our curves gets sharper. Our curves still have a much smoother transition as compared to Zhu et al., 2023

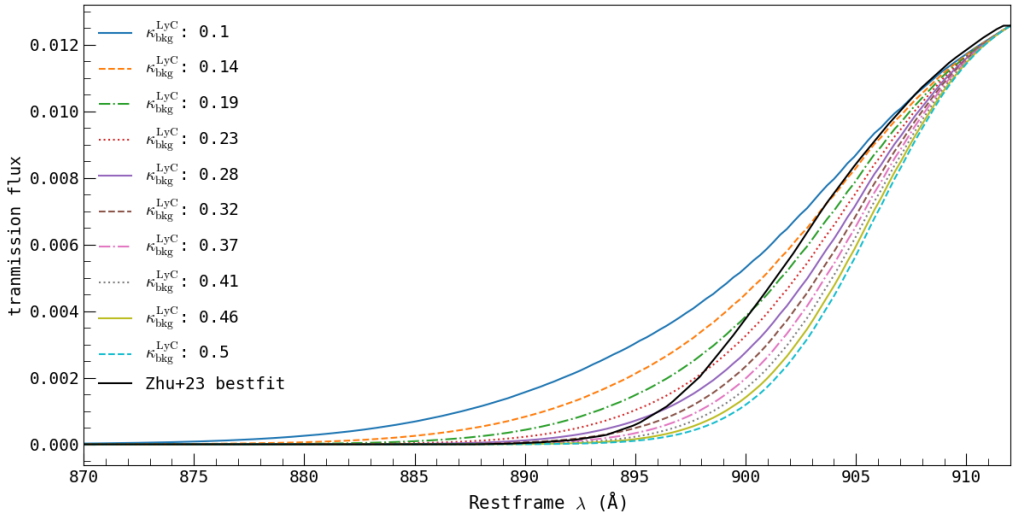


Figure A.11. In this figure, we plot our models of the normalized transmission flux as a function of the $\kappa_{\rm bg}^{\rm LyC}$. We assume the fiducial values for our model parameters, i.e., $x_{\rm HI}$ 0.15, $M_{\rm min}$ $10^{8.78}M_\odot$, t_q 10^6yrs, and $M_{\rm qso}$ $10^{52}M_\odot$
, at redshift $z = 6$.

B
My Publications

The work in this thesis comprises of the following papers:

Sharma, Yash, Davies, Frederick, Gaikwad, Prakash, Nasir, Fahad, and Bosman, Sarah (Apr. 2025). "Behavior of the Ly Damping Wings as a Function of Reionization Topology". In: *The Astrophysical Journal* 983, p. 118. DOI: 10.3847/1538-4357/adbe6d (cit. on pp. 54, 143).

Sharma, Yash, Davies, Frederick, "Constraining *Lyα* Damping Wings Using Fisher Matrix", submitted to The Astrophysical Journal, and under review.

Sharma, Yash, Davies, Frederick, "Understanding the short LyC MFP at $z = 6$", in preparation.

References

Abel, Tom, Bryan, Greg L, and Norman, Michael L (2000). "The formation and fragmentation of primordial molecular clouds". In: *The Astrophysical Journal* 540.1, p. 39 (cit. on p. 7).

Aghanim, Nabila, Akrami, Yashar, Ashdown, Mark, Aumont, Jonathan, Baccigalupi, Carlo, Ballardini, Mario, Banday, Anthony J, Barreiro, RB, Bartolo, N, Basak, S, et al. (2020). "Planck 2018 results-VI. Cosmological parameters". In: *Astronomy & Astrophysics* 641, A6 (cit. on pp. 5, 7, 56, 80).

Alpher, Ralph A, Bethe, Hans, and Gamow, George (1948). "The origin of chemical elements". In: *Physical Review* 73.7, p. 803 (cit. on p. 6).

Amaro-Seoane, Pau, Andrews, Jeff, Arca Sedda, Manuel, Askar, Abbas, Baghi, Quentin, Balasov, Razvan, Bartos, Imre, Bavera, Simone S, Bellovary, Jillian, Berry, Christopher PL, et al. (2023). "Astrophysics with the laser interferometer space antenna". In: *Living Reviews in Relativity* 26.1, p. 2 (cit. on p. 7).

Bajtlik, Stanislaw, Duncan, Robert C, and Ostriker, Jeremiah P (1988). "Quasar ionization of Lyman-alpha clouds-The proximity effect, a probe of the ultraviolet background at high redshift". In: *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 327, April 15, 1988, p. 570-583.* 327, pp. 570–583 (cit. on p. 110).

Barkana, Rennan and Loeb, Abraham (2001). "In the beginning: the first sources of light and the reionization of the universe". In: *Physics reports* 349.2, pp. 125–238 (cit. on pp. 9, 30, 31, 46, 57).

Barnett, Rhys, Warren, SJ, Mortlock, Daniel John, Cuby, J-G, Conselice, C, Hewett, PC, Willott, CJ, Auricchio, NATALIA, Balaguera-Antolínez, A, Baldi, Marco, et al. (2019). "Euclid preparation-V. Predicted yield of redshift 7< z< 9 quasars from the wide survey". In: *Astronomy & Astrophysics* 631, A85 (cit. on p. 76).

Barry, N, Wilensky, M, Trott, CM, Pindor, B, Beardsley, AP, Hazelton, BJ, Sullivan, IS, Morales, MF, Pober, JC, Line, J, et al. (2019). "Improving the epoch of reionization power spectrum results from Murchison Widefield Array season 1 observations". In: *The Astrophysical Journal* 884.1, p. 1 (cit. on p. 43).

Battaglia, Nick, Trac, Hy, Cen, Renyue, and Loeb, Abraham (2013). "Reionization on large scales. I. A parametric model constructed from radiation-hydrodynamic simulations". In: *The Astrophysical Journal* 776.2, p. 81 (cit. on p. 52).

Becker, George D., D'Aloisio, Anson, Christenson, Holly M., Zhu, Yongda, Worseck, Gábor, and Bolton, James S. (Dec. 2021). "The mean free path of ionizing photons at $5 < z < 6$: evidence for rapid evolution near reionization". In: 508.2, pp. 1853–1869. DOI: 10.1093/mnras/stab2696. arXiv: 2103.16610 [astro-ph.CO] (cit. on pp. 111, 115, 117).

Becker, George D., Rauch, Michael, and Sargent, Wallace L. W. (June 2007). "The Evolution of Optical Depth in the Lyα Forest: Evidence Against Reionization at z$\sim$6". In: 662.1, pp. 72–93. DOI: 10.1086/517866. arXiv: astro-ph/0607633 [astro-ph] (cit. on p. 91).

Becker, Robert H, Fan, Xiaohui, White, Richard L, Strauss, Michael A, Narayanan, Vijay K, Lupton, Robert H, Gunn, James E, Annis, James, Bahcall, Neta A, Brinkmann, J, et al. (2001). "Evidence for Reionization at z 6: Detection of a Gunn-Peterson Trough in az= 6.28 Quasar". In: *The Astronomical Journal* 122.6, p. 2850 (cit. on pp. 47, 107).

Behroozi, Peter S and Silk, Joseph (2015). "A simple technique for predicting high-redshift galaxy evolution". In: *The Astrophysical Journal* 799.1, p. 32 (cit. on p. 61).

Biscoveanu, Sylvia, Talbot, Colm, Thrane, Eric, and Smith, Rory (Dec. 2020). "Measuring the Primordial Gravitational-Wave Background in the Presence of Astrophysical Foregrounds". In: *Phys. Rev. Lett.* 125 (24), p. 241101. DOI: 10.1103/PhysRevLett.125.241101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.241101 (cit. on p. 7).

Bolton, James S., Puchwein, Ewald, Sijacki, Debora, Haehnelt, Martin G., Kim, Tae-Sun, Meiksin, Avery, Regan, John A., and Viel, Matteo (Jan. 2017). "The Sherwood simulation suite: overview and data comparisons with the Lyman α forest at redshifts $2 \leq z \leq 5$". In: 464.1, pp. 897–914. DOI: 10.1093/mnras/stw2397. arXiv: 1605.03462 [astro-ph.CO] (cit. on p. 117).

Bond, J. R., Cole, S., Efstathiou, G., and Kaiser, N. (Oct. 1991). "Excursion Set Mass Functions for Hierarchical Gaussian Fluctuations". In: 379, p. 440. DOI: 10.1086/170520 (cit. on pp. 24, 28, 57, 88).

Bosman, Sarah EI, Davies, Frederick B, Becker, George D, Keating, Laura C, Davies, Rebecca L, Zhu, Yongda, Eilers, Anna-Christina, D'Odorico, Valentina, Bian, Fuyan, Bischetti, Manuela, et al. (2022). "Hydrogen reionization ends by z= 5.3: Lyman-α optical depth measured by the XQR-30 sample". In: *Monthly Notices of the Royal Astronomical Society* 514.1, pp. 55–76 (cit. on pp. 8, 127).

