EFFICIENT HYDRODYNAMICAL SIMULATIONS OF THE INTERGALACTIC MEDIUM AND PARAMETER ESTIMATION

A Thesis

Submitted to the

Tata Institute of Fundamental Research
(A Deemed University)

for the degree of Doctor of Philosophy
in Physics

by

Prakash Gaikwad

NCRA • TIFR

National Centre for Radio Astrophysics
Tata Institute of Fundamental Research
2017

Dedication

I would like to dedicate my thesis to my beloved parents.

Acknowledgements

Jonathan: Chiang (Elder), this world isn't heaven at all, is it?
Elder: You are learning again, Jonathan Seagull.
Jonathan: Well, what happens from here? Where are we going? Is there no such place as heaven?
Elder: No, Jonathan, there is no such place. Heaven is not a place, and it is not a time. Heaven is being perfect (silence for a moment). You are a very fast flier, aren't you?
Jonathan: I $\cdots$ I enjoy speed
Elder: You will begin to touch heaven, Jonathan, in the moment that you touch perfect speed. And that isn't flying a thousand miles an hour, or a million, or flying at the speed of light. Because any number is a limit, and perfection doesn't have limits. Perfect speed, my son, is being there. - "Jonathan Livingston Seagull" by Richard Bach

My Ph.D. experience can be well summarized by the above excerpt from the book "Jonathan Livingston Seagull". The thesis presented here is not only my work but contributed by many people. I would like to use this opportunity to express my sincere gratitude to them. First, I would like to express my special appreciation and thanks to my Ph.D. supervisor Dr. Tirthankar Roy Choudhury and collaborator Prof. Raghunathan Srianand (IUCAA) for the continuous support of my Ph.D. study and related research, for their patience, motivation, and immense knowledge. Their guidance, comments and suggestions helped me in all the time of research and writing of this thesis. I would like to thank them for encouraging my research and for allowing me to grow as a research scientist. Their advice on both research as well as on my career has been priceless.

I would also like to thank my committee members, Dr. Divya Oberoi, Dr. Poonam Chandra, Dr. Yogesh Wadadekar and Dr. Nissim Kanekar for serving as my committee members and for useful suggestions on my Ph.D. work. It was a fantastic learning experience while discussing the majority of my research work with Dr. Aseem Paranjape, Prof. Kandaswamy Subramanian, Dr. Neeraj Gupta (all are IUCAA faculty) and Dr. Vikram Khaire (NCRA). I thank them for their comments and suggestions to improve the quality of the work presented in this thesis. All the faculty members in NCRA are very friendly and happily helped me in difficulties. I would like to pay my sincere gratitude to them. I am also grateful to all the NCRA staff including the administrative, scientific, technical, canteen etc for their unfailing support and assistance during my Ph.D. work.

A good support system is important to surviving and staying sane during Ph.D. Thanks to Rohit, J.H.N.S Aditya, Raghunath, Narendra and Arun with whom I enjoyed playing *counter strike* and watching many insane movies like *The Agrenz* and *Gunda*. My Ph.D would have been difficult without the support from my friends Atul, Avishek, Biny, Dinesh, Nayana, Omkar, Preetish, Peter, Sushma, Prakash, Sravani, Narendra, Gunjan, Minazur, Aditya, Atrideb, Barnali, Rajeshwari, Tanvir, Niladri, Isha. I also want to thank Sushan (wife of T. R. Choudhury), Veena (wife of R. Srianand), Swapnali (wife of Vikram) and Dr. Dhram vir Lal (Faculty at NCRA) for giving many parties and many memorable events in my life.

The significant part of this thesis is based on numerical methods. I would like to thank the *python* community and especially the programmers developing and maintaining the packages like *matplotlib, mpi4py, numpy, scipy, hdf5, h5py, numba, mayavi* etc. A special thank to *stackoverflow community* where most of my questions were answered by anonymous friends.

In the darkest hour of my Ph.D. the books like *Jonathan Livingston Seagull* by Richard Bach, *Ek hota Carver* (Biography of George Washington Carver) by Veena Gavankar, *Oh! Life relax please, Karma yoga, Personal excellence through the bhagavad gita* by Swami Sukhabodhananda, *7 Habits of Highly Effective People* by Steven Covey and speeches by Shivaji rao bhosale, Babasaheb Purandare were like lighthouse and guided me through difficult times. I thank them all for such inspirational work.

Finally, I would like to thank and dedicate this thesis to my parents, brother, sister and two important historical figures Mahatma Jyotiba Phule and Dr. Babasaheb Ambedkar who are the greatest source of inspiration.

Declaration

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions. The work was done under the guidance of Dr. Tirthankar Roy Choudhury, at the Tata Institute of Fundamental Research, Mumbai.

(Prakash Gaikwad)

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

(Dr. Tirthankar Roy Choudhury)

Date: December 6, 2017

Abstract

The intergalactic medium (IGM) is imprinted with rich information of the cosmological initial conditions and the astrophysical processes that occurred after the formation of large scale structures. The sensitivity of the QSO absorption spectra to low density gas in IGM is suitable for probing the physical conditions prevailing in the IGM. With the advent of high quality observations of QSO absorption spectra, an efficient method to simulate the IGM would be useful for parameter estimation. In this thesis we have developed set of tools which are used in the post-processing step of hydrodynamical simulations to efficiently model the observed properties of the IGM. The main outcomes of the thesis are as follows:

1. We have developed a module named “Code for Ionization and Temperature Evolution (CITE)” for calculating the IGM temperature evolution from high to low redshifts by post-processing the smoothed particle hydrodynamic simulation GADGET-2. In particular, we model the thermal and ionization histories that are not computed self-consistently by default in GADGET-2. We capture the effect of pressure smoothing by running GADGET-2 at an elevated temperature floor and using an appropriate smoothing kernel. We further developed a module for “Generating Lyman Alpha forest Spectra in Simulation” (GLASS). Using GLASS, we calculated the Lyα transmitted flux that has signal-to-noise ratio (SNR), spectral resolution and line spread function (LSF) effects similar to the observational data. We show that the line-of-sight density, velocity, temperature field and Lyα transmitted flux is

remarkably similar to those obtained from self-consistent simulations like GADGET-3. Our method, using GADGET-2 + CITE + GLASS, allows us to model various observed properties of the Lyα forest that enables an efficient parameter estimation.

2. The statistics based on parameters obtained using Voigt profile fitting of Lyα forest are useful in deriving thermal history and equation of state of the IGM. However, the Voigt profile decomposition is usually subjective, laborious and time consuming process. We have developed a parallel code called "VoIgt profile Parameter Estimation Routine (VIPER)" for automatically fitting the H I Ly-α forest seen in the spectra of QSOs. We obtained the H I column density distribution function (CDDF) and line width (b) parameter distribution for $z < 0.45$ using spectra of 82 QSOs obtained using Cosmic Origins Spectrograph on board Hubble Space Telescope (HST-COS) and VIPER. Consistency of these with the existing measurements in the literature validate our code. The parallel and automated nature of VIPER allows us to simultaneously fit large number of simulated spectra and to explore a wide parameter space efficiently.

3. We validated our method of evolving thermal and ionization state of the IGM using CITE with other simulations in the past and full hydrodynamic simulations like GADGET-3. For low resolution (similar to that achieved in HST-COS data at $z < 0.5$) simulation, we show the consistency of our method with other simulations in the literature by comparing with three metrics: (i) thermal history parameters, (ii) distribution of baryons in phase diagram at $z = 0$ and (iii) the correlation between baryon overdensity Δ vs H I column density, $N_{\rm HI}$. On the other hand, for high resolution (typically achieved in echelle data) simulations (at $2 \leq z \leq 4$), the dynamical evolution of SPH particles at finite pressure is an important effect. In this case, we validate our procedure by comparing different statistics derived from our

method with those derived using self-consistent simulations with GADGET-3. These statistics are: line of sight density field power spectrum (PS), flux probability distribution function (PDF), flux PS, wavelet statistics, curvature statistics, H I CDDF, linewidth (b) distribution and b versus $\log \mathrm{N}_{\mathrm{HI}}$ scatter. For the temperature floor of 10^4 K and typical signal-to-noise of 25, the results agree well within 1σ level. Moreover for a given cosmology, we gain a factor of $\sim N$ in computing time for modelling the intergalactic medium under $N \gg 1$ different thermal histories. Our method is computationally less expensive and flexible thus allowing us to explore a large parameter space. In addition, our method allows us to simulate the non-equilibrium evolution of thermal and ionization state of the gas and include heating due to non-standard sources like cosmic rays and high energy γ-rays from Blazars.

4. We constrain the H I photoionization rate (Γ_{HI}) at $z \lesssim 0.45$ in four redshift bins by comparing the flux PDF, flux PS and CDDF of the Lyα forest data along 82 QSO sightlines obtained using HST-COS with models generated from GADGET-2 and post-processed with CITE, GLASS and VIPER. Our method allows rigorous estimation of the error covariance matrix for various statistical quantities of interest. We explore large parameter space and perform χ^2 minimization to obtain Γ_{HI}. We notice that the b parameters from the simulations are smaller than what are derived from the observations. We show the observed b parameter distribution and b vs $\log \mathrm{N}_{\mathrm{HI}}$ scatter can be reproduced in simulation by introducing sub-grid scale turbulence. However, it has very little influence on the derived Γ_{HI}. We find that the best-fit $\Gamma_{\mathrm{HI}}(z)$ increases with z and follows $(4 \pm 0.1) \times 10^{-14} (1+z)^{4.99 \pm 0.12}$ s^{-1}. At any given z the typical uncertainties $\Delta\Gamma_{\mathrm{HI}}/\Gamma_{\mathrm{HI}}$ are ~ 25 per cent which contains not only the statistical errors but also those arising from possible degeneracy with the thermal history of the IGM and cosmological parameters and uncertainties in

fitting the QSO continuum. These values of Γ_{HI} are consistent with the hydrogen ionizing ultra-violet (UV) background being dominated mainly by QSOs without needing any contribution from the non-standard sources of the UV photons.

In order to synchronize with large flow of observational data from the upcoming surveys, IGM simulations needs to be efficient, flexible and sufficiently accurate to probe large parameter space. *The tools and techniques developed in this thesis provide a starting point to efficiently simulate the IGM.*

Synopsis

Name : Prakash Gaikwad
Title of the thesis : Efficient hydrodynamical simulations of the intergalactic medium and parameter estimation
Supervisor : Dr. Tirthankar Roy Choudhury

1 Introduction

According to the standard big bang model, the Universe was extremely hot at the beginning and was composed of a soup of elementary particles tightly coupled to photons. As the Universe expanded, the photons decoupled from baryons allowing the electrons, protons and neutrons to combine and form neutral hydrogen (H I) and helium (He I) atoms. This important event in the history of the Universe is known as "epoch of recombination". The small scale initial perturbations, generated by quantum fluctuations in the early Universe, started to grow via gravitational instability to eventually form stars and galaxies. There was also a significant amount of diffuse and clumpy gas around galaxies that failed to be a part of galaxies. This all-pervading gas that is not associated with galaxies is known as the intergalactic medium (IGM). The IGM is a crucial ingredient of any theory of structure formation and evolution because, by definition, all the baryons were part of IGM at sufficiently early times. The IGM also has signatures of various feedback processes associated with the galaxy formation in the form of presence of metals, thermal history etc.After the epoch of recombination, the IGM went through two major phase changes: (i) the photoionization of H I and He I at $z \sim 6 - 15$ by radiation from stellar light escaping the early galaxies and (ii) the photoionization of He II at $z \sim 3$ by radiation from accreting blackholes in Quasi-Stellar Objects (QSOs). The temperature of the low density gas in the IGM is expected to increase during the reionization processes because of residual photoheating. Subsequently at later epochs, the IGM is maintained at highly ionized state by the ultra-violet background (UVB). This UVB is contributed by radiation from QSOs and stars in galaxies (see for example, Fardal et al. 1998; Haardt and Madau 1996; Khaire and Srianand 2015a; Miralda-Escude and Ostriker 1990; Shull et al. 1999). Thus the IGM is imprinted with rich information on the cosmological initial conditions and the astrophysical processes that occurred after the formation of large scale structures.