Bowman, Judd D., Rogers, Alan E. E., Monsalve, Raul A., Mozdzen, Thomas J., and Mahesh, Nivedita (2018). "An absorption profile centred at 78 megahertz in the sky-averaged spectrum". In: *Nature* 555.7694, pp. 67–70. DOI: 10.1038/nature25792. arXiv: 1810.05912 [astro-ph.CO] (cit. on p. 43).

Bromm, Volker (2013). "Formation of the first stars". In: *Reports on Progress in Physics* 76.11, p. 112901 (cit. on p. 33).

Bromm, Volker, Coppi, Paolo S, and Larson, Richard B (1999). "Forming the first stars in the universe: the fragmentation of primordial gas". In: *The Astrophysical Journal* 527.1, p. L5 (cit. on p. 7).

Bryan, Greg L. and Norman, Michael L. (Mar. 1998). "Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons". In: 495.1, pp. 80–99. DOI: 10.1086/305262. arXiv: astro-ph/9710107 [astro-ph] (cit. on pp. 24, 30).

Cain, Christopher, D'Aloisio, Anson, Lopez, Garett, Gangolli, Nakul, and Roth, Joshua T. (June 2024). "On the rise and fall of galactic ionizing output at the end of reionization". In: 531.1, pp. 1951–1970. DOI: 10.1093/mnras/stae1223. arXiv: 2311.13638 [astro-ph.CO] (cit. on pp. 127, 131).

Calverley, Alexander P, Becker, George D, Haehnelt, Martin G, and Bolton, James S (2011). "Measurements of the ultraviolet background at 4.6< z< 6.4 using the quasar proximity effect". In: *Monthly Notices of the Royal Astronomical Society* 412.4, pp. 2543–2562 (cit. on pp. 110, 111).

Carilli, CL, Wang, Ran, Fan, X, Walter, F, Kurk, J, Riechers, D, Wagg, J, Hennawi, J, Jiang, L, Menten, KM, et al. (2010). "Ionization near zones associated with quasars at z 6". In: *The Astrophysical Journal* 714.1, p. 834 (cit. on p. 111).

Carroll, Sean M, Press, William H, and Turner, Edwin L (1992). "The cosmological constant". In: *In: Annual review of astronomy and astrophysics. Vol. 30 (A93-25826 09-90), p. 499-542.* 30, pp. 499–542 (cit. on p. 16).

Cen, Renyue and Haiman, Zoltan (2000). "Quasar Strömgren spheres before cosmological reionization". In: *The Astrophysical Journal* 542.2, p. L75 (cit. on pp. 64, 88).

Charlton, Jane and Churchill, Chris (2001). "Quasistellar Objects: Intervening Absorption Lines". In: *Encyclopedia of Astronomy & Astrophysics*. CRC Press, pp. 1–11 (cit. on p. 44).

Choudhury, T Roy and Ferrara, Andrea (2006). "Updating reionization scenarios after recent data". In: *Monthly Notices of the Royal Astronomical Society: Letters* 371.1, pp. L55–L59 (cit. on p. 50).

Choudhury, Tirthankar Roy (2022). "A short introduction to reionization physics". In: *General Relativity and Gravitation* 54.9, p. 102 (cit. on pp. 31, 38, 51).

Choudhury, Tirthankar Roy and Paranjape, Aseem (2018). "Photon number conservation and the large-scale 21 cm power spectrum in seminumerical models of reionization". In: *Monthly Notices of the Royal Astronomical Society* 481.3, pp. 3821–3837 (cit. on p. 52).

Ciardi, Benedetta and Ferrara, Andrea (2005). "The first cosmic structures and their effects". In: *Space Science Reviews* 116.3, pp. 625–705 (cit. on pp. 33, 37).

Collaboration, The LIGO Scientific, Aasi, J, Abbott, B P, Abbott, R, Abbott, T, Abernathy, M R, Ackley, K, Adams, C, Adams, T, Addesso, P, Adhikari, R X, Adya, V, Affeldt, C, Aggarwal, N, Aguiar, O D, Ain, A, Ajith, P, Alemic, A, Allen, B, Amariutei, D, Anderson, S B, Anderson, W G, Arai, K, Araya, M C, Arceneaux, C, Areeda, J S, Ashton, G, Ast, S, Aston, S M, Aufmuth, P, Aulbert, C, Aylott, B E, Babak, S, Baker, P T, Ballmer, S W, Barayoga, J C, Barbet, M, Barclay, S, Barish, B C, Barker, D, Barr, B, Barsotti, L, Bartlett, J, Barton, M A, Bartos, I, Bassiri, R, Batch, J C, Baune, C, Behnke, B, Bell, A S, Bell, C, Benacquista, M, Bergman, J, Bergmann, G, Berry, C P L, Betzwieser, J, Bhagwat, S, Bhandare, R, Bilenko, I A, Billingsley, G, Birch, J, Biscans, S, Biwer, C, Blackburn, J K, Blackburn, L, Blair, C D, Blair, D, Bock, O, Bodiya, T P, Bojtos, P, Bond, C, Bork, R, Born, M, Bose, Sukanta, Brady, P R, Braginsky, V B, Brau, J E, Bridges, D O, Brinkmann, M, Brooks, A F, Brown, D A, Brown, D D, Brown, N M, Buchman, S, Buikema, A, Buonanno, A, Cadonati, L, Calderón Bustillo, J, Camp, J B, Cannon, K C, Cao, J, Capano, C D, Caride, S, Caudill, S, Cavaglià, M, Cepeda, C, Chakraborty, R, Chalermsongsak, T, Chamberlin, S J, Chao, S, Charlton, P, Chen, Y, Cho, H S, Cho, M, Chow, J H, Christensen, N, Chu, Q, Chung, S, Ciani, G, Clara, F, Clark, J A, Collette, C, Cominsky, L, Constancio, M, Cook, D, Corbitt, T R, Cornish, N, Corsi, A, Costa, C A, Coughlin, M W, Countryman, S, Couvares, P, Coward, D M, Cowart, M J, Coyne, D C, Coyne, R, Craig, K, Creighton, J D E, Creighton, T D, Cripe, J, Crowder, S G, Cumming, A, Cunningham, L, Cutler, C, Dahl, K, Dal Canton, T, Damjanic, M, Danilishin, S L, Danzmann, K, Dartez, L, Dave, I, Daveloza, H, Davies, G S, Daw, E J, DeBra, D, Del Pozzo, W, Denker, T, Dent, T, Dergachev, V, DeRosa, R T, DeSalvo, R, Dhurandhar, S, D´ıaz, M, Di Palma, I, Dojcinoski, G, Dominguez, E, Donovan, F, Dooley, K L, Doravari, S, Douglas, R, Downes, T P, Driggers, J C, Du, Z, Dwyer, S, Eberle, T, Edo, T, Edwards, M, Edwards, M, Effler, A, Eggenstein, H.-B, Ehrens, P, Eichholz, J, Eikenberry, S S, Essick, R, Etzel, T, Evans, M, Evans, T, Factourovich, M, Fairhurst, S, Fan, X, Fang, Q, Farr, B, Farr, W M, Favata, M, Fays, M, Fehrmann, H, Fejer, M M, Feldbaum, D, Ferreira, E C, Fisher, R P, Frei, Z, Freise, A, Frey, R, Fricke, T T, Fritschel, P, Frolov, V V, Fuentes-Tapia, S, Fulda, P, Fyffe, M, Gair, J R, Gaonkar, S, Gehrels, N, Gergely´, L Á, Giaime, J A, Giardina, K D, Gleason, J, Goetz, E, Goetz, R, Gondan, L, González, G, Gordon, N, Gorodetsky, M L, Gossan, S, Goßler, S, Gräf, C, Graff, P B, Grant, A, Gras, S, Gray, C, Greenhalgh, R J S, Gretarsson, A M, Grote, H, Grunewald, S, Guido, C J, Guo, X, Gushwa, K, Gustafson, E K, Gustafson, R, Hacker, J, Hall, E D, Hammond, G, Hanke, M, Hanks, J, Hanna, C, Hannam, M D, Hanson, J, Hardwick, T, Harry, G M, Harry, I W, Hart, M, Hartman, M T, Haster, C-J, Haughian, K, Hee, S, Heintze, M, Heinzel, G, Hendry, M, Heng, I S, Heptonstall, A W, Heurs, M, Hewitson, M, Hild, S, Hoak, D, Hodge, K A, Hollitt, S E, Holt, K, Hopkins, P, Hosken, D J, Hough, J, Houston, E, Howell, E J, Hu, Y M, Huerta, E, Hughey, B, Husa, S, Huttner, S H, Huynh, M, Huynh-Dinh, T, Idrisy, A, Indik, N, Ingram, D R, Inta, R, Islas, G, Isler, J C, Isogai, T, Iyer, B R, Izumi, K, Jacobson, M, Jang, H, Jawahar, S, Ji, Y, Jiménez-Forteza, F, Johnson, W W, Jones, D I, Jones, R, Ju, L, Haris, K, Kalogera, V, Kandhasamy, S, Kang, G, Kanner, J B, Katsavounidis, E,