The Lyα forest absorption seen in the spectra of the luminous distant QSOs is one of the most sensitive tools to study the physical conditions prevailing in IGM. It traces the distribution of H I in the Universe at mildly non-linear overdensities ($\Delta \lesssim 10$, Bi and Davidsen 1997; Choudhury et al. 2001; Croft et al. 1997; Miralda-Escudé et al. 1996). The observed properties of the Lyα forest are sensitive to fluctuations in the cosmic density and velocity fields and physical conditions such as the temperature, turbulence and ionizing radiation prevailing in the IGM (Cen et al. 1994; Hernquist et al. 1996; Zhang et al. 1995). In recent years, improvements in observational techniques and cosmological simulations have made it possible to constrain the cosmological and astrophysical parameters from the Lyα forest. These simulations can incorporate different complex astrophysical processes such as the radiative heating and cooling of the IGM, large scale shocks due to structure formation and various feedback processes that drives the interactions between IGM and galaxies (Davé et al. 2010; Kollmeier et al. 2006; McDonald et al. 2006; Schaye et al. 2010; Viel et al. 2013). While the current state-of-the-art hydrodynamical simulations are extremely useful for probing the physical properties of the IGM, the computational expenses severely limit their usage for constraining the unknown model parameters and their associated errors.

Various approaches have been introduced to keep the computational expense within manageable limits while exploring the large parameter space. These approaches involve, for example, (i) expanding the statistical quantities of interest in a Taylor series around a chosen best-guess value (Viel et al. 2009; Viel and

Haehnelt 2006), (ii) running simulations on a carefully chosen grid in the parameter space and then interpolating between their outputs (McDonald et al. 2005) and (iii) deriving scaling relations between different parameters from a limited number of hydrodynamical simulations to study parameter degeneracy (Bolton and Haehnelt 2007; Bolton et al. 2005; Faucher-Giguère et al. 2008b). Since many of the parameters, particularly those related to the thermal state of the IGM, are poorly understood, obtaining robust constraints would require exploring a sufficiently wide range of parameter values. It is thus crucial to develop newer methods of simulating the IGM that are efficient, flexible and at the same time sufficiently accurate. This forms one of the main motivation of this thesis.

2 Outline

The main scientific contribution of this thesis is to develop an efficient way of modelling the IGM as a post-processing step of the cosmological N-body hydrodynamical simulation GADGET-2 and to apply these models to constrain astrophysical parameters from observations using various statistics. For this purpose, we have developed following post-processing tools:

(i) A module that accounts for the effect of thermal, ionization evolution and dynamic pressure of the gas.

(ii) A tool to generate the Lyα forest by shooting random sightlines through the simulation box.

(iii) A module that automatically decomposes the Lyα forest in multi-component Voigt profiles.

(iv) Different codes to derive the various Lyα flux statistics such as flux probability distribution function (PDF), flux power spectrum (PS), wavelet statistics (Lidz et al. 2010) and curvature statistics (Becker et al. 2011).

Using these tools, we model various observed properties of the IGM and compare with observations to constrain astrophysical parameters with appropriate errorbars.

The thesis consists of six Chapters. Chapter 1 contains the introduction to the subjects with highlights to the major objective of the thesis. In Chapters 2 and 3, we develop the necessary tools to simulate and analyze the Lyα forest. In Chapter 4, we show the consistency of these tools with other simulations in the literature. In Chapter 5, we apply these tools and constrain the ionizing UV radiation background (UVB) at $z \lesssim 0.5$ using Lyα forest observations obtained using the Cosmic Origins Spectrograph on-board Hubble Space Telescope (HST-COS). Finally we summarize the main results of the thesis in Chapter 6. We summarize each chapter in the following.

2.1 Code for Ionization and Temperature Evolution (CITE)

We use the publicly available smoothed particle hydrodynamic (SPH) code GADGET-2[1] (Springel 2005) to generate the density and velocity field in a periodic box. We store the GADGET-2 snapshots from high redshifts (e.g. $z = 6$) to low redshifts ($z = 2$) in intervals of $\Delta z = 0.1$. An unique identification number assigned to each particle in GADGET-2 can be used to trace its density and temperature evolution. The modelling of the Lyα optical depth involves shooting of random sightlines through simulation box and the calculation of three baryonic fields, namely, (i) the overdensity $\Delta_{\rm b}$ (hereafter simply Δ), (ii) the peculiar velocity v and (iii) the temperature T along these sighlines. Hui and Gnedin (1997) have shown that the temperature and density of the low density gas exposed to photo-heating by UVB follows a tight power-law correlation given by,

$$T = T_0 \, \Delta^{\gamma-1} \tag{1}$$

where T_0 is mean IGM temperature and γ is slope of equation of state. However, the temperature and density obtained from GADGET-2 do not follow this relation. It is because GADGET-2 does not include processes like photoheating and radiative cooling. This shortcoming has been addressed in subsequent versions of the GADGET, e.g., GADGET-3 (as discussed in Bolton et al. 2006) where one can perform the simulation in presence of a UVB. However, performing a GADGET-3

---

[1] `http://www.mpa.mpa-garching.mpg.de/gadget/`

simulation is computationally expensive especially while probing large parameter space.

As a complimentary tool, we develope a "Code for Ionization and Temperature Evolution" (CITE) to model the thermal and ionization evolution of particles in the post-processing step of GADGET-2. The basic equation governing the evolution of IGM temperature (Hui and Gnedin 1997) is,

$$\frac{dT}{dt} = \left(-2HT + \frac{2T}{3\Delta}\frac{d\Delta}{dt} + \frac{dT_{\mathrm{shock}}}{dt} \right) + \frac{dT_{\mathrm{IE}}}{dt} + \frac{dT_{\mathrm{other}}}{dt} \ . \tag{2}$$

In the above equation the first three terms on right hand side represent, respectively, the rate of cooling due to Hubble expansion, adiabatic heating and/or cooling arising from the evolution of the densities of gas particles and the change in temperature because of the structure formation shocks. These three mechanisms are taken into account in the default run of the GADGET-2. The fourth and fifth terms on the right hand side of Eq. 2.6, corresponding to change in temperature due to change in internal energy per particle and change in temperature due to other heating/cooling processes respectively, are not accounted for in GADGET-2.

The method followed by us in CITE to account for radiative heating and cooling processes is as follows:

(i) At high redshifts ($z_1 = 6$) we assume power-law $T - \Delta$ relation given in Eq. 1 for unshocked particles. We solve the equilibrium ionization evolution equation for a given UVB to calculate various ionization fractions of H and He.

(ii) Given the ionization fractions and the temperatures, the last two terms on the right hand side of Eq. 2.6 can be calculated.

(iii) We obtain the temperature of the particles in the next time step ($z_2 = z_1 - \Delta z$) accounting for all the five terms in Eq. 2.6.

(iv) For redshift z_2, we solve the equilibrium (or non-equilibrium, if desired) ionization evolution equation to calculate various ionization fractions.

(v) We repeat the steps (ii)-(iv) to obtain the temperature of the particle at subsequent redshifts.

We find that the resulting $T - \Delta$ relation (i.e., that obtained after applying CITE) for the particles is power-law with T_0 and γ consistent with those from GADGET-3. CITE offers many advantages for the IGM studies as given below:

(1) Because CITE is based on post-processing the GADGET-2 output, the method is computationally less expensive.

(2) CITE allows us to explore large thermal history parameter space without performing the full SPH simulation from high-z.

(3) CITE provides flexibility to (i) simulate the non-equilibrium evolution of thermal and ionization state of the gas, (ii) incorporate the radiative cooling for a wide range of metallicities and (iii) include heating due to non-standard sources like cosmic rays and high energy γ-rays from Blazars.

Despite offering flexibility, the obvious shortcoming of CITE is that the diffuse gas is evolved dynamically at effectively zero pressure (because of its low temperature), rather than the pressure it would have if it were at $T \sim 10^4$ K typical of photoionized gas. Thus the dynamical impact of the diffuse IGM pressure (especially at small scales) is not modelled self-consistently in CITE. However, we expect that the effect of dynamic pressure is not important for moderate to low resolution (gas particle mass $\delta m \sim 1.26 \times 10^7\ h^{-1}\ \mathrm{M}_\odot$, pixel size $\delta x \sim 48.8 h^{-1}$ ckpc) Lyα forest studies typically achieved with HST-COS at low-z ($z < 0.5$). On the other hand, for studying the high-z ($2 \leq z \leq 4$) Lyα forest one usually uses higher resolution echelle data. When we use appropriate high resolution (gas particle mass $\delta m \sim 1.01 \times 10^5\ h^{-1}\ \mathrm{M}_\odot$, pixel size $\delta x \sim 9.77$ ckpc) simulation boxes, we notice that the density (Δ) and velocity (v) fields are smoother for GADGET-3 as compared to those from GADGET-2 (known as "pressure smoothing"). This is because the temperature of the SPH particles in GADGET-2 is not calculated self-consistently, the unshocked gas is effectively evolved at zero pressure in GADGET-2, and the local Jeans length of the particles, responsible for pressure smoothing in GADGET-3, is comparable to the resolution of the simulation box.

The effect of dynamical pressure can be captured by running GADGET-2 at an elevated temperature floor $\sim 10^4$ K, post-processing with CITE, and smoothing (in 3 dimensions) the density and velocity fields over a local Jeans scale (over which pressure smoothing effects are important) with a modified SPH kernel.

We further develop a module for "Generating Lyman Alpha forest Spectra in Simulation" (GLASS). Using GLASS, we calculated the Lyα transmitted flux that has signal-to-noise ratio (SNR), spectral resolution and line spread function (LSF) effects similar to the observational data. We show that the line-of-sight density, velocity, temperature field and Lyα transmitted flux is remarkably similar to those obtained from self-consistent simulations like GADGET-3. Our method to simulate the Lyα forest is computationally less expensive, flexible and accurate to within 5 percent. The details of CITE can be found in Chapter 2 of the thesis (see also Gaikwad et al. 2017a,b).

2.2 VoIgt profile Parameter Estimation Routine (VIPER)

The observations of the Lyα forest have regularly been used to constrain cosmological and astrophysical parameters related to IGM physics. Usually constraining these parameters involves comparing different statistics of the Lyα forest derived from observed spectra with those from the simulated ones. These statistics are broadly divided into two cases. In the first case, Lyα transmitted flux is treated as a continuous field quantity. In particular, the mean flux, the flux PDF and the flux PS have been used to constrain cosmological parameters (Choudhury et al. 2001; Viel et al. 2009, 2006, 2004a,b), and astrophysical parameters such as T_0, γ and $\Gamma_{\rm HI}$ (Becker et al. 2011; Lidz et al. 2010). In the second case, Lyα forest is decomposed into multiple Voigt profiles. The line width distribution function calculated from Voigt profile fitting is sensitive to the thermal history and the energy injected by various astrophysical processes in the form of heat and turbulent motions in the IGM (Davé et al. 2001; Schaye et al. 1999, 2000). Similarly, the column density distribution function (CDDF) calculated from Voigt profile decomposition is sensitive to $\Gamma_{\rm HI}$ (Gurvich et al. 2017; Kollmeier et al. 2014; Shull et al. 2015) and cosmological parameters (Shull et al. 2012). While statistics based on parameters obtained using Voigt profile fitting are useful in deriving thermal history and equation of state of the IGM, the Voigt profile decomposition is usually subjective, laborious and time consuming process. Therefore, a large parameter space exploration in simulations is usually difficult.

We develop a parallel processing module "VoIgt profile Parameter Estimation Routine" (VIPER) to fit the Lyα forest with multiple Voigt profiles automatically. In VIPER, the blended and saturated features are fitted simultaneously with multi-component Voigt profiles. An objective criteria based on information theory is used to find the number of Voigt profiles needed to describe the Lyα forest. Here we briefly outline the steps involved in VIPER:

(i) First the crude significance level (CSL) is calculated from flux and SNR per pixel as a function of wavelength.

(ii) The maxima (peaks) in CSL above a certain threshold (CSL $\geq$ 1.5) are identified.

(iii) The minima are found on either side of maxima to enclose the peaks in region accounting for blending effects.

(iv) In each region as many component as necessary are fitted based on an objective criteria namely "Akaike Information Criteria with Correction" (AICC).

(v) Finally the rigorous significance level (RSL) of each line is calculated to account for LSF effect and lines are accepted if RSL $>$ 4.

We fit 82 HST-COS Lyα forest spectra using VIPER and compiled a Lyα line catalog called "VIPER line catalog". We show that the Voigt profiles fitted to observed spectra using VIPER match well, in terms of number of components and the values of fitted parameter along with the errorbar, with those fitted using semi-numerical method in Danforth et al. (2016). The median b parameter from VIPER (32.9 ± 20.8 km s^{-1}) is consistent with that from Danforth et al. (2016, 33.9 ± 18.3 km s^{-1}). Also, the median $\log \mathrm{N_{HI}}$ from VIPER (13.39 ± 0.61) is in good agreement with that from Danforth et al. (2016, 13.38 ± 0.63). We calculate the appropriate redshift path length $\Delta z(\mathrm{N_{HI}})$ and the sensitivity curve from the HST-COS data. Subsequently, we calculate the CDDF after accounting for the incompleteness of the sample. We show that our calculated CDDF in the redshift range ($0.075 \leq z \leq 0.45$) is consistent (KS test p-value is 0.83) with that of Danforth et al. (2016) CDDF in the redshift range ($0 \leq z \leq 0.75$). The parallel and automated nature of VIPER allows us to simultaneously fit large number of simulated spectra and to explore a wide parameter space efficiently. The details of VIPER are given in Chapter 3 of the thesis (also refer to Gaikwad et al. 2017c).

2.3 Validation of CITE

Chapter 4 of the thesis focuses on validating our method of evolving thermal and ionization state of the IGM using CITE. For low resolution (similar to that achieved in HST-COS data) simulation, we show the consistency of our method with other simulations in the literature by comparing with three metrics:

(i) thermal history parameters: our simulation predicts $T_0 \sim 5000$ K and $\gamma \sim 1.6$ in the redshift range $z = 0.1$ to 0.45. These values are shown to be insensitive to our choice of T_0 and γ at an initial redshift, $z_1 = 2.1$;

(ii) distribution of baryons in phase diagram at $z = 0$: We find ~ 34 per cent of baryons are in diffuse phase, ~ 29 per cent in warm hot intergalactic medium (WHIM), ~ 18 per cent in hot halo and ~ 19 per cent in condensed phase and

(iii) the correlation between baryon overdensity Δ vs H I column density, $\mathrm{N_{HI}}$, in the redshift range $0.2 < z < 0.3$: we find $\Delta = 34.8 \pm 5.9\,(\mathrm{N_{HI}}/10^{14})^{0.770 \pm 0.022}$.

We show that all these predictions compare well with those of low-z simulations in the literature (Davé et al. 2010; Smith et al. 2011) that include different feedback processes at varied levels. Note that the feedback processes such as galactic winds or AGN feedback are not incorporated in our simulations.

On the other hand, for high resolution (typically achieved in echelle data) simulations (at $2 \leq z \leq 4$), the dynamical evolution of SPH particles at finite pressure is an important effect. It is well known that the pressure smoothing in GADGET-3 is not only decided by the instantaneous density and temperature of the particles but also to some extent by the thermal history (Gnedin and Hui 1998; Kulkarni et al. 2015). To illustrate this, we perform three high resolution simulations (gas particle mass $\delta m \sim 1.01 \times 10^5\ h^{-1}\ \mathrm{M}_\odot$, pixel size $\delta x \sim 9.77$ ckpc) with same initial conditions (i) G2-LTF: GADGET-2 with low temperature ($T \sim 100$ K) floor in which local Jeans length is decided by instantaneous density and temperature and (ii) G2-HTF: GADGET-2 with high temperature ($T \sim 10^4$ K) floor in which even the unshocked gas is evolved at a pressure appropriate for a photoionized gas at $T = 10^4$ K and (iii) GADGET-3: a reference model for comparison with G2-LTF and G2-HTF model.

First, we obtain the evolution of thermal history parameters T_0 and γ by estimating the temperature of the SPH particles from CITE. The T_0 and γ evolution from our model is in very good agreement with that from GADGET-3. CITE also provides us with enough flexibility to solve the non-equilibrium ionization evolution equation. The T_0 and γ evolution for non-equilibrium case is considerably different (T_0 is larger by ~ 60 percent and γ is smaller by 15 percent at $z = 3.7$) than that for equilibrium case. The T_0 evolution for non-equilibrium case from our method is consistent with that from Puchwein et al. (2015, difference less than 2.5 percent). Using our module GLASS, we generate the Lyα forest spectra by shooting random sightlines through simulation box in all the 3 models. The resulting Lyα forest spectra along sightline are remarkably similar in the G2-HTF and GADGET-3 methods. We compare the G2-LTF and G2-HTF with the GADGET-3 model using 8 different statistics, namely: (i) 1D density field PS, (ii) flux PDF, (iii) flux PS, (iv) wavelet PDF (Lidz et al. 2010), (v) curvature PDF (Becker et al. 2011), (vi) CDDF, (vii) linewidth (b) distribution and (viii) b vs $\log \mathrm{N_{HI}}$ correlation, at four different redshift $z = 2.5, 3.0, 3.5$ and 4.0. We have developed a statistical module to calculate the 8 statistics mentioned above. Treating the GADGET-3 model as the reference, we demonstrate that the Γ_{HI} can be recovered, using flux PDF and flux PS statistics, within 1σ statistical uncertainty using the G2-HTF model. We find that the G2-HTF model is in general very good agreement (within 1σ) with GADGET-3 model at all redshifts. Using enhanced Haardt and Madau (2012) photo-heating rates, we obtain a thermal history such that T_0 is increased by a factor of ~ 2. We show that our method for such significantly different thermal history is consistent (in 1σ) with GADGET-3 simulation. We refer the reader to Gaikwad et al. (2017a,b) for detail analysis.

2.4 Application of CITE and VIPER: H I photoionization rate at low-z

In Chapter 5 of the thesis, we constrain the UVB (at $z < 2$) responsible for maintaining highly ionized state of IGM using the tools CITE, GLASS and VIPER developed in Chapters 2 and 3. The accurate characterization UVB is important for modeling the thermal and ionization state of the IGM (Becker and Bolton 2013; Becker et al. 2011; Boera et al. 2014; Lidz et al. 2010), measuring the baryon content in the IGM (Shull et al. 2012), and deriving metal abundances in the IGM using ionization corrections (Carswell et al. 2002; Peeples et al. 2014; Shull et al. 2014; Songaila 2001).

The stellar contribution to the UVB depends crucially on the average fraction of ionizing photons escaping the galaxies known as the escape fraction ($f_{\rm esc}$). The theoretical modelling of $f_{\rm esc}$ is difficult as it depends on various physical factors such as the galaxy mass, morphology, composition of the interstellar medium (ISM), spatial distribution of gas and supernova rates (Cen and Kimm 2015; Ricotti and Shull 2000; Roy et al. 2015). Measuring $f_{\rm esc}$ directly from observations too is quite challenging. Only the handful of individual galaxies are detected to show escaping ionizing photons, 8 at low$-z$ (Bergvall et al. 2006; Borthakur et al. 2014; Izotov et al. 2016a,c) and 2 at high-z (Shapley et al. 2016; Vanzella et al. 2016). The 3σ upper limits on the average escape fraction at low-z is ~ 0.01 to 0.2 (Cowie et al. 2009; Leitet et al. 2013; Siana et al. 2010) and at high $z \sim 0.04$ to 0.15.

One way of constraining $f_{\rm esc}$ (and hence the stellar contribution to the UVB) is by using measured $\Gamma_{\rm HI}$ (Inoue et al. 2006; Khaire et al. 2016). This can be done by simulating the observed properties of the Lyα forest (far away from the proximity of QSOs) such as CDDF (Kollmeier et al. 2014; Shull et al. 2015). To observe the Lyα forest at low-z (i.e., $z < 0.5$), an UV spectrograph on a space telescope is needed for the Lyα forest observations. Thanks to the HST-COS we have good quality observations of the low-z Lyα forest. These observations have been used by Kollmeier et al. (2014); Shull et al. (2015) to place constraints on $\Gamma_{\rm HI}$. It turns out that there is a tension between the inferred $\Gamma_{\rm HI}$ by Kollmeier et al. (2014) and Shull et al. (2015), where both use Lyα forest data by Danforth et al. (2016)

but different simulations. The inferred $\Gamma_{\rm HI}$ values disagree by a factor of ~ 2.5. Given such a wide disagreement, it is worth taking an independent closer look at the $\Gamma_{\rm HI}$ measurements at low-z using the Lyα forest with a careful analysis of the systematics in the data as well as modeling uncertainties.

We measure $\Gamma_{\rm HI}$ at $z \leq 0.45$ from a sample of 82 QSO spectra obtained with HST-COS and hydrodynamical simulations using GADGET-2 post-processed with CITE and GLASS. The basic steps for constraining any astrophysical quantity from Lyα forest are

(i) to identify a statistics sensitive to the quantity,

(ii) derive the statistics from Lyα forest observations,

(iii) derive the same statistics by modeling the Lyα forest in simulation treating the quantity of interest as a free parameter,

(iv) compute χ^2 between the statistics derived from observation and simulation and

(v) the best fit value and the associated statistical error on the quantity is obtained from χ^2 parabola.

For a fair comparison, we mimic the simulated Lyα forest as close to observations as possible in terms of the SNR, resolution and LSF. The spectra generated using our method are remarkably similar to the observed spectra. We use three statistics (i.e., flux PDF, flux PS and CDDF) and χ^2 minimization using appropriate covariance matrices to compare the observations with the model predictions. We measured $\Gamma_{\rm HI}$ in four different redshift bins (of $\Delta z = 0.1$) centered at $z = 0.1125, 0.2, 0.3, 0.4$ using constraints from these statistics mentioned above. We estimated the associated errors by varying thermal history parameters, cosmological parameters and continuum fitted to the observed spectrum. Due to limited wavelength range covered in the HST-COS spectrum used in this study, the $\Gamma_{\rm HI}$ measurement for the highest redshift bin (i.e $z = 0.4$) is likely to be affected by the contamination of Lyβ forest absorption from higher-z. We contaminated our simulated Lyα forest at $z = 0.4$ by Lyβ forest from $z = 0.6$ and corrected for the effect of Lyβ contamination in our $\Gamma_{\rm HI}$ measurement for this z bin. The measured Γ_{12} ($\equiv \Gamma_{\rm HI}/10^{-12}$ s^{-1}) values at redshift bins $z = 0.1125, 0.2, 0.3, 0.4$ are $0.066 \pm 0.015,\ 0.100 \pm 0.021,\ 0.145 \pm 0.037,\ 0.210 \pm 0.052$, respectively.

We find that the b parameters of Voigt profile components from simulations are typically underestimated as compared to observations. This difference can be rectified by including the Gaussian distributed line width parameter $b_{\rm turb}$ ($\mu = 20$ km s^{-1}and $\sigma = 10$ km s^{-1}) at each pixel in the simulation (Muzahid et al. 2012; Tripp et al. 2008). The resulting line width distribution from simulations matches roughly with observed line width distribution, scatter and lower envelope of the b vs $\log {\rm N_{HI}}$ distribution. However, the CDDF has little effect of additional $b_{\rm turb}$ ($<$ 7 percent) and the $\Gamma_{\rm HI}$ constraints are mildly affected ($<$ 9 percent). On the other hand if we consider additional heating effect (Viel et al. 2017) for the excess broadening then the $\Gamma_{\rm HI}$ obtained will be slightly reduced (roughly scale as $T^{-0.7}$).