Katzman, W, Kaufer, H, Kaufer, S, Kaur, T, Kawabe, K, Kawazoe, F, Keiser, G M, Keitel, D, Kelley, D B, Kells, W, Keppel, D G, Key, J S, Khalaidovski, A, Khalili, F Y, Khazanov, E A, Kim, C, Kim, K, Kim, N G, Kim, N, Kim, Y.-M, King, E J, King, P J, Kinzel, D L, Kissel, J S, Klimenko, S, Kline, J, Koehlenbeck, S, Kokeyama, K, Kondrashov, V, Korobko, M, Korth, W Z, Kozak, D B, Kringel, V, Krishnan, B, Krueger, C, Kuehn, G, Kumar, A, Kumar, P, Kuo, L, Landry, M, Lantz, B, Larson, S, Lasky, P D, Lazzarini, A, Lazzaro, C, Le, J, Leaci, P, Leavey, S, Lebigot, E O, Lee, C H, Lee, H K, Lee, H M, Leong, J R, Levin, Y, Levine, B, Lewis, J, Li, T G F, Libbrecht, K, Libson, A, Lin, A C, Littenberg, T B, Lockerbie, N A, Lockett, V, Logue, J, Lombardi, A L, Lormand, M, Lough, J, Lubinski, M J, Lück, H, Lundgren, A P, Lynch, R, Ma, Y, Macarthur, J, MacDonald, T, Machenschalk, B, MacInnis, M, Macleod, D M, Magaña-Sandoval, F, Magee, R, Mageswaran, M, Maglione, C, Mailand, K, Mandel, I, Mandic, V, Mangano, V, Mansell, G L, Márka, S, Márka, Z, Markosyan, A, Maros, E, Martin, I W, Martin, R M, Martynov, D, Marx, J N, Mason, K, Massinger, T J, Matichard, F, Matone, L, Mavalvala, N, Mazumder, N, Mazzolo, G, McCarthy, R, McClelland, D E, McCormick, S, McGuire, S C, McIntyre, G, McIver, J, McLin, K, McWilliams, S, Meadors, G D, Meinders, M, Melatos, A, Mendell, G, Mercer, R A, Meshkov, S, Messenger, C, Meyers, P M, Miao, H, Middleton, H, Mikhailov, E E, Miller, A, Miller, J, Millhouse, M, Ming, J, Mirshekari, S, Mishra, C, Mitra, S, Mitrofanov, V P, Mitselmakher, G, Mittleman, R, Moe, B, Mohanty, S D, Mohapatra, S R P, Moore, B, Moraru, D, Moreno, G, Morriss, S R, Mossavi, K, Mow-Lowry, C M, Mueller, C L, Mueller, G, Mukherjee, S, Mullavey, A, Munch, J, Murphy, D, Murray, P G, Mytidis, A, Nash, T, Nayak, R K, Necula, V, Nedkova, K, Newton, G, Nguyen, T, Nielsen, A B, Nissanke, S, Nitz, A H, Nolting, D, Normandin, M E N, Nuttall, L K, Ochsner, E, O'Dell, J, Oelker, E, Ogin, G H, Oh, J J, Oh, S H, Ohme, F, Oppermann, P, Oram, R, O'Reilly, B, Ortega, W, O'Shaughnessy, R, Osthelder, C, Ott, C D, Ottaway, D J, Ottens, R S, Overmier, H, Owen, B J, Padilla, C, Pai, A, Pai, S, Palashov, O, Pal-Singh, A, Pan, H, Pankow, C, Pannarale, F, Pant, B C, Papa, M A, Paris, H, Patrick, Z, Pedraza, M, Pekowsky, L, Pele, A, Penn, S, Perreca, A, Phelps, M, Pierro, V, Pinto, I M, Pitkin, M, Poeld, J, Post, A, Poteomkin, A, Powell, J, Prasad, J, Predoi, V, Premachandra, S, Prestegard, T, Price, L R, Principe, M, Privitera, S, Prix, R, Prokhorov, L, Puncken, O, Pürrer, M, Qin, J, Quetschke, V, Quintero, E, Quiroga, G, Quitzow-James, R, Raab, F J, Rabeling, D S, Radkins, H, Raffai, P, Raja, S, Rajalakshmi, G, Rakhmanov, M, Ramirez, K, Raymond, V, Reed, C M, Reid, S, Reitze, D H, Reula, O, Riles, K, Robertson, N A, Robie, R, Rollins, J G, Roma, V, Romano, J D, Romanov, G, Romie, J H, Rowan, S, Rüdiger, A, Ryan, K, Sachdev, S, Sadecki, T, Sadeghian, L, Saleem, M, Salemi, F, Sammut, L, Sandberg, V, Sanders, J R, Sannibale, V, Santiago-Prieto, I, Sathyaprakash, B S, Saulson, P R, Savage, R, Sawadsky, A, Scheuer, J, Schilling, R, Schmidt, P, Schnabel, R, Schofield, R M S, Schreiber, E, Schuette, D, Schutz, B F, Scott, J, Scott, S M, Sellers, D, Sengupta, A S, Sergeev, A, Serna, G, Sevigny, A, Shaddock, D A, Shahriar, M S, Shaltev, M, Shao, Z, Shapiro, B, Shawhan, P, Shoemaker, D H, Sidery, T L, Siemens, X, Sigg, D, Silva, A D, Simakov, D, Singer, A, Singer, L, Singh, R, Sintes, A M, Slagmolen, B J J, Smith, J R, Smith, M R, Smith, R J E, Smith-Lefebvre, N D, Son, E J, Sorazu, B, Souradeep, T, Staley, A, Stebbins, J, Steinke, M, Steinlechner, J, Steinlechner, S, Steinmeyer, D, Stephens, B C, Steplewski, S, Stevenson, S, Stone, R, Strain, K A, Strigin, S, Sturani, R, Stuver, A L, Summerscales, T Z, Sutton, P J, Szczepanczyk, M, Szeifert, G, Talukder, D, Tanner, D B, Tápai, M, Tarabrin, S P, Taracchini, A, Taylor, R, Tellez, G, Theeg, T, Thirugnanasambandam, M P, Thomas, M, Thomas, P, Thorne, K A, Thorne, K S, Thrane, E, Tiwari, V, Tomlinson, C, Torres, C V,

Torrie, C I, Traylor, G, Tse, M, Tshilumba, D, Ugolini, D, Unnikrishnan, C S, Urban, A L, Usman, S A, Vahlbruch, H, Vajente, G, Valdes, G, Vallisneri, M, Veggel, A A van, Vass, S, Vaulin, R, Vecchio, A, Veitch, J, Veitch, P J, Venkateswara, K, Vincent-Finley, R, Vitale, S, Vo, T, Vorvick, C, Vousden, W D, Vyatchanin, S P, Wade, A R, Wade, L, Wade, M, Walker, M, Wallace, L, Walsh, S, Wang, H, Wang, M, Wang, X, Ward, R L, Warner, J, Was, M, Weaver, B, Weinert, M, Weinstein, A J, Weiss, R, Welborn, T, Wen, L, Wessels, P, Westphal, T, Wette, K, Whelan, J T, Whitcomb, S E, White, D J, Whiting, B F, Wilkinson, C, Williams, L, Williams, R, Williamson, A R, Willis, J L, Willke, B, Wimmer, M, Winkler, W, Wipf, C C, Wittel, H, Woan, G, Worden, J, Xie, S, Yablon, J, Yakushin, I, Yam, W, Yamamoto, H, Yancey, C C, Yang, Q, Zanolin, M, Zhang, Fan, Zhang, L, Zhang, M, Zhang, Y, Zhao, C, Zhou, M, Zhu, X J, Zucker, M E, Zuraw, S, and Zweizig, J (Mar. 2015). "Advanced LIGO". In: *Classical and Quantum Gravity* 32.7, p. 074001. DOI: 10.1088/0264-9381/32/7/074001. URL: https://dx.doi.org/10.1088/0264-9381/32/7/074001 (cit. on p. 7).

Davies, Frederick B., Hennawi, Joseph F., Bañados, Eduardo, Lukić, Zarija, Decarli, Roberto, Fan, Xiaohui, Farina, Emanuele P., Mazzucchelli, Chiara, Rix, Hans-Walter, Venemans, Bram P., Walter, Fabian, Wang, Feige, and Yang, Jinyi (Sept. 2018). "Quantitative Constraints on the Reionization History from the IGM Damping Wing Signature in Two Quasars at z > 7". In: 864.2, 142, p. 142. DOI: 10.3847/1538-4357/aad6dc. arXiv: 1802.06066 [astro-ph.CO] (cit. on pp. 55, 71, 90, 135).