Our final quoted errors in the $\Gamma_{\rm HI}$ measurements include possible uncertainties coming from the statistical uncertainty ($\sim$ 14 per cent), cosmic variance ($\sim$ 3 per cent), cosmological parameters uncertainty ($\sim$ 10 per cent) and continuum uncertainty (systematic uncertainty $\sim$ 7 per cent). Uncertainty in $\Gamma_{\rm HI}$ due to uncertainty in thermal history parameters, over the range considered here, is small and within statistical uncertainty. As expected based on UVB models, even in the small redshift range covered in our study the measured $\Gamma_{\rm HI}$ shows a rapid evolution with z. We fit the redshift evolution of Γ_{12} as $\Gamma_{12} = 0.040 \pm 0.001\,(1+z)^{4.99\pm0.12}$ at $0.075 \leq z \leq 0.45$. The $\Gamma_{\rm HI}(z)$ obtained here are consistent with the measurement of Shull et al. (2015) however our $\Gamma_{\rm HI}$ measurement at $z = 0.1$ is factor ~ 2.7 smaller than Kollmeier et al. (2014).

The $\Gamma_{\rm HI}$ measurement at any z_1 depends on the emissivities of the ionizing sources at $z \geq z_1$ and Lyman continuum opacity of the IGM. We considered the updated emissivities of QSOs and galaxies (with $f_{\rm esc}$ as a free parameter) and two different H I column density distribution as a function of z obtained by Haardt and Madau (2012) and Inoue et al. (2014) and obtained $\Gamma_{\rm HI}$ using Khaire and Srianand (2015a) UVB code. We find that for, both H I distributions, our derived $\Gamma_{\rm HI}(z)$ is consistent with being contributed only by QSOs. This is true even if we allow for variations in the UV spectral index of QSOs. We also find the maximum 3σ upper limit on $f_{\rm esc}$ at $z < 2$, allowing for uncertainty in far UV spectral index and cloud distribution $f(N_{\rm HI}, z)$ of Inoue et al. (2014), is 0.008. This is consistent with 3σ upper limits on average $f_{\rm esc}$ (i.e ≤ 0.02) obtained by stacking samples of galaxies probing average galaxy mass $M \geq 10^{9.3} M_{\odot}$.

Our measurements suggest that the contribution of low mass galaxies to average $f_{\rm esc}$ will also be small. *Our study confirms that there is no crisis at low redshift in accounting for the observed Lyman continuum photons using standard known luminous astronomical sources as claimed by* Kollmeier et al. (2014). Thus our $\Gamma_{\rm HI}(z)$ measurement can in turn be used to place a strong constraint on the contributions of decaying dark matter to the low-z UVB. For more details we refer the reader to Gaikwad et al. (2017b,c).

2.5 Future Outlook

In Chapter 6, we provide the general outlook and future directions related to IGM science. With the rapid advancement in observations and simulations, IGM science is entering a new phase that will provide new opportunities for other areas of astrophysics and cosmology to exploit the IGM as a tool for testing models of interest. The upcoming surveys with ground based telescopes like Thirty Meter Telescope[1] (TMT), Giant Magellan Telescope[2] (GMT), Giant Segmented Mirror Telescope[3] (GSMT), Extremely Large Telescope[4] (ELT), Sloan Digital Sky Survey[5] (SDSS) and space telescope like James Webb Space Telescope[6] (JWST), promises to settle a number of outstanding questions in IGM science such as

(i) circumstances that led to reionization of H I (at $z \sim 6 - 15$) and He II (at $z \sim 3$),

(ii) connection of IGM with galaxy (in the form of Galactic winds, various feedback processes),

(iii) the signal from Baryon Acoustic Oscillations (BAO),

---

[1] `http://www.tmt.sorg`

[2] `http://www.gmto.org`

[3] `http://www.gsmt.noao.edu`

[4] `http://www.eso.org/projects/e-elt/`

[5] `http://www.sdss.org`

[6] `http://www.jwst.nasa.gov/`

(iv) 3D matter distribution (tomography) in the IGM,

(v) metal enrichment of the IGM, and

(vi) constraints on cosmological models such as dark energy equation of state, mass of warm dark matter particles.

In order to synchronize with flow of observational data from these surveys, IGM simulations needs to be efficient, flexible and sufficiently accurate to probe large parameter space. *The tools and techniques developed in this thesis provide a starting point to efficiently simulate the IGM.* With large amount of observational data together with efficient simulations, the IGM science will continue to flourish for many years to come.

List of publications

- **Published manuscripts**
  1. *Intergalactic Lyman continuum photon budget in the past 5 billion years*, Gaikwad P., Khaire V., Choudhury T. R. & Srianand R. 2017, MNRAS, 466, 838
  2. *VoIgt profile Parameter Estimation Routine* (VIPER)*: H I photoionization rate at $z < 0.5$*, Gaikwad P., Srianand R., Choudhury T. R. & Khaire V. 2017, MNRAS, 467, 3172
- **Accepted manuscripts**
  1. *Efficient adiabatic hydrodynamical simulations of the high-redshift intergalactic medium*, Gaikwad P., Choudhury T. R., Srianand R. & Khaire V., arXiv:1705.05374

Contents

List of Figures

List of Tables

1 | Introduction

According to our current understanding, the universe was extremely hot, dense and composed of soup of elementary particles (electrons and ions collectively known as baryons) tightly coupled to photons (see Fig. 1.1). At early times the photons were scattered by these electrons to give a spectrum close to that of a blackbody in the early universe (Dicke et al. 1965; Penzias and Wilson 1965). As the universe expanded, the density of the universe dropped and photons were scattered less frequently. The decoupling of photons from baryons allowed the elementary particles to combine and form neutral hydrogen (H I) and helium (He I). As a result, photons eventually could propagate freely to the observer from the surface of "last-scattering". This important milestone in the history of the universe is known as "epoch of recombination".

Today, we see the all pervading radiation from the epoch of recombination in microwave regime of the spectrum and is commonly known as "Cosmic Microwave Background (CMB)" (Dicke et al. 1965; Hinshaw et al. 2013; Komatsu et al. 2011; Mather et al. 1994; Penzias and Wilson 1965; Planck Collaboration et al. 2014, 2016). The temperature map of the CMB suggests that the initial density field was non-uniform with an amplitude of fluctuation $\sim 10^{-5}$ (Komatsu et al. 2011; Mather et al. 1994; Planck Collaboration et al. 2016). After the recombination, the universe cooled below few thousand degree due to adiabatic expansion and entered a 'dark age' in which the universe was filled with CMB photons, dark matter particles and mainly neutral gas (baryons). The slightly overdensed regions in the initial density field started to grow via gravitational instability.

Primordial nucleosynthesis is expected to produce very little heavy elements, hence the baryonic content of the universe during the dark ages mostly consists of H and He (H / He $\sim 1/12$ by number). The recombination process requires the encounter of an electron with a proton. With the expansion of the universe, the rate of recombination is expected to decrease as the number densities of electrons and protons are reduced by recombination. The number of free electrons is frozen when the recombination rate becomes much lower than the expansion rate. These residual free electrons (ionization fraction $X_{\rm e} \sim 10^{-3} - 10^{-5}$) could in principle act as a catalyst for the formation of molecular hydrogen which may be the main source of cooling for formation of first generation luminous objects like population III stars (Bromm and Larson 2004; Ciardi and Ferrara 2005; Loeb and Barkana 2001).

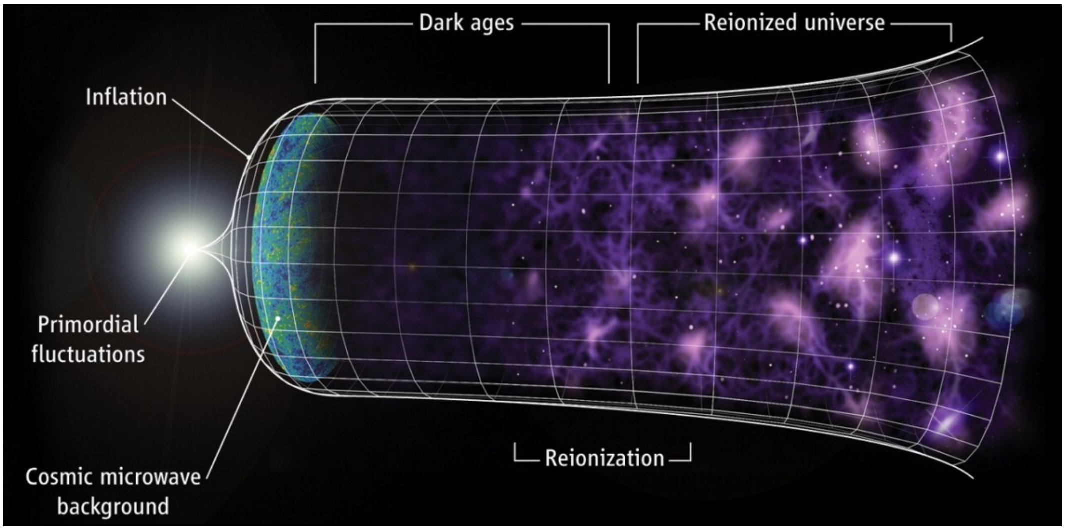

Figure 1.1: Schematic diagram illustrating the evolution of the Universe in the standard bigbang cosmology. The early Universe was extremely hot, dense and composed of soup of elementary particles tightly coupled to photons. The photons decoupled from baryons at epoch of Cosmic Microwave Background (CMB). The primordial fluctuations in the early Universe started to grow via gravitational instabilites during the cosmic dark ages. The cosmic dark age ended when some perturbations in the cosmic density field evolve into the nonlinear regime and collapse to form stars and galaxies. The photons from these young stars and galaxies photo-ionized the IGM. (Image credit: Faucher-Giguère et al. 2008a)

The 'dark age' is expected to end when some perturbations in the cosmic density field evolve into the nonlinear regime and collapse to form objects like stars and galaxies. There was also a significant amount of diffuse and clumpy gas around galaxies that failed to be a part of galaxies. This all-pervading gas that is not associated with galaxies is known as the intergalactic medium (IGM). On the other hand, the diffuse and clumpy gas directly associated with galaxies but not locked in stars is called as "Interstellar Medium" (ISM). The IGM is a crucial ingredient of any theory of galaxy formation and evolution because by definition all the baryons were part of IGM during dark age.

The first generation of stars (population III stars) are expected to be massive, short lived and copiously emitting ultra-violet (UV) photons. The UV photons from these population III stars in early galaxies ionized the surrounding H I, He I to H II, He II respectively (ionization potential of H I, He I is 13.6 eV, 24.6 eV respectively). As the time progressed, these bubbles started to grow and eventually overlapped with each other to fill the entire universe[1]. This is the first major phase transition (after the epoch of recombination) in the universe known as "Epoch of Reionization" (Barkana and Loeb 2001; Choudhury and Ferrara 2006; Ciardi and Ferrara 2005; Furlanetto et al. 2006; Loeb and Barkana 2001; Morales and Wyithe 2010; Zaroubi 2013, EoR). Since the ionization potential of He II is high ($E \sim 54.4$ eV), the energy of UV photons from stars is not sufficient to completely ionize the He II. However, the process of blackhole accretion in Quasi Stellar Objects (QSO) emit the photons at $E \geq 54.4$ eV. Since it takes some time to build the sufficient QSO population in the universe, the epoch of complete ionization of He II is separated from H I reionization. At present, observations suggest that H I and He I are completely ionized at $z \sim 6$ (Becker et al. 2001; Bolton et al. 2011; Fan et al. 2001, 2006; Khaire et al. 2016; Madau et al. 1999; Planck Collaboration et al. 2016; Robertson et al. 2010) whereas He II reionization is completed around $z \sim 3$ (Kriss et al. 2001; Shull et al. 2010, 2004; Worseck et al. 2014).

The temperature of the low density gas in the IGM is expected to increase during the reionization processes because of photoheating (Becker et al. 2011; Hui and Gnedin 1997; Lidz et al. 2010; Schaye et al. 2000). Subsequently at later epochs, the IGM is maintained at a highly ionized state by the ultra-violet

---

[1] Some of the highly overdense regions still remain neutral due to self-shielding.

background (UVB). This UVB is contributed by radiation from QSOs and stars in galaxies (see for example, Fardal et al. 1998; Haardt and Madau 1996; Khaire and Srianand 2015a; Miralda-Escude and Ostriker 1990; Shull et al. 1999). Thus thermal and ionization state of the IGM is closely related to nature of ionizing sources. The IGM also has signatures of various feedback processes associated with the galaxy formation in the form of presence of metals (elements heavier than He), thermal history etc. Furthermore the matter distribution in the IGM traces the cosmic density field on large scale (Bi 1993; Bi and Davidsen 1997; Bi et al. 1992; Choudhury et al. 2001; Croft et al. 1997, 1998; Miralda-Escudé et al. 1996). Thus the IGM is imprinted with rich information of the cosmological initial conditions and the astrophysical processes that occurred after the formation of large scale structures (McDonald et al. 2000, 2005; Penton et al. 2000; Phillips et al. 2001; Seljak et al. 2006; Shull et al. 2012; Storrie-Lombardi et al. 1996; Tegmark et al. 2004; Viel et al. 2009; Viel and Haehnelt 2006; Viel et al. 2006, 2004a,b).

The absorption line features seen in the spectra of luminous distant QSO are one of the most sensitive tools to study the physical conditions prevailing in the IGM (Croft et al. 1998; Gunn and Peterson 1965; Weinberg et al. 1998). The QSO absorption spectroscopy has several advantages (Khare 2013):

- It is highly sensitive to the absorption by even the low density gas and a good probe of the cosmic density field in the mildly non-linear regime.
- It probes thermal, chemical and ionization state of the IGM along the sightline.
- It covers large redshift range and allows to study the redshift evolution of the physical properties of the IGM.
- It allows to study the kinematics of the gas i.e., inflow or outflow of the gas associated with the galaxies.
- If the QSO sightline passes through the environment of galaxy, one can study the connection between ISM and the IGM.

- It also allows one to study the properties of the background QSO itself such as intensity of ionizing radiation from QSO, anisotropies in the density field around QSOs and radiation field of the QSOs etc.

Despite its numerous advantages, QSO absorption spectroscopy suffers from following limitations,

- It depends on the brightness of background source. Faint QSOs are difficult to observe at high spectral resolution. Also one may avoid dusty sightlines that have important clues on galaxy evolution due to the way QSOs are selected based on their colors. This limits the sample size.

- Since QSO absorption spectroscopy probes the IGM along a sightline, it contains information about the medium along the pencil beam. Thus the connection between QSO absorption systems and 3D distribution of cosmological density field (or with galaxies) is not straightforward.

However, with upcoming 30 m class telescopes (Skidmore et al. 2015), it would be possible to obtain spectra from not only faint QSO but with faint galaxies as background source. Since number density of galaxies is much more than QSO, it will also be possible to sample the region densely enough in the longitudinal direction to study the absorption systems in the global context. This will allow one to construct the 3D H I map (known as IGM tomography) from Lyα forest. Hence some of the above mentioned limitations can be circumvent by next generation telescopes.

1.0.1 QSO absorption spectra

The ground state electronic transitions ($1s$ state) of H I are important as they have potential to trace the H I distribution in the IGM. Of particular interest, the transition of electron from $n = 1$ state to higher excited state by absorption of photon, commonly referred as Lyman series transitions, are important. The allowed Lyman series transition of H I: $1s \rightarrow 2p$, $1s \rightarrow 3p$, $1s \rightarrow 4p$, $1s \rightarrow 5p$, are known as Lyα, Lyβ, Lyγ, Lyδ transition respectively. These transitions are resonant transitions. The Lyα transition is the strongest transition in Lyman

Table 1.1: Specifications of H I, He I and He II Lyman series transition

Line	Transition	Rest Wavelength (Å)	Oscillator Strength (f_{ij})	QM Damping constant (Γ in s^{-1})
H I Lyα	$1s \rightarrow 2p$	1215.6701	0.416	6.265×10^8
H I Lyβ	$1s \rightarrow 3p$	1025.7223	0.079	1.897×10^8
H I Lyγ	$1s \rightarrow 4p$	972.5368	0.029	8.127×10^7
H I Lyδ	$1s \rightarrow 5p$	949.7431	0.0139	4.204×10^7
H I Lyω	$1s \rightarrow 6p$	937.8035	0.0078	2.450×10^7
He I Lyα	$1s \rightarrow 2p$	584.334	0.285	1.799×10^9
He I Lyβ	$1s \rightarrow 3p$	537.0296	0.0752	5.663×10^8
He II Lyα	$1s \rightarrow 2p$	303.7822	0.416	1.003×10^{10}
He II Lyβ	$1s \rightarrow 3p$	256.317	0.079	2.667×10^9

series. Table 1.1 summarizes specifications of the Lyman series transition for H I, He I and He II (Verner et al. 1994). The rest wavelength, oscillator strength and quantum mechanical damping constant (Γ in s^{-1}) is given in column 3, 4, and 5 respectively. In this work we denote H I Lyman series transition simply as Lyα, Lyβ etc. Similarly, He II transitions are referred to as He II Lyα, He II Lyβ etc. The absorption features produced by these species are commonly observed in QSO absorption spectra.

Figure 1.2 shows a typical QSO absorption spectrum towards a QSO HE2347-4342 obtained using VLT telescope. The peak in the flux at wavelength $\lambda =$ 4725 Å (emission redshift $z_{\rm em} = 2.885$) corresponds to Lyα emission from the QSO. The absorption features blueward of the emission wavelength are produced by the absorption of Lyα photons by intervening H I gas between observer and QSO. These absorption features are collectively known as Lyα forest. The intrinsic emission from QSO, that is if there was no absorption by the intervening H I, is shown by magenta line (also known as continuum fit to the QSO). The region enclosed in black dashed vertical line in the top and middle panels is enlarged in the middle panel and bottom panel respectively which shows that the Lyα forest consists of many absorption features blended together.

Occasionally, the sightline also passes through the dense H I region. Figure 1.3

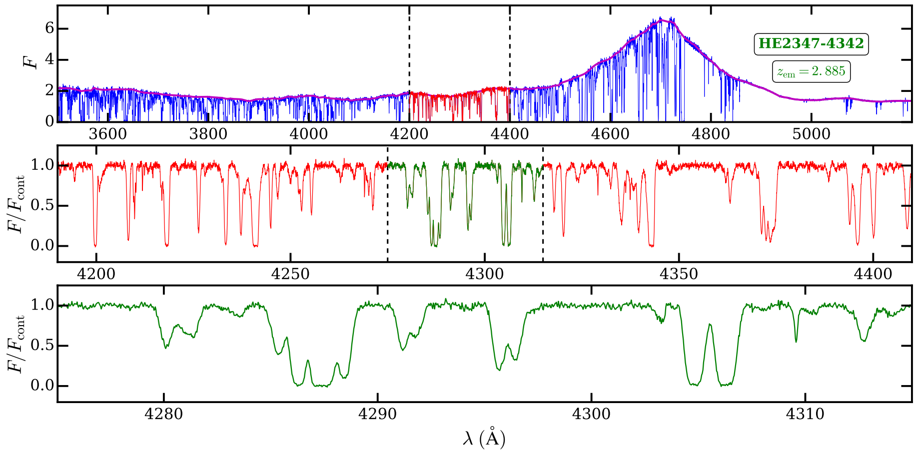

Figure 1.2: A typical example of Lyα forest absorption towards QSO HE2347-4342. The emission redshift of the QSO is $z_{\rm em} = 2.885$. The continuum fit to the QSO is shown by magenta line. The region enclosed in black dashed vertical line in top and middle is enlarged in middle panel and bottom panel respectively. One can see that Lyα forest consists of many absorption features isolated or blended together. Note that we have normalized the flux by continuum in the middle and bottom panel.

shows a typical QSO absorption spectrum towards a QSO PKS0454+039 obtained using Faint Object Spectrograph (FOS) on board Hubble Space Telescope (HST). The redshifted Lyα emission (at $z_{\rm em} = 1.34$) of the QSO is marked by black dashed vertical line. In addition to Lyα forest, this sightline also contains two strong absorbers, (i) damped Lyα absorber (DLA) at $z = 0.86$ and (ii) partial Lyman limit system (LLS) at $z = 1.15$, such that they can absorb the H I ionizing photons. The complete absorption of H I ionizing photons in DLA produces the break at $\lambda = 912 \times (1+z) = 912 \times 1.86 \sim 1700$ Å. In contrast, the column density of H I in partial LLS is not sufficient to completely absorb the H I ionizing photons. Hence a partial break is produced by the partial LLS at $\lambda = 912 \times 2.15 \sim 1950$ Å. The DLA system also produces a range of metal line absorption in the spectrum as an example CIV absorption system is marked in the figure.

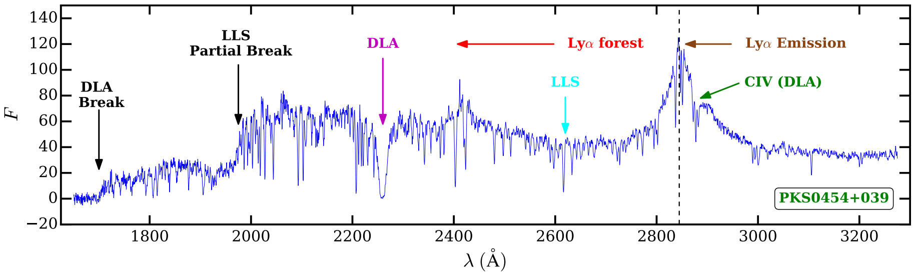

Figure 1.3: Typical example of absorption spectrum towards QSO PKS0454+039 obtained using Faint Object Spectrograph (FOS) on board Hubble Space Telescope (HST). The Lyα emission of the QSO at $\lambda = 2844.66$ Å is indicated by black dashed vertical line which corresponds to the emission redshift $z_{\rm em} = 1.34$. The two strong absorbers at $z = 0.86$ and $z = 1.15$ are examples of damped Lyα absorber (DLA) and partial Lyman limit system (LLS) respectively. The DLA produces the break at $\lambda = 1700$ Å corresponding to complete absorption of H I ionizing photons. Whereas, partial LLS produces the partial break at $\lambda = 1950$ Å because H I column density is not sufficient to absorb all H I ionizing photons. The DLA also produces the metal line (C IV) absorption systems redward of Lyα emission. The common absorption features seen blueward of Lyα emission is known as Lyα forest (figure similar to Charlton and Churchill 2000).

1.0.2 Physical properties of absorber and absorption line profile

Let us assume that the electron in an atom absorbs the photons of energy $h\nu_{12}$ such that two energy levels are sharply defined (i.e. neglecting uncertainty principle for the time being) $E_2 - E_1 = h\nu_{12}$. The line profile function for an observer in the frame of an absorbing atom is given by (refer to Draine 2010; Mo et al. 2010, for more details),

$$\phi(\nu) = \delta(\nu_{obs} - \nu_{12}) \tag{1.1}$$

But according to uncertainty principle each energy level E_i has a spread of energy ΔE_i due to the finite lifetimes of this level. The natural line broadening parameter in the rest frame of atom is given by Lorentzian,

$$\mathcal{L}(\nu_{obs}) = \frac{\pi\, e^2}{m_e\, c}\, f_{12}\, \frac{(\Gamma/4\pi^2)}{(\Gamma/4\pi)^2 + (\nu_{obs} - \nu_{12})^2} \tag{1.2}$$

where Γ is quantum mechanical damping constant listed in Table 1.1.