Davies, Frederick B., Hennawi, Joseph F., and Eilers, Anna-Christina (Mar. 2020). "Time-dependent behaviour of quasar proximity zones at z ∼ 6". In: 493.1, pp. 1330–1343. DOI: 10.1093/mnras/stz3303. arXiv: 1903.12346 [astro-ph.CO] (cit. on pp. 91, 102, 131).

Dayal, Pratika and Ferrara, Andrea (2018). "Early galaxy formation and its large-scale effects". In: *Physics Reports* 780-782. Early galaxy formation and its large-scale effects, pp. 1–64. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2018.10.002. URL: https://www.sciencedirect.com/science/article/pii/S0370157318302266 (cit. on pp. 33, 37).

Dayal, Pratika, Ferrara, Andrea, Dunlop, James S, and Pacucci, Fabio (2014). "Essential physics of early galaxy formation". In: *Monthly Notices of the Royal Astronomical Society* 445.3, pp. 2545–2557 (cit. on p. 61).

DeBoer, David R., Parsons, Aaron R., Aguirre, James E., Alexander, Paul, Ali, Zaki S., Beardsley, Adam P., Bernardi, Gianni, Bowman, Judd D., Bradley, Richard F., Carilli, Chris L., Cheng, Carina, de Lera Acedo, Eloy, Dillon, Joshua S., Ewall-Wice, Aaron, Fadana, Gcobisa, Fagnoni, Nicolas, Fritz, Randall, Furlanetto, Steve R., Glendenning, Brian, Greig, Bradley, Grobbelaar, Jasper, Hazelton, Bryna J., Hewitt, Jacqueline N., Hickish, Jack, Jacobs, Daniel C., Julius, Austin, Kariseb, MacCalvin, Kohn, Saul A., Lekalake, Telalo, Liu, Adrian, Loots, Anita, MacMahon, David, Malan, Lourence, Malgas, Cresshim, Maree, Matthys, Martinot, Zachary, Mathison, Nathan, Matsetela, Eunice, Mesinger, Andrei, Morales, Miguel F., Neben, Abraham R., Patra, Nipanjana, Pieterse, Samantha, Pober, Jonathan C., Razavi-Ghods, Nima, Ringuette, Jon, Robnett, James, Rosie, Kathryn, Sell, Raddwine, Smith, Craig, Syce, Angelo, Tegmark, Max, Thyagarajan, Nithyanandan, Williams, Peter K. G., and Zheng, Haoxuan (Apr. 2017). "Hydrogen Epoch of Reionization Array (HERA)". In: 129.974, p. 045001. DOI: 10.1088/1538-3873/129/974/045001. arXiv: 1606.07473 [astro-ph.IM] (cit. on p. 43).

Doroshkevich, AG and Zel'dovich, Ya B (1964). "The Development of Perturbations of Arbitrary From in a Homogeneous Medium at Low Pressure". In: *Soviet Astronomy, Vol. 7, p. 615* 7, p. 615 (cit. on p. 18).

Ďurovčíková, Dominika, Eilers, Anna-Christina, Chen, Huanqing, Satyavolu, Sindhu, Kulkarni, Girish, Simcoe, Robert A, Keating, Laura C, Haehnelt, Martin G, and Bañados, Eduardo (2024). "Chronicling the Reionization History at 6 z 7 with Emergent Quasar Damping Wings". In: *The Astrophysical Journal* 969.2, p. 162 (cit. on pp. 69, 90).

Eilers, Anna-Christina, Hennawi, Joseph F, Davies, Frederick B, and Simcoe, Robert A (2021). "Detecting and Characterizing Young Quasars. II. Four Quasars at z 6 with Lifetimes< 104 Yr". In: *The Astrophysical Journal* 917.1, p. 38 (cit. on pp. 67, 89).

Eilers, Anna-Christina, Mackenzie, Ruari, Pizzati, Elia, Matthee, Jorryt, Hennawi, Joseph F, Zhang, Haowen, Bordoloi, Rongmon, Kashino, Daichi, Lilly, Simon J, Naidu, Rohan P, et al. (2024). "EIGER VI. The Correlation Function, Host Halo Mass and Duty Cycle of Luminous Quasars at z\gtrsim6". In: *arXiv preprint arXiv:2403.07986* (cit. on p. 64).

Fan, Xiaohui, Strauss, Michael A, Becker, Robert H, White, Richard L, Gunn, James E, Knapp, Gillian R, Richards, Gordon T, Schneider, Donald P, Brinkmann, J, and Fukugita, Masataka (2006). "Constraining the evolution of the ionizing background and the epoch of reionization with z 6 quasars. II. A sample of 19 quasars". In: *The Astronomical Journal* 132.1, p. 117 (cit. on p. 111).

Fisher, Ronald Aylmer (1925). "Theory of statistical estimation". In: *Mathematical proceedings of the Cambridge philosophical society*. Vol. 22. 5. Cambridge University Press, pp. 700–725 (cit. on p. 81).

Furlanetto, Steven R and Oh, S Peng (2005). "Taxing the rich: recombinations and bubble growth during reionization". In: *Monthly Notices of the Royal Astronomical Society* 363.3, pp. 1031–1048 (cit. on pp. 116, 118).

Furlanetto, Steven R, Zaldarriaga, Matias, and Hernquist, Lars (2004). "The growth of H II regions during reionization". In: *The Astrophysical Journal* 613.1, p. 1 (cit. on p. 56).

Gaikwad, Prakash, Haehnelt, Martin G., Davies, Fredrick B., Bosman, Sarah E. I., Molaro, Margherita, Kulkarni, Girish, D'Odorico, Valentina, Becker, George D., Davies, Rebecca L., Nasir, Fahad, Bolton, James S., Keating, Laura C., Iršič, Vid, Puchwein, Ewald, Zhu, Yongda, Asthana, Shikhar, Yang, Jinyi, Lai, Samuel, and Eilers, Anna-Christina (Nov. 2023). "Measuring the photoionization rate, neutral fraction, and mean free path of H I ionizing photons at $4.9 \leq z \leq 6.0$ from a large sample of XShooter and ESI spectra". In: 525.3, pp. 4093–4120. DOI: 10.1093/mnras/stad2566. arXiv: 2304.02038 [astro-ph.CO] (cit. on pp. 118, 127, 131).

Garaldi, Enrico, Kannan, Rahul, Smith, Aaron, Springel, Volker, Pakmor, Rüdiger, Vogelsberger, Mark, and Hernquist, Lars (2022). "The THESAN project: properties of the intergalactic medium and its connection to reionization-era galaxies". In: *Monthly Notices of the Royal Astronomical Society* 512.4, pp. 4909–4933 (cit. on p. 51).

Gehlot, BK, Mertens, FG, Koopmans, LVE, Brentjens, MA, Zaroubi, S, Ciardi, B, Ghosh, A, Hatef, M, Iliev, Ilian T, Jelić, Vibor, et al. (2019). "The first power spectrum limit on the 21-cm signal of neutral hydrogen during the Cosmic Dawn at z= 20–25 from LOFAR". In: *Monthly Notices of the Royal Astronomical Society* 488.3, pp. 4271–4287 (cit. on p. 43).

Geil, Paul M, Mutch, Simon J, Poole, Gregory B, Angel, Paul W, Duffy, Alan R, Mesinger, Andrei, and Wyithe, J Stuart B (2016). "Dark-ages reionization and galaxy formation simulation V: morphology and statistical signatures of reionization". In: *Monthly Notices of the Royal Astronomical Society* 462.1, pp. 804–817 (cit. on p. 52).

Ghara, Raghunath, Choudhury, T Roy, and Datta, Kanan K (2015). "21 cm signal from cosmic dawn: imprints of spin temperature fluctuations and peculiar velocities". In: *Monthly Notices of the Royal Astronomical Society* 447.2, pp. 1806–1825 (cit. on p. 51).

Giroux, Mark L., Sutherland, Ralph S., and Shull, J. Michael (Nov. 1994). "Multiphase Gas in Quasar Absorption-Line Systems". In: 435, p. L97. DOI: 10.1086/187603 (cit. on p. 57).

Greig, Bradley and Mesinger, Andrei (2017). "Simultaneously constraining the astrophysics of reionization and the epoch of heating with 21CMMC". In: *Monthly Notices of the Royal Astronomical Society* 472.3, pp. 2651–2669 (cit. on pp. 62, 88).

— (2018). "21CMMC with a 3D light-cone: the impact of the co-evolution approximation on the astrophysics of reionization and cosmic dawn". In: *Monthly Notices of the Royal Astronomical Society* 477.3, pp. 3217–3229 (cit. on pp. 56, 88).

Greig, Bradley, Mesinger, Andrei, and Bañados, Eduardo (Apr. 2019). "Constraints on reionization from the z = 7.5 QSO ULASJ1342+0928". In: 484.4, pp. 5094–5101. DOI: 10.1093/mnras/stz230. arXiv: 1807.01593 [astro-ph.CO] (cit. on p. 55).