In a gas at kinetic temperature T, individual atoms are in random motion away from or towards the observer. Consider ensemble of such atom that are constantly in motion e.g. due to thermal motion. The line profile function for an observer in this frame is given by,

$$\phi(\nu) = \delta[\nu_{obs} - \nu_{12}(1 + v/c)] \tag{1.3}$$

if atom is moving towards source $v > 0$, if atom is moving away from source $v < 0$. The velocity distribution along sightline, $f(v)\,dv$, for such a system can be given by Maxwellian distribution as,

$$\begin{aligned} f(v)\,dv &= \left(\frac{2k_BT}{\pi\, m_e}\right)^{1/2} \exp\left[-\frac{m_e v^2}{2k_BT}\right] dv \\ b^2 &= \frac{2k_BT}{m_e} \\ f(v)\,dv &= \frac{1}{\sqrt{\pi}\, b} \exp\left[-\frac{v^2}{b^2}\right] dv \end{aligned} \tag{1.4}$$

In presence of thermal broadening, the net profile is convolution of Lorentzian profile with Gaussian profile due to thermal broadening given by,

$$\begin{aligned} \sigma(\nu_{obs}) &= \frac{I\, c}{\sqrt{\pi} b}\, V(t; B, A) \\ \text{where,} \quad I &= \frac{\pi e^2 f_{12}}{m_e\, c\, \nu_{12}} \end{aligned} \tag{1.5}$$

For Lyα absorption, $I = I_\alpha = 4.45 \times 10^{-18}\ \ \text{cm}^2$. The quantity $V(t; B, A)$ is Voigt function given by,

$$V(t; B, A) = \frac{A}{\pi} \int_{-\infty}^{\infty} \frac{\exp(-t^2)}{A^2 + (B - t)^2}\ dt \tag{1.6}$$

with,

$$\begin{aligned} A &= \frac{c\Gamma}{4\pi\nu_{12} b} \\ B &= \frac{c}{b}\, \frac{\nu - \nu_{12}}{\nu_{12}} \quad . \end{aligned} \tag{1.7}$$

The Voigt function can also be expressed as a real part of “Faddeeva function” $W(x + i\, y)$. The Voigt function is easy to implement using *wofz* function in *python*'s *scipy* package. The optical depth (τ) of the Lyα absorption is related to absorption cross-section as,

$$\tau = \sigma(\nu_{obs}) \int n_{\rm HI} \; dl = \sigma(\nu_{obs}) \, {\rm N_{HI}} \tag{1.8}$$

where $\rm N_{HI}$ is H I column density and we have assumed that the absorption cross-section does not depend on the path length. The transmitted Lyα flux $F = \exp(-\tau)$ is an observable quantity.

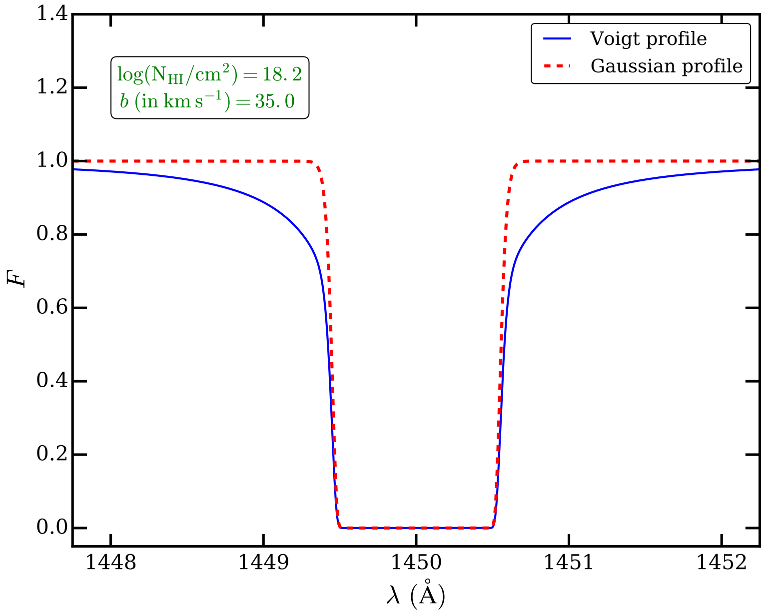

Figure 1.4: Illustration of absorption line profile for H I Lyα line ($\log \rm N_{HI} = 18.2$ and $b = 35$ km s^{-1}). The Voigt profile (blue solid curve) is convolution of Gaussian profile (due to thermal motion of atoms) and Lorentzian (due to quantum mechanical finite lifetime of energy state). Figure shows that the center of the absorption line profile is dominated by thermal broadening (red dashed line). The difference (mainly in the wings) between the Voigt profile and Gaussian is due to the Lorentzian profile.

Fig. 1.4 shows contribution of Gaussian and Lorentzian to the Lyα flux calculated using Voigt function. The center of the absorption line profile is mainly dominated by Gaussian while the wings are dominated by the Lorentzian profile. The Lyα forest, LLS and DLA discussed in the previous section are usually defined as the absorption systems with $\rm N_{HI}/cm^{-2} < 10^{17.2}$, $10^{17.2} \leq \rm N_{HI}/cm^{-2} <$

$10^{20.3}$ and $10^{20.3} \leq \mathrm{N_{HI}/cm^{-2}}$ respectively. But as evident from the Fig. 1.2, the Lyα forest is composed of many absorption features. Naturally the question is how are these continuous absorption features produced in QSO absorption spectra?

1.0.3 Gunn-Peterson Effect

Consider a photon of frequency ν_e emitted from QSO at redshift z_e. The observed frequency of the photon (if not absorbed) at earth is $\nu_0 = \nu_e/(1+z_e)$. This photon passes through IGM and is absorbed by H I at redshift z. The Lyα absorption cross-section for this photon is given by,

$$\sigma[\nu_0(1+z)] = \frac{\pi\, e^2\, f}{m_e\, c}\, \phi[\nu_0(1+z)] \tag{1.9}$$

where e, m_e, c is charge on electron, mass of electron and speed of light respectively. $f = 0.4162$ is upward oscillator strength for the transition. $\phi[\nu_0(1+z)]$ is the normalized profile function ($\int \phi(\nu)\, d\nu = 1$) and is expressed in the form of Voigt function Eq. 1.6. Assuming that the profile function is sharply peaked at Lyα resonant frequency we can write,

$$\sigma[\nu_0(1+z)] = \frac{\pi\, e^2\, f}{m_e\, c}\, g[\nu_0(1+z) - \nu_\alpha] \tag{1.10}$$

where ν_α is laboratory frequency of the Lyα transition. The Lyα optical depth at observed frequency is given by,

$$\tau = \int_0^{z_e} \sigma[\nu_0(1+z)]\, n_{\mathrm{HI}}(z)\, \frac{dl}{dz} \tag{1.11}$$

where $dl/dz = cH^{-1}(z)/(1+z)$ is the line element in Friedmann-Lemaître-Robertson-Walker (FLRW) metric, $H(z)$ is Hubble parameter at redshift z (see Appendix 1.7.1 for definitions). The Lyα optical depth can then be written as,

$$\tau = \frac{\pi\, e^2\, f\, c}{m_e\, c} \int_0^{z_e} \left\{ \frac{n_{\mathrm{HI}}(z)}{\nu_0(1+z)\, H(z)} \right\} g[\nu_0(1+z) - \nu_\alpha]\, d[(1+z)\nu_0]$$

The function g is strongly peaked at zero, its width depend on kinetic temperature of the gas. Even if we take temperature of $10^5 K$, the corresponding velocity width is 60 $\mathrm{km\,s^{-1}}$which is small as compared to redshift range. Thus we can take the term in braces out of the integral evaluated at $(1+z) = \nu_\alpha/\nu_0$. The integral would be 1 because it is like dirac delta function. The Lyα optical depth under this approximation (known as Gunn-Peterson optical depth) is given by,

$$\tau_{GP} = \frac{\pi\, e^2\, f}{m_e\, c}\, \frac{c\, n_{\rm HI}(z)}{\nu_\alpha\, H(z)} = I_\alpha\, \frac{c\, n_{\rm HI}(z)}{H(z)} \tag{1.12}$$

The above equation can be rearranged to give H I neutral fraction at redshift z as,

$$\begin{aligned} f_{\rm HI} &= \frac{n_{\rm HI}(z)}{\overline{n}_{\rm H}(z)} = \frac{H(z)\,\tau_{GP}}{c\, I_\alpha\, \overline{n}_{\rm H}(z)} \\ f_{\rm HI} &= \frac{3.241h \times 10^{-18}\, E(z)\, \tau_{GP}}{3\times10^{10} \times 4.45\times10^{-18} \times 2\times10^{-7} \times (1+z)^3} \\ f_{\rm HI} &= 1.214\times10^{-4}\, \tau_{GP}\, h\, E(z)\, (1+z)^{-3} \\ \tau_{GP} &= 8.237\times10^{3}\, f_{\rm HI}\, h^{-1}\, E(z)^{-1}\, (1+z)^3 \quad . \end{aligned} \tag{1.13}$$

where $E(z)$ is defined in Appendix 1.7.1 (see Eq. 1.50). In a homogeneous and expanding Universe, the above equation is applicable to all the wavelengths blueward of QSO Lyα emission redshift. If we assume that the universe is completely neutral between $z = 0$ to 6, the Gunn-Peterson optical depth is $\tau_{GP} \sim 9 \times 10^3$ or Lyα flux $F = e^{-\tau_{GP}} \sim 0$. Thus one should see an absorption trough blueward of QSO emission redshift at $z < 6$. This effect is known as “Gunn-Peterson effect”. However, as shown in Fig. 1.2 and Fig. 1.3, the transmitted Lyα flux is non-zero at $z < 6$. The observations of Lyα forest suggests that the effective optical depth ($\tau_{\rm eff} = -\ln\langle e^{-\tau}\rangle$, angle brackets indicate average along different sightlines) strongly evolves with redshift as (Fan et al. 2006),

$$\tau_{\rm eff} = (0.85 \pm 0.06)\, \times \left(\frac{1+z}{5}\right)^{4.3\pm0.3} \quad \text{for } z \leq 5.5 \tag{1.14}$$

This suggests that the fraction of neutral hydrogen in the Universe must be smaller than 10^{-4} at $z < 6$. In other words, the Universe is highly ionized at $z < 6$. At present, we think that the Universe is predominantly ionized by photons coming from young stars, galaxies and QSO.

1.1 Models of Lyα forest

Since the discovery of Lyα forest over the large redshift range, many models have been proposed for the existence of Lyα forest. Any model of Lyα forest should address the question: how these systems are sustained over the history of the universe? The early models treat Lyα forest to be consists of discrete clouds of gas either supported by pressure in hot medium or gravitationally confined in dark matter halo (for more details see Mo et al. 2010).