Greig, Bradley, Mesinger, Andrei, Davies, Frederick B., Wang, Feige, Yang, Jinyi, and Hennawi, Joseph F. (June 2022). "IGM damping wing constraints on reionization from covariance reconstruction of two $z \gtrsim 7$ QSOs". In: 512.4, pp. 5390–5403. DOI: 10.1093/mnras/stac825. arXiv: 2112.04091 [astro-ph.CO] (cit. on p. 55).

Greig, Bradley, Mesinger, Andrei, Haiman, Zoltan, and Simcoe, Robert A (2017). "Are we witnessing the epoch of reionization at z= 7.1 from the spectrum of J1120+ 0641?" In: *Monthly Notices of the Royal Astronomical Society* 466.4, pp. 4239–4249 (cit. on p. 55).

Gunn, James E and Peterson, Bruce A (1965). "On the Density of Neutral Hydrogen in Intergalactic Space." In: *Astrophysical Journal, vol. 142, p. 1633-1636* 142, pp. 1633–1636 (cit. on pp. 46, 55, 65).

Guth, Alan H (1981). "Inflationary universe: A possible solution to the horizon and flatness problems". In: *Physical Review D* 23.2, p. 347 (cit. on p. 5).

Hazard, C., Mackey, M. B., and Shimmins, A. J. (Mar. 1963). "Investigation of the Radio Source 3C 273 By The Method of Lunar Occultations". In: *Nature* 197, pp. 1037–1039 (cit. on p. 43).

HERA Collaboration, Abdurashidova, Zara, Adams, Tyrone, Aguirre, James E., Alexander, Paul, Ali, Zaki S., Baartman, Rushelle, Balfour, Yanga, Barkana, Rennan, Beardsley, Adam P., Bernardi, Gianni, Billings, Tashalee S., Bowman, Judd D., Bradley, Richard F., Breitman, Daniela, Bull, Philip, Burba, Jacob, Carey, Steve, Carilli, Chris L., Cheng, Carina, Choudhuri, Samir, DeBoer, David R., de Lera Acedo, Eloy, Dexter, Matt, Dillon, Joshua S., Ely, John, Ewall-Wice, Aaron, Fagnoni, Nicolas, Fialkov, Anastasia, Fritz, Randall, Furlanetto, Steven R., Gale-Sides, Kingsley, Garsden, Hugh, Glendenning, Brian, Gorce, Adélie, Gorthi, Deepthi, Greig, Bradley, Grobbelaar, Jasper, Halday, Ziyaad, Hazelton, Bryna J., Heimersheim, Stefan, Hewitt, Jacqueline N., Hickish, Jack, Jacobs, Daniel C., Julius, Austin, Kern, Nicholas S., Kerrigan, Joshua, Kittiwisit, Piyanat, Kohn, Saul A., Kolopanis, Matthew, Lanman, Adam, La Plante, Paul, Lewis, David, Liu, Adrian, Loots, Anita, Ma, Yin-Zhe, MacMahon, David H. E., Malan, Lourence, Malgas, Keith, Malgas, Cresshim, Maree, Matthys, Marero, Bradley, Martinot, Zachary E., McBride, Lisa, Mesinger, Andrei, Mirocha, Jordan, Molewa, Mathakane, Morales, Miguel F., Mosiane, Tshegofalang, Muñoz, Julian B., Murray, Steven G., Nagpal, Vighnesh, Neben, Abraham R., Nikolic, Bojan, Nunhokee, Chuneeta D., Nuwegeld, Hans, Parsons, Aaron R., Pascua, Robert, Patra, Nipanjana, Pieterse, Samantha, Qin, Yuxiang, Razavi-Ghods, Nima, Robnett, James, Rosie, Kathryn, Santos, Mario G., Sims, Peter, Singh, Saurabh, Smith, Craig, Swarts, Hilton, Tan, Jianrong, Thyagarajan, Nithyanandan, Wilensky, Michael J., Williams, Peter K. G., van Wyngaarden, Pieter, and Zheng, Haoxuan (Mar. 2023). "Improved Constraints on the 21 cm EoR Power Spectrum and the X-Ray Heating of the IGM with HERA Phase I Observations". In: 945.2, 124, p. 124. DOI: 10.3847/1538-4357/acaf50 (cit. on p. 43).

Hui, Lam and Gnedin, Nickolay Y (1997). "Equation of state of the photoionized intergalactic medium". In: *Monthly Notices of the Royal Astronomical Society* 292.1, pp. 27–42 (cit. on pp. 46, 57).

Hutter, Anne, Dayal, Pratika, Yepes, Gustavo, Gottlöber, Stefan, Legrand, Laurent, and Ucci, Graziano (2021). "Astraeus I: the interplay between galaxy formation and reionization". In: *Monthly Notices of the Royal Astronomical Society* 503.3, pp. 3698–3723 (cit. on p. 52).

Iliev, Ilian T, Mellema, Garrelt, Pen, U-L, Merz, Hugh, Shapiro, Paul R, and Alvarez, Marcelo A (2006). "Simulating cosmic reionization at large scales–I. The geometry of reionization". In: *Monthly Notices of the Royal Astronomical Society* 369.4, pp. 1625–1638 (cit. on p. 51).

Kannan, Rahul, Garaldi, E, Smith, A ea, Pakmor, R, Springel, V, Vogelsberger, M, and Hernquist, L (2022). "Introducing the thesan project: radiation-magnetohydrodynamic simulations of the epoch of reionization". In: *Monthly Notices of the Royal Astronomical Society* 511.3, pp. 4005–4030 (cit. on p. 51).

Katz, Harley, Kimm, Taysun, Haehnelt, Martin G, Sijacki, Debora, Rosdahl, Joakim, and Blaizot, Jeremy (2019). "Tracing the sources of reionization in cosmological radiation hydrodynamics simulations". In: *Monthly Notices of the Royal Astronomical Society* 483.1, pp. 1029–1041 (cit. on p. 51).

Keating, Laura C, Bolton, James S, Cullen, Fergus, Haehnelt, Martin G, Puchwein, Ewald, and Kulkarni, Girish (2024). "JWST observations of galaxy damping wings during reionization interpreted with cosmological simulations". In: *Monthly Notices of the Royal Astronomical Society*, stae1530 (cit. on pp. 64, 70).

Khrykin, Ilya S., Hennawi, Joseph F., Worseck, Gábor, and Davies, Frederick B. (July 2021). "The first measurement of the quasar lifetime distribution". In: 505.1, pp. 649–662. DOI: 10.1093/mnras/stab1288. arXiv: 2102.04477 [astro-ph.GA] (cit. on p. 89).

Kirkman, David and Tytler, David (2008). "The transverse proximity effect in the z 2 Lyman α forest suggests quasi-stellar object episodic lifetimes of 1 Myr". In: *Monthly Notices of the Royal Astronomical Society* 391.3, pp. 1457–1471 (cit. on p. 108).

Kist, Timo, Hennawi, Joseph F, and Davies, Frederick B (2024). "Quantifying the Precision of IGM Damping Wing Measurements Towards Quasars". In: *arXiv preprint arXiv:2406.12071* (cit. on pp. 50, 55, 65, 79, 90, 102, 130).

Klessen, Ralf S and Glover, Simon CO (2023). "The first stars: formation, properties, and impact". In: *Annual Review of Astronomy and Astrophysics* 61, pp. 65–130 (cit. on p. 7).

Kuhlen, Michael and Faucher-Giguère, Claude-André (2012). "Concordance models of reionization: implications for faint galaxies and escape fraction evolution". In: *Monthly Notices of the Royal Astronomical Society* 423.1, pp. 862–876 (cit. on p. 61).

Kušmić, Samir, Finlator, Kristian, Keating, Laura, and Huscher, Ezra (May 2022). "Assuming Ionization Equilibrium and the Impact on the Ly Forest Power Spectrum during the End of Reionization at 8 z 5". In: *The Astrophysical Journal* 931.1, p. 46. DOI: 10.3847/1538-4357/ac66e3. URL: https://dx.doi.org/10.3847/1538-4357/ac66e3 (cit. on p. 39).

Lacey, Cedric and Cole, Shaun (1993). "Merger rates in hierarchical models of galaxy formation". In: *Monthly Notices of the Royal Astronomical Society* 262.3, pp. 627–649 (cit. on pp. 28, 57, 88).

Larson, Rebecca L, Finkelstein, Steven L, Kocevski, Dale D, Hutchison, Taylor A, Trump, Jonathan R, Haro, Pablo Arrabal, Bromm, Volker, Cleri, Nikko J, Dickinson, Mark, Fujimoto, Seiji, et al. (2023). "A CEERS discovery of an accreting supermassive black hole 570 Myr after the Big Bang: identifying a progenitor of massive z> 6 quasars". In: *The Astrophysical Journal Letters* 953.2, p. L29 (cit. on p. 8).

Lusso, E., Worseck, G., Hennawi, J. F., Prochaska, J. X., Vignali, C., Stern, J., and O'Meara, J. M. (June 2015). "The first ultraviolet quasar-stacked spectrum at z 2.4 from WFC3". In: 449.4, pp. 4204–4220. DOI: 10.1093/mnras/stv516. arXiv: 1503.02075 [astro-ph.GA] (cit. on pp. 91, 111).

Ly, Alexander, Marsman, Maarten, Verhagen, Josine, Grasman, Raoul, and Wagenmakers, Eric-Jan (2017). *A Tutorial on Fisher Information*. arXiv: 1705.01064 [math.ST]. URL: https://arxiv.org/abs/1705.01064 (cit. on pp. 81, 84).

Lynden-Bell, D. (1967). "Statistical Mechanics of Violent Relaxation in Stellar Systems". In: *Monthly Notices of the Royal Astronomical Society* 136, pp. 101–121. DOI: 10.1093/mnras/136.1.101 (cit. on p. 23).

Madau, Piero and Dickinson, Mark (2014). "Cosmic star-formation history". In: *Annual Review of Astronomy and Astrophysics* 52, pp. 415–486 (cit. on p. 8).

Mason, Charlotte A, Treu, Tommaso, Dijkstra, Mark, Mesinger, Andrei, Trenti, Michele, Pentericci, Laura, De Barros, Stephane, and Vanzella, Eros (2018). "The universe is reionizing at z 7: bayesian inference of the IGM neutral fraction using Lyα emission from galaxies". In: *The Astrophysical Journal* 856.1, p. 2 (cit. on pp. 55, 97).

Mason, Charlotte A., Muñoz, Julian B., Greig, Bradley, Mesinger, Andrei, and Park, Jaehong (Sept. 2023). "21CMFISH: Fisher-matrix framework for fast parameter forecasts from the cosmic 21-cm signal". In: 524.3, pp. 4711–4728. DOI: 10.1093/mnras/stad2145. arXiv: 2212.09797 [astro-ph.CO] (cit. on pp. 102, 103, 130).

McQuinn, Matthew, Oh, S. Peng, and Faucher-Giguère, Claude-André (Dec. 2011). "On Lyman-limit Systems and the Evolution of the Intergalactic Ionizing Background". In: 743.1, 82, p. 82. DOI: 10.1088/0004-637X/743/1/82. arXiv: 1101.1964 [astro-ph.CO] (cit. on pp. 116, 118).

Mellema, Garrelt, Iliev, Ilian T, Pen, Ue-Li, and Shapiro, Paul R (2006). "Simulating cosmic reionization at large scales–II. The 21-cm emission features and statistical signals". In: *Monthly notices of the royal astronomical society* 372.2, pp. 679–692 (cit. on p. 51).

Mertens, Florent G, Mevius, Maaijke, Koopmans, Leon VE, Offringa, AR, Mellema, Garrelt, Zaroubi, Saleem, Brentjens, MA, Gan, H, Gehlot, Bharat Kumar, Pandey, VN, et al. (2020). "Improved upper limits on the 21 cm signal power spectrum of neutral hydrogen at z 9.1 from LOFAR". In: *Monthly Notices of the Royal Astronomical Society* 493.2, pp. 1662–1685 (cit. on p. 43).

Mesinger, Andrei and Dijkstra, Mark (2008). "Ultraviolet radiative feedback during the advanced stages of reionization". In: *Monthly Notices of the Royal Astronomical Society* 390.3, pp. 1071–1080 (cit. on p. 57).

Mesinger, Andrei and Furlanetto, Steven (2007). "Efficient simulations of early structure formation and reionization". In: *The Astrophysical Journal* 669.2, p. 663 (cit. on pp. 63, 89).

Mesinger, Andrei, Furlanetto, Steven, and Cen, Renyue (2011). "21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal". In: *Monthly Notices of the Royal Astronomical Society* 411.2, pp. 955–972 (cit. on pp. 52, 56, 59, 91).

Mesinger, Andrei and Furlanetto, Steven R (2008). "Lyα damping wing constraints on inhomogeneous reionization". In: *Monthly Notices of the Royal Astronomical Society* 385.3, pp. 1348–1358 (cit. on pp. 52, 65).

Miralda-Escudé, Jordi (1998). "Reionization of the intergalactic medium and the damping wing of the Gunn-Peterson trough". In: *The Astrophysical Journal* 501.1, p. 15 (cit. on pp. 48, 55).

Mitra, Sourav, Choudhury, T Roy, and Ferrara, Andrea (2011). "Reionization constraints using principal component analysis". In: *Monthly Notices of the Royal Astronomical Society* 413.3, pp. 1569–1580 (cit. on p. 51).

Mitra, Sourav, Choudhury, T Roy, and Ferrara, Andrea (2012). "Joint quasar–cosmic microwave background constraints on reionization history". In: *Monthly Notices of the Royal Astronomical Society* 419.2, pp. 1480–1488 (cit. on p. 51).

— (2015). "Cosmic reionization after Planck". In: *Monthly Notices of the Royal Astronomical Society: Letters* 454.1, pp. L76–L80 (cit. on pp. 51, 61).

Mo, Houjun, Van den Bosch, Frank, and White, Simon (2010). *Galaxy formation and evolution*. Cambridge University Press (cit. on pp. 9, 24, 27, 33, 38).

Morey, Karna A., Eilers, Anna-Christina, Davies, Frederick B., Hennawi, Joseph F., and Simcoe, Robert A. (Nov. 2021). "Estimating the Effective Lifetime of the z 6 Quasar Population from the Composite Proximity Zone Profile". In: 921.1, 88, p. 88. DOI: 10.3847/1538-4357/ac1c70. arXiv: 2108.10907 (cit. on pp. 67, 89).

Mortlock, Daniel (2016). "Quasars as probes of cosmological reionization". In: *Understanding the Epoch of Cosmic Reionization: Challenges and Progress*, pp. 187–226 (cit. on pp. 50, 65).

Mortlock, Daniel J, Warren, Stephen J, Venemans, Bram P, Patel, Mitesh, Hewett, Paul C, McMahon, Richard G, Simpson, Chris, Theuns, Tom, Gonzáles-Solares, Eduardo A, Adamson, Andy, et al. (2011). "A luminous quasar at a redshift of z= 7.085". In: *Nature* 474.7353, pp. 616–619 (cit. on pp. 55, 64, 88).

Muñoz, Julian B., Qin, Yuxiang, Mesinger, Andrei, Murray, Steven G., Greig, Bradley, and Mason, Charlotte (Apr. 2022). "The impact of the first galaxies on cosmic dawn and reionization". In: 511.3, pp. 3657–3681. DOI: 10.1093/mnras/stac185. arXiv: 2110.13919 [astro-ph.CO] (cit. on pp. 56, 92).

Murray, Steven G., Greig, Bradley, Mesinger, Andrei, Muñoz, Julian B., Qin, Yuxiang, Park, Jaehong, and Watkinson, Catherine A. (2020). "21cmFAST v3: A Python-integrated C code for generating 3D realizations of the cosmic 21cm signal." In: *Journal of Open Source Software* 5.54, p. 2582. DOI: 10.21105/joss.02582. URL: https://doi.org/10.21105/joss.02582 (cit. on pp. 56, 91).

Mutch, Simon J, Geil, Paul M, Poole, Gregory B, Angel, Paul W, Duffy, Alan R, Mesinger, Andrei, and Wyithe, J Stuart B (2016). "Dark-ages reionization and galaxy formation simulation–III. Modelling galaxy formation and the epoch of reionization". In: *Monthly Notices of the Royal Astronomical Society* 462.1, pp. 250–276 (cit. on pp. 52, 61).

Narlikar, Jayant Vishnu (2002). *An introduction to cosmology*. Cambridge University Press (cit. on pp. 9, 38).

Nyquist, Harry (1928). "Certain Topics in Telegraph Transmission Theory". In: *Transactions of the American Institute of Electrical Engineers* 47.2, pp. 617–644. DOI: 10.1109/T-AIEE.1928.5055024 (cit. on p. 59).

Ocvirk, Pierre, Aubert, Dominique, Sorce, Jenny G, Shapiro, Paul R, Deparis, Nicolas, Dawoodbhoy, Taha, Lewis, Joseph, Teyssier, Romain, Yepes, Gustavo, Gottlöber, Stefan, et al. (2020). "Cosmic Dawn II (CoDa II): a new radiation-hydrodynamics simulation of the self-consistent coupling of galaxy formation and reionization". In: *Monthly Notices of the Royal Astronomical Society* 496.4, pp. 4087–4107 (cit. on p. 51).