- **Pressure confinement model:** The observations of Lyα forest at $z \geq 2$ suggests that the temperature of IGM is $\sim 10^4$ K (from doppler b parameter $\sim 13\ \mathrm{km\,s^{-1}}$). One postulation was that Lyα clouds are dense, cooler clouds confined by pressure in hot, ambient and tenuous medium all pervading in IGM. This model was partially motivated by the observations of hot X-ray background at that time (Ikeuchi and Ostriker 1986; Ostriker and Cowie 1981; Ostriker and Ikeuchi 1983; Sargent et al. 1980). Absorption lines with different column density in Lyα forest corresponds to range of cloud masses. In this model, the massive clouds are Jeans-unstable whereas very small clouds are evaporated by heat conduction. However, pressure confinement model has three main problems (i) recent observations suggests that the X-ray background is contributed by discrete source as compared to diffuse hot medium as required for this model, (ii) the column density of H I varies from $10^{13} \leq \mathrm{N_{HI}/cm^{-2}} \leq 10^{16}$. Such a wide range in $\mathrm{N_{HI}}$, requires 9 orders of magnitude cloud mass range or a factor ~ 100 change in pressure and (iii) the correlated Lyα absorptions in sightline towards double QSOs shows that absorber sizes are over 100 kpc which constrained their densities (n_{H}) to be very low, resulting in a high ionization fraction from photoionization. These effects are difficult to realize in pressure confinement model. Although failed to explain Lyα forest, this model is important because cold, dense clouds in the halos of galaxies may have formed via local-instabilities and in pressure equilibrium with hot gas in halos (Mo and Miralda-Escude 1996).
- **Gravitational confinement model:** Since the large scale structures in the Universe are formed via gravitational instability around initial density

peaks. Melott (1980) and Black (1981) proposed that the Lyα clouds are confined by gravity. However the Lyα clouds could not be self-gravitating because such clouds are unstable in the presence of effective cooling. The model suggested that Lyα clouds are hosted by 'mini halos' where the potential is not enough to cool the photo-ionized gas ($T \sim 10^4$ K, $b \sim 13$ km s^{-1}) but sufficient to prevent the photo-ionized gas from escaping (circular velocity of the mini-halos, v_c, are ~ 30 km s^{-1}, Ikeuchi 1986; Rees 1986). This model was successful in explaining wide range in N$_{\rm HI}$ due to large density gradient, evolution in number density of absorber and shape of H I column density distribution function. Although successful, the mini halo model is an oversimplification of hierarchical clustering where collapse is highly non-spherical forming sheets, filaments and nodes.

- **Density field fluctuations model:** In hierarchical structure formation scenario, in which low mass object collapse first and high mass object later, the Universe at any given epoch consists of dark matter halos, sheets, filaments and structures with density in mildly non-linear regime. Any sightline passing through these structures can produce QSO absorption systems if neutral gas column density is sufficiently high (Bi 1993; Bi and Davidsen 1997; Bi et al. 1992; Choudhury et al. 2001; Croft et al. 1997). Thus instead of treating Lyα absorbers as discrete clouds, this model suggests fluctuations in cosmic density field as origin of Lyα forest. This model has been very successful because physics at these densities is relatively simple, model prediction can be worked out in details and are consistent with observations. In the next two section, we are going to derive the basic formalism of this model.

1.2 Lyα forest as fluctuations in Optical depth

In hierarchical structure formation model, the Lyα forest (see Fig. 1.3) is shown to arise from fluctuations in the optical depth along a sightline (Bi 1993; Bi and Davidsen 1997; Bi et al. 1992; Choudhury et al. 2001; Croft et al. 1997, 1998; Meiksin 2009; Miralda-Escudé et al. 1996). To elaborate this, consider a sightline passing through the cosmic density field. Let x be the comoving co-ordinate along the sightline. The optical depth at an observed frequency ν_{obs} produced by the Lyα absorption from the intervening H I gas can be written as,

$$\tau(\nu_{\rm obs}) = \int_0^\infty n_{\rm HI}(x)\, \sigma_{Ly\alpha}(\nu)\, \frac{dx}{1+z} \tag{1.15}$$

$$\nu = \nu_{\rm obs}(1+z) + \nu_{\rm obs}(1+z)\frac{v_r(x)}{c} \tag{1.16}$$

where $n_{\rm HI}(x)$ is number density of neutral hydrogen, $\sigma_{Ly\alpha}(\nu)$ is Lyα absorption cross-section at frequency ν given by Eq. 1.5 and z is redshift of the Lyα absorber which is related to the comoving co-ordinate. The first term on the right hand side of Eq. 1.16 corresponds to Hubble expansion where as second term represents Doppler shift in frequency due to peculiar velocity, $v_r(x)$, of Lyα absorber.

The IGM can be assumed to be in photoionization equilibrium with UVB (Choudhury et al. 2001; Theuns et al. 1998b; Weinberg et al. 1998). This is a good approximation as time scales of radiative processes are smaller as compared to dynamical and Hubble time scales (see Appendix 1.7.2 and 1.7.3). Under this assumption we can write,

$$n_{\rm HI} = \frac{n_{\rm e}\, n_{\rm p}\, \alpha(T)}{\Gamma_{\rm HI}} = \frac{\mu_{\rm e}\, n_{\rm H}^2\, \alpha(T)}{\Gamma_{\rm HI}} \tag{1.17}$$

where $\Gamma_{\rm HI}$ is H I photoionization rate, $n_{\rm p}, n_{\rm e}$ number density of proton and electron respectively, $\alpha(T)$ is recombination rate coefficient that in general depends on temperature and $\mu_e = (4-2Y)/(4-3Y)$ is mean molecular weight of the electrons with $Y = 0.24$ is helium fraction by mass. For the second equality, we assumed that the number of free protons are contributed mainly by H I i.e., $n_{\rm p} \approx n_{\rm H}$ and is given by,

$$n_{\rm H} = 1.55 \times 10^{-7}\ {\rm cm}^{-3}\ (1+z)^3\ \Delta \left[\frac{\Omega_b\, h^2}{0.2208} \times \frac{1.2195}{\mu_b} \times \frac{1-Y}{1-0.24}\right] \tag{1.18}$$

where we used Eq. 1.48, cosmological parameters given in Table 1.3, $\mu_b = 4/(4-3Y)$ is mean molecular weight of baryons.

Fig. 1.5 summarizes the dependence of cosmological and astrophysical parameters on the modeling of the Lyα optical depth. It is evident from Fig. 1.5, Eq. 1.15 to Eq.1.18 and Eq. 1.5 to Eq. 1.7 that the Lyα optical depth can be calculated if we know the line of sight (i) overdensity field (Δ), (ii) peculiar velocity field (v_r), (iii) temperature field (T) and (iv) ionization correction. Note that the transmitted Lyα flux $F = \exp(-\tau)$ is a observable quantity. Thus modeling of the Lyα forest involves the estimation of these fields from the simulation.

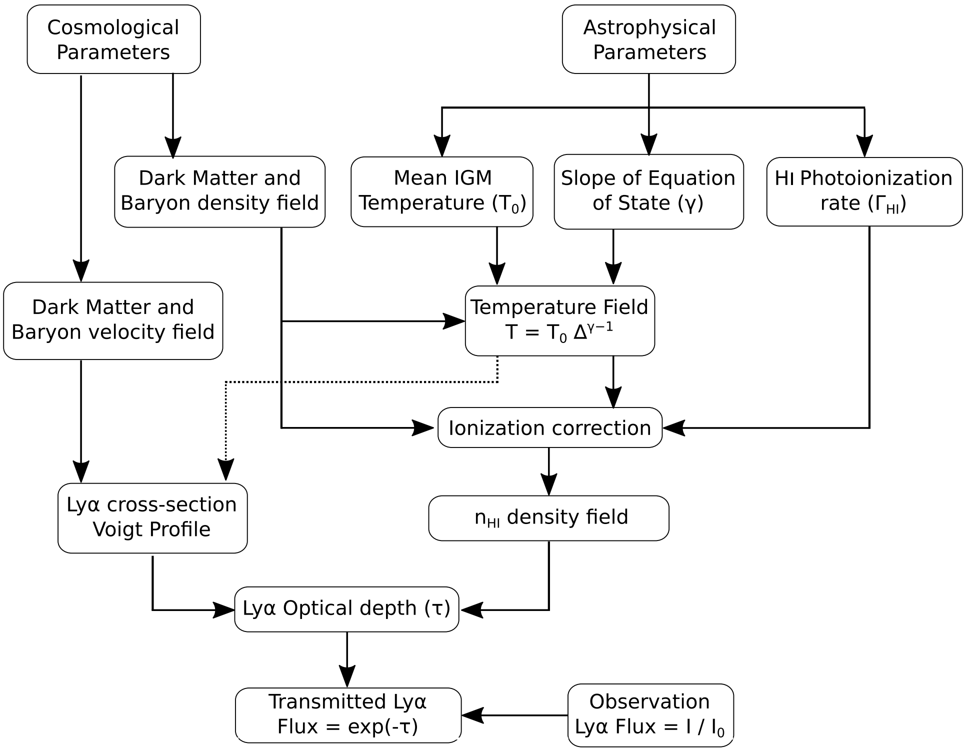

Figure 1.5: Flow chart showing the dependence of cosmological and astrophysical parameters on modeling of the Lyα optical depth.

1.2.1 Temperature evolution of the IGM

The estimation of the IGM temperature is crucial in accurate modeling of the Lyα forest. The IGM temperature has 3 important effects on transmitted Lyα flux.

1. The higher temperature of the IGM broadens the absorption line profile through doppler broadening (see Eq. 1.5).

2. The recombination rate coefficient ($\alpha \propto T^{-0.726}$, for Case-A which is good for optically thin gas) depends on the IGM temperature (see Eq. 1.17). The larger the temperature fewer the recombinations and hence highly ionized IGM.

3. The temperature of the IGM also affects the shape of the density field due to finite pressure of the gas. In general, density field in hot IGM is smoother as compared to cold IGM (Gnedin and Hui 1998; Kulkarni et al. 2015; Peeples et al. 2010).

The temperature evolution of the IGM is governed by various cosmological and astrophysical processes in the Universe. The basic temperature evolution equation of the IGM is given by (Hui and Gnedin 1997),

$$\frac{dT}{dt} = -2HT + \frac{2T}{3\Delta}\frac{d\Delta}{dt} + \frac{dT_{shock}}{dt} + \frac{T}{\sum_i X_i}\frac{d\sum_i X_i}{dt} + \frac{2}{3\,k_B\,n_b}\frac{dQ}{dt} \tag{1.19}$$

The first three terms on the right hand side represent, respectively, the rate of cooling due to Hubble expansion, adiabatic heating and/or cooling arising from the evolution of the densities of gas particles and the change in temperature because of shocks due to structure formation. The fourth term on the right hand side represents the change in internal energy per particle arising from the change in the number of particles. The last term accounts for the radiative heating and cooling processes. The last two terms in Eq.1.19 depends on the ionization fraction and requires to solve the ionization evolution equation of the IGM.

1.2.2 Ionization evolution of the IGM

Consider a primordial composition of IGM gas exposed to extragalactic UVB (see Appendix 1.7.2 for more details) which is assumed to originate from QSO and/or young galaxies. The main heating source for the gas is photo-ionization in which excess energy of photons is carried away by electrons as kinetic energy. The primary cooling processes relevant for IGM are two body interaction in which gas loses its energy by radiation. It is safe to ignore the three body interactions in IGM as the densities of gas under consideration are too low. The interaction of free electrons with CMB photons, collision between atoms and radiation from accelerated electrons can allow the gas to cool down by Compton, collisional and Bremsstrahlung cooling respectively. The cooling rates for various radiative processes can be calculated for a given temperature. The cooling rates are usually expressed in terms of cooling function defined as (Katz et al. 1996; Sutherland and Dopita 1993; Weinberg et al. 1997),

$$\Lambda(T) \equiv \frac{\mathscr{C}}{n_{\mathrm{H}}^2} \tag{1.20}$$

where $\mathscr{C}$ is the total cooling rate per unit volume and n_{H} is number density of total H. The unit of $\mathscr{C}$ is ergs cm^{-3} s^{-1} whereas, unit of $\Lambda(T)$ is ergs cm^3 s^{-1}. Note that for optically thin gas, $\Lambda(T)$ is independent of gas density. The radiative cooling function can be calculated by summing the cooling rates given in Appendix 1.7.4 (see Table 1.4, Katz et al. 1996; Theuns et al. 1998b).