Ocvirk, Pierre, Gillet, Nicolas, Shapiro, Paul R, Aubert, Dominique, Iliev, Ilian T, Teyssier, Romain, Yepes, Gustavo, Choi, Jun-Hwan, Sullivan, David, Knebe, Alexander, et al. (2016). "Cosmic Dawn (CoDa): the first radiation-hydrodynamics simulation of reionization and galaxy formation in the Local Universe". In: *Monthly Notices of the Royal Astronomical Society* 463.2, pp. 1462–1485 (cit. on p. 51).

Okamoto, Takashi, Gao, Liang, and Theuns, Tom (2008). "Mass loss of galaxies due to an ultraviolet background". In: *Monthly Notices of the Royal Astronomical Society* 390.3, pp. 920–928 (cit. on p. 57).

Paciga, Gregory, Albert, Joshua, Bandura, Kevin, Chang, Tzu-Ching, Gupta, Yashwant, Hirata, Christopher, Odegova, Julia, Pen, Ue-Li, Peterson, Jeffrey B, and Roy, Jayanta (2013). "A refined foreground-corrected limit on the HI power spectrum at z= 8.6 from the GMRT Epoch of Reionization Experiment". In: *Mon. Not. Roy. Astron. Soc.* 433.arXiv: 1301.5906, p. 639 (cit. on p. 43).

Padmanabhan, Thanu (2000). *Theoretical astrophysics: volume 3, galaxies and cosmology*. Vol. 3. Cambridge University Press (cit. on pp. 9, 32, 38).

Park, Jaehong, Mesinger, Andrei, Greig, Bradley, and Gillet, Nicolas (2019). "Inferring the astrophysics of reionization and cosmic dawn from galaxy luminosity functions and the 21-cm signal". In: *Monthly Notices of the Royal Astronomical Society* 484.1, pp. 933–949 (cit. on pp. 56, 57, 61, 62, 66, 89, 92, 97, 102, 103, 130).

Peacock, John A (1998). *Cosmological physics*. Cambridge university press (cit. on pp. 9, 26, 38).

Peebles, P. J. E. (1993). *Principles of Physical Cosmology*. Princeton, NJ: Princeton University Press. ISBN: 9780691019338 (cit. on pp. 2, 9, 48).

Penzias, Arno A and Wilson, Robert W (1979). "A measurement of excess antenna temperature at 4080 MHz". In: *A Source Book in Astronomy and Astrophysics, 1900–1975*. Harvard University Press, pp. 873–876 (cit. on p. 7).

Pizzati, Elia, Hennawi, Joseph F, Schaye, Joop, Schaller, Matthieu, Eilers, Anna-Christina, Wang, Feige, Frenk, Carlos S, Elbers, Willem, Helly, John C, Mackenzie, Ruari, et al. (2024). "A unified model for the clustering of quasars and galaxies at z\approx6". In: *arXiv preprint arXiv:2403.12140* (cit. on p. 64).

Press, William H. and Schechter, Paul (Feb. 1974). "Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation". In: 187, pp. 425–438. DOI: 10.1086/152650 (cit. on pp. 24, 28, 57, 88).

Pritchard, Jonathan R, Loeb, Abraham, and Wyithe, J Stuart B (2010). "Constraining reionization using 21-cm observations in combination with CMB and Lyα forest data". In: *Monthly Notices of the Royal Astronomical Society* 408.1, pp. 57–70 (cit. on p. 51).

Prochaska, J. Xavier, O'Meara, John M., and Worseck, Gabor (July 2010). "A Definitive Survey for Lyman Limit Systems at z ∼3.5 with the Sloan Digital Sky Survey". In: 718.1, pp. 392–416. DOI: 10.1088/0004-637X/718/1/392. arXiv: 0912.0292 [astro-ph.CO] (cit. on p. 107).

Prochaska, J. Xavier, Worseck, Gabor, and O'Meara, John M. (Nov. 2009). "A Direct Measurement of the Intergalactic Medium Opacity to H I Ionizing Photons". In: 705.2, pp. L113–L117. DOI: 10.1088/0004-637X/705/2/L113. arXiv: 0910.0009 [astro-ph.CO] (cit. on p. 109).

Puchwein, Ewald, Bolton, James S, Keating, Laura C, Molaro, Margherita, Gaikwad, Prakash, Kulkarni, Girish, Haehnelt, Martin G, Iršič, Vid, Šoltinskỳ, Tomáš, Viel, Matteo, et al. (2023). "The Sherwood–Relics simulations: overview and impact of patchy reionization and pressure smoothing on the intergalactic medium". In: *Monthly Notices of the Royal Astronomical Society* 519.4, pp. 6162–6183 (cit. on p. 51).

Qin, Yuxiang, Mesinger, Andrei, Park, Jaehong, Greig, Bradley, and Muñoz, Julian B (2020). "A tale of two sites–I. Inferring the properties of minihalo-hosted galaxies from current observations". In: *Monthly Notices of the Royal Astronomical Society* 495.1, pp. 123–140 (cit. on pp. 56, 92).

Qin, Yuxiang, Mesinger, Andrei, Prelogović, David, Becker, George, Bischetti, Manuela, Bosman, Sarah, Davies, Frederick, D'Odorico, Valentina, Gaikwad, Prakash, Haehnelt, Martin, Keating, Laura, Lai, Samuel, Ryan-Weber, Emma, Satyavolu, Sindhu, Walter, Fabian, and Zhu, Yongda (Apr. 2025). "Percent-level timing of reionisation: Self-consistent, implicit-likelihood inference from XQR-30+ Lyα forest data". In: 42, e049,

e049. DOI: 10.1017/pasa.2025.35. arXiv: 2412.00799 [astro-ph.CO] (cit. on pp. 116, 127, 131).

Rauch, Michael (1998). "The Lyman alpha forest in the spectra of quasistellar objects". In: *Annual Review of Astronomy and Astrophysics* 36.1, pp. 267–316 (cit. on p. 46).

Roth, Joshua T., D'Aloisio, Anson, Cain, Christopher, Wilson, Bayu, Zhu, Yongda, and Becker, George D. (June 2024). "The effect of reionization on direct measurements of the mean free path". In: 530.4, pp. 5209–5219. DOI: 10.1093/mnras/stae1194. arXiv: 2311.06348 [astro-ph.CO] (cit. on pp. 128, 132).

Sandage, Allan (1965). "The Existence of a Major New Constituent of the Universe: the Quasistellar Galaxies." In: *Astrophysical Journal, vol. 141, p. 1560* 141, p. 1560 (cit. on p. 43).

Santos, Mario G, Ferramacho, Luis, Silva, MB, Amblard, A, and Cooray, A (2010). "Fast large volume simulations of the 21-cm signal from the reionization and pre-reionization epochs". In: *Monthly Notices of the Royal Astronomical Society* 406.4, pp. 2421–2432 (cit. on p. 52).

Satyavolu, Sindhu, Kulkarni, Girish, Keating, Laura C., and Haehnelt, Martin G. (May 2023). "The need for obscured supermassive black hole growth to explain quasar proximity zones in the epoch of reionization". In: 521.2, pp. 3108–3126. DOI: 10.1093/mnras/stad729. arXiv: 2209.08103 [astro-ph.GA] (cit. on pp. 102, 131).

Schervish, Mark J (2012). *Theory of statistics*. Springer Science & Business Media (cit. on p. 81).

Schindler, Jan-Torge, Bañados, Eduardo, Connor, Thomas, Decarli, Roberto, Fan, Xiaohui, Farina, Emanuele Paolo, Mazzucchelli, Chiara, Nanni, Riccardo, Rix, Hans-Walter, Stern, Daniel, et al. (2023). "The Pan-STARRS1 z> 5.6 Quasar Survey. III. The z 6 Quasar Luminosity Function". In: *The Astrophysical Journal* 943.1, p. 67 (cit. on pp. 76, 97, 103).

Schmidt, Maarten (1963). "3C273: a star-like object with large red-shift."" In: *A Century of Nature, Twenty One Discoveries That Changed Science and the World* (cit. on p. 43).

Scoccimarro, Roman (1998). "Transients from initial conditions: a perturbative analysis". In: *Monthly Notices of the Royal Astronomical Society* 299.4, pp. 1097–1118 (cit. on p. 56).

Shannon, Claude E. (1949). "Communication in the Presence of Noise". In: *Proceedings of the IRE* 37.1, pp. 10–21. DOI: 10.1109/JRPROC.1949.232969 (cit. on p. 59).

Shapiro, Paul R and Giroux, Mark L (1987). "Cosmological H II regions and the photoionization of the intergalactic medium". In: *Astrophysical Journal, Part 2-Letters to the Editor (ISSN 0004-637X), vol. 321, Oct. 15, 1987, p. L107-L112. Research supported by the Robert A. Welch Foundation.* 321, pp. L107–L112 (cit. on pp. 64, 88).

Shapiro, Paul R., Giroux, Mark L., and Babul, Arif (May 1994). "Reionization in a Cold Dark Matter Universe: The Feedback of Galaxy Formation on the Intergalactic Medium". In: 427, p. 25. DOI: 10.1086/174120 (cit. on p. 57).

Sharma, Yash, Davies, Frederick, Gaikwad, Prakash, Nasir, Fahad, and Bosman, Sarah (Apr. 2025). "Behavior of the Ly Damping Wings as a Function of Reionization Topology". In: *The Astrophysical Journal* 983, p. 118. DOI: 10.3847/1538-4357/adbe6d (cit. on pp. 54, 143).

Sheth, Ravi K and Tormen, Giuseppe (1999). "Large-scale bias and the peak background split". In: *Monthly Notices of the Royal Astronomical Society* 308.1, pp. 119–126 (cit. on pp. 28, 57, 88).

— (2002). "An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier". In: *Monthly Notices of the Royal Astronomical Society* 329.1, pp. 61–75 (cit. on pp. 26, 27, 29).

Singh, Saurabh, Nambissan T., Jishnu, Subrahmanyan, Ravi, Udaya Shankar, N., Girish, B. S., Raghunathan, A., Somashekar, R., Srivani, K. S., and Sathyanarayana Rao, Mayuri (2022). "On the detection of a cosmic dawn signal in the radio background". In: *Nature Astron.* 6.5, pp. 607–617. DOI: 10.1038/s41550-022-01610-5. arXiv: 2112.06778 [astro-ph.CO] (cit. on p. 43).

Sobacchi, Emanuele and Mesinger, Andrei (2013a). "How does radiative feedback from an ultraviolet background impact reionization?" In: *Monthly Notices of the Royal Astronomical Society* 432.4, pp. 3340–3348 (cit. on p. 57).

— (2013b). "The depletion of gas in high-redshift dwarf galaxies from an inhomogeneous reionization". In: *Monthly Notices of the Royal Astronomical Society: Letters* 432.1, pp. L51–L55 (cit. on p. 57).

Songaila, Antoinette and Cowie, Lennox L. (Oct. 2010). "The Evolution of Lyman Limit Absorption Systems to Redshift Six". In: 721.2, pp. 1448–1466. DOI: 10.1088/0004-637X/721/2/1448. arXiv: 1007.3262 [astro-ph.CO] (cit. on p. 107).

Springel, Volker and Hernquist, Lars (2003). "The history of star formation in a Λ cold dark matter universe". In: *Monthly Notices of the Royal Astronomical Society* 339.2, pp. 312–334 (cit. on p. 57).

Stengler-Larrea, Erik A., Boksenberg, Alec, Steidel, Charles C., Sargent, W. L. W., Bahcall, John N., Bergeron, Jacqueline, Hartig, George F., Jannuzi, Buell T., Kirhakos, Sofia, Savage, Blair D., Schneider, Donald P., Turnshek, David A., and Weymann, Ray J. (May 1995). "The Hubble Space Telescope Quasar Absorption Line Key Project. V. Redshift Evolution of Lyman Limit Absorption in the Spectra of a Large Sample of Quasars". In: 444, p. 64. DOI: 10.1086/175582 (cit. on p. 107).

Sun, Guochao and Furlanetto, Steven R (2016). "Constraints on the star formation efficiency of galaxies during the epoch of reionization". In: *Monthly Notices of the Royal Astronomical Society* 460.1, pp. 417–433 (cit. on p. 61).

Tegmark, Max (1997). "Measuring cosmological parameters with galaxy surveys". In: *Physical Review Letters* 79.20, p. 3806 (cit. on p. 82).

Trac, Hy and Cen, Renyue (2007). "Radiative transfer simulations of cosmic reionization. I. Methodology and initial results". In: *The Astrophysical Journal* 671.1, p. 1 (cit. on p. 51).

Trott, Cathryn M, Jordan, CH, Midgley, S, Barry, Nichole, Greig, B, Pindor, B, Cook, JH, Sleap, Greg, Tingay, SJ, Ung, Daniel, et al. (2020). "Deep multiredshift limits on Epoch of Reionization 21 cm power spectra from four seasons of Murchison Widefield Array observations". In: *Monthly Notices of the Royal Astronomical Society* 493.4, pp. 4711–4727 (cit. on p. 43).

Umeda, Hiroya, Ouchi, Masami, Nakajima, Kimihiko, Harikane, Yuichi, Ono, Yoshiaki, Xu, Yi, Isobe, Yuki, and Zhang, Yechi (2024). "JWST Measurements of Neutral Hydrogen Fractions and Ionized Bubble Sizes at z= 7–12 Obtained with Lyα Damping Wing Absorptions in 27 Bright Continuum Galaxies". In: *The Astrophysical Journal* 971.2, p. 124 (cit. on p. 70).

Wang, Feige, Davies, Frederick B., Yang, Jinyi, Hennawi, Joseph F., Fan, Xiaohui, Barth, Aaron J., Jiang, Linhua, Wu, Xue-Bing, Mudd, Dale M., Bañados, Eduardo, Bian, Fuyan, Decarli, Roberto, Eilers, Anna-Christina, Farina, Emanuele Paolo, Venemans, Bram, Walter, Fabian, and Yue, Minghao (June 2020). "A Significantly Neutral Intergalactic Medium Around the Luminous z = 7 Quasar J0252-0503". In: 896.1, 23, p. 23. DOI: 10.3847/1538-4357/ab8c45. arXiv: 2004.10877 [astro-ph.GA] (cit. on pp. 48, 49, 55).

Worseck, Gábor, Prochaska, J. Xavier, O'Meara, John M., Becker, George D., Ellison, Sara L., Lopez, Sebastian, Meiksin, Avery, Ménard, Brice, Murphy, Michael T., and Fumagalli, Michele (Dec. 2014). "The Giant Gemini GMOS survey of $z_{em} > 4.4$

quasars - I. Measuring the mean free path across cosmic time". In: 445.2, pp. 1745–1760. DOI: 10.1093/mnras/stu1827. arXiv: 1402.4154 [astro-ph.CO] (cit. on pp. 116, 118, 126, 127, 131).

Wyithe, J Stuart B and Loeb, Abraham (2003). "Reionization of hydrogen and helium by early stars and quasars". In: *The Astrophysical Journal* 586.2, p. 693 (cit. on p. 50).

Yue, Bin, Ferrara, Andrea, and Xu, Yidong (2016). "On the faint-end of the high-z galaxy luminosity function". In: *Monthly Notices of the Royal Astronomical Society* 463.2, pp. 1968–1979 (cit. on p. 61).

Zel'dovich, Yakov B. (1970). "Gravitational instability: An approximate theory for large density perturbations". In: *Astronomy & Astrophysics* 5. Available via NASA ADS, pp. 84–89. URL: https://ui.adsabs.harvard.edu/abs/1970A%26A.....5...84Z (cit. on pp. 16, 58).

Zhou, Yihao, Chen, Huanqing, Di Matteo, Tiziana, Ni, Yueying, Croft, Rupert A. C., and Bird, Simeon (Feb. 2024). "Modeling quasar proximity zones in a realistic cosmological environment with a self-consistent light curve". In: 528.2, pp. 3730–3744. DOI: 10.1093/mnras/stae172. arXiv: 2309.11571 [astro-ph.CO] (cit. on pp. 102, 131).

Zhu, Yongda, Becker, George D., Christenson, Holly M., D'Aloisio, Anson, Bosman, Sarah E. I., Bakx, Tom, D'Odorico, Valentina, Bischetti, Manuela, Cain, Christopher, Davies, Frederick B., Davies, Rebecca L., Eilers, Anna-Christina, Fan, Xiaohui, Gaikwad, Prakash, Haehnelt, Martin G., Keating, Laura C., Kulkarni, Girish, Lai, Samuel, Ma, Hai-Xia, Mesinger, Andrei, Qin, Yuxiang, Satyavolu, Sindhu, Takeuchi, Tsutomu T., Umehata, Hideki, and Yang, Jinyi (Oct. 2023). "Probing Ultralate Reionization: Direct Measurements of the Mean Free Path over $5 < z < 6$". In: 955.2, 115, p. 115. DOI: 10.3847/1538-4357/aceef4. arXiv: 2308.04614 [astro-ph.CO] (cit. on pp. 117, 118, 120–128, 131, 144).

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst have und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den Yash Mohan Sharma

Declaration:

I certify that I wrote this work independently and did not use any sources or aids other than those specified.

Heidelberg, the Yash Mohan Sharma

Concerning the usage of artifical intelligence

The author employed large language model-based services such as Grammarly to correct the language and improve the cohesiveness of the text throughout this thesis.

While the correction of language keeps the original text mostly untouched (substitution of single words), the improvement of language may have exchanged several words and the structure of sentences. The author wants to emphasize that Grammarly was used at no point to write sections, generate new sentences, create, or change scientific content, and the closeness to any cited works has been manually rechecked sentence by sentence for any alterations of the original text. However, the author can not be sure whether a certain phrasing in improved sentences belongs to related works, as Grammarly relies on “in-context learning”.