$$\mathscr{C} = \sum_{i=1}^{11} c_i(T, z, \mathrm{X}) \tag{1.21}$$

In order to obtain the cooling rates given in Table 1.4, we need to calculate ionic abundances of hydrogen and helium. These abundances can be obtained by solving the ionization evolution equations (Theuns et al. 1998b),

$$\begin{aligned} \frac{d\mathrm{H\textsc{ii}}}{dt} &= \alpha_{\mathrm{HII}}\, n_e\, \mathrm{H\textsc{ii}} - \mathrm{H\textsc{i}}\,(\Gamma_{\gamma\mathrm{H\textsc{i}}} + \Gamma_{e\mathrm{H\textsc{i}}}\, n_e) \\ \frac{d\mathrm{He\textsc{i}}}{dt} &= \alpha_{\mathrm{HeII}}\, n_e\, \mathrm{He\textsc{ii}} - \mathrm{He\textsc{i}}\,(\Gamma_{\gamma\mathrm{He\textsc{i}}} + \Gamma_{e\mathrm{He\textsc{i}}}\, n_e) \\ \frac{d\mathrm{He\textsc{iii}}}{dt} &= -\alpha_{\mathrm{HeIII}}\, n_e\, \mathrm{He\textsc{iii}} + \mathrm{He\textsc{ii}}\,(\Gamma_{\gamma\mathrm{He\textsc{ii}}} + \Gamma_{e\mathrm{He\textsc{ii}}}\, n_e) \end{aligned} \tag{1.22}$$

where Γ_γ, Γ_{eX} and $\alpha_{\mathrm{X}}(T)$ is photoionization rate, collisional ionization rate and recombination rate coefficient for specie $\mathrm{X} \equiv [\mathrm{H\textsc{i}}, \mathrm{H\textsc{ii}}, \mathrm{He\textsc{i}}, \mathrm{He\textsc{ii}}, \mathrm{He\textsc{iii}}]$ and symbol $x \equiv [\mathrm{H\textsc{i}}, \mathrm{H\textsc{ii}}, \mathrm{He\textsc{i}}, \mathrm{He\textsc{ii}}, \mathrm{He\textsc{iii}}]$ in above equation corresponds to $x = n_x/n_{\mathrm{H}}$ where n_x is number density of specie x. The photoionization rate is usually obtained from the UVB model that has contribution from QSO and/or young galaxies (see Appendix 1.7.2 for more details). The collisional ionization rates and recombination rates are given in Table 1.5 (Theuns et al. 1998b).

The above equations are supplemented with closure conditions,

$$\begin{aligned} &\mathrm{H\textsc{i}} + \mathrm{H\textsc{ii}} = 1 \\ &\mathrm{He\textsc{i}} + \mathrm{He\textsc{ii}} + \mathrm{He\textsc{iii}} = y \\ &\mathrm{H\textsc{ii}} + \mathrm{He\textsc{ii}} + 2\mathrm{He\textsc{iii}} = e \end{aligned} \tag{1.23}$$

where $y = Y\, m_{\rm H}/[m_{\rm He}\,(1-Y)]$ is He abundance by number and $m_{\rm H}, m_{\rm He}$ is the atomic mass of H, He respectively. Fig. 1.6 shows the cooling rates as a function temperature for primordial composition of gas under the assumption of collisional equilibrium i.e., we set $\Gamma_{\gamma\rm X}$ in Eq. 1.22 to zero. The cooling curve is dominated by collisional excitation at $T < 3 \times 10^5$ K whereas at $T > 10^6$ K the Bremsstrahlung cooling is dominant. This curve changes with metallicity (Sutherland and Dopita 1993; Wiersma et al. 2009).

In the presence of photoionization, Eq. 1.22 and Eq. 1.23 can be easily solved with initial guess value $n_{\rm e} \approx n_{\rm H}$. The relative abundances of the species depends on density, temperature and UVB model. In addition to this photoionization process injects thermal energy into gas, since free electrons carry excess energy. The total excess energy per unit volume per unit time (ergs cm^{-3} s^{-1}) injected

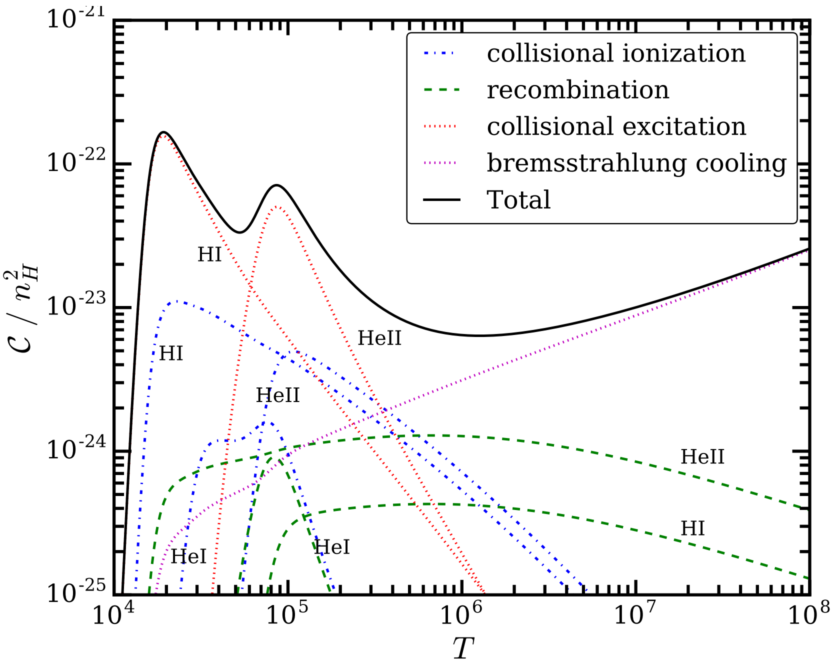

Figure 1.6: Cooling rate as a function of temperature for primordial (H and He) composition of gas in collisional equilibrium. The black solid curve represents total cooling rate. The total cooling rate is dominated by collisional excitation rates at low temperatures ($T < 3 \times 10^5$ K) and by free-free emission at high temperatures ($T > 10^6$ K).

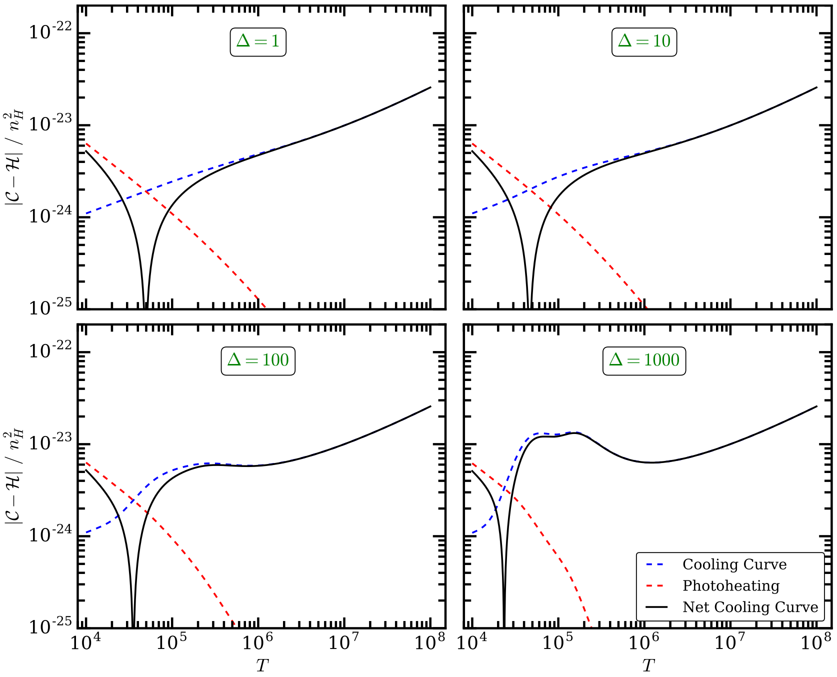

Figure 1.7: Net cooling rate as a function of temperature for primordial (H and He) composition of gas in ionization equilibrium with HM12 UVB. Results are shown for different overdensities as indicated in each subpanel. The blue and red dashed line indicates cooling rate and the rate of heating by photoionization respectively. The absolute net cooling rate is is shown by black solid curve.

in electrons from photoionization is given by,

$$\mathscr{H} = (\text{H\,I}\ \epsilon_{\gamma\text{HI}} \ + \ \text{He\,I}\ \epsilon_{\gamma\text{HeI}} \ + \ \text{He\,II}\ \epsilon_{\gamma\text{HeII}})/n_{\text{H}} \tag{1.24}$$

where $\epsilon_{\gamma\text{X}}$ is photo-heating rates for species X. In presence of photo-heating the net cooling rate ($|\mathscr{C} - \mathscr{H}|/n_{\text{H}}^2$) is no longer independent of density. Fig. 1.7 shows the net cooling rate (Λ_{net}) as a function of temperature for primordial composition of gas in ionization equilibrium with Haardt and Madau (2012, hereafter HM12) UVB for different overdensity given in each sub-panel. At $T < 10^4$ K the photo-heating of the gas is efficient. The net cooling rate follows the total photo-heating rates of the gas. Whereas at $T > 10^5$ K the net cooling rate follows the cooling rates. The residual heating due to photoionization is sub-dominant at $T > 10^5$ K.

The net heat gain (or negative heat loss, as given in 5th term of Eq. 1.19) per unit volume by the gas particles from the surrounding radiation field is given by,

$$\frac{dQ}{dt} = \mathscr{H} - \mathscr{C} = \Lambda_{\rm net}(T)\, n_{\rm H}^2 \tag{1.25}$$

Thus it is straightforward now to calculate the 4th and 5th term in the Eq. 1.19 and to solve for the temperature of the gas in IGM.

1.2.3 Temperature density ($T - \Delta$) relation in IGM

In general the temperature evolution equation (Eq. 1.19) is non-trivial to solve analytically. However, it is still possible to get the physical insight of the low density IGM temperature by equating (heating - cooling) time or simply net heating time ($t_{\rm heat}$) with Hubble time ($t_{\rm Hubble}$ Theuns et al. 1998b). The net heating time for gas at a temperature T is given by (see Appendix 1.7.5 and Theuns et al. (1998b) for more details),

$$\begin{aligned} t_{\rm heat} &= \frac{3\, k_{\rm B}\, T}{2\mu}\, \frac{m_{\rm H}}{\rho\, (1-Y)^2\, (\mathscr{H} - \mathscr{C})} \\ t_{\rm heat} &= \frac{K_1\, T}{\Delta\, (K_2\, T^{-\beta} - K_3\, T\, \Delta^{-1})} \end{aligned} \tag{1.26}$$

where $k_{\rm B}$, μ and ρ is Boltzmann constant, mean molecular weight and density of the gas respectively. The factors K_1, K_2, K_3 depends on the redshift, cosmological and astrophysical parameter as given in Appendix 1.7.5. The recombination rate (for X $\equiv$ H II, He III) depends on temperature as $\alpha_{\rm X} \propto T^{-\beta}$ (assuming $\beta \sim 0.7$ which is close to Case-A coefficient, $\beta = 0.726$, for optically thin gas). Equating net heating time with Hubble time (i.e., $t_{\rm heat} \approx t_{\rm Hubble}$) and rearranging the terms gives,

$$T = \left[\frac{K_2\, t_{\rm Hubble}}{K_1 + K_3\, t_{\rm Hubble}}\right]^{\frac{1}{1+\beta}}\ \Delta^{\frac{1}{1+\beta}} \tag{1.27}$$

$$T = T_0\, \Delta^{\gamma-1} \tag{1.28}$$

where T_0 is mean IGM temperature i.e., temperature of the IGM for mean cosmic density ($\Delta = 1$) and γ is known as slope of $T - \Delta$ relation. Parameters T_0 and γ are usually known as thermal history parameters. Thus it is clear from Eq. 1.28 that the low density IGM follows a power law $T - \Delta$ relation. Typically at $z \sim 6$ one can show that $T_0 \sim 10000$ K and $\gamma \sim 1.6$.

1.2.4 Effect of reionization on thermal history parameters: Qualitative approach

The temperature of the IGM during reionization process is determined by the neutral fractions ($n_{\rm HI}$, $n_{\rm HeI}$ during H I reionization and $n_{\rm HeII}$ during He II reionization). In particular the excess energy per unit volume per unit time injected in electrons due to photoionization depends on $n_{\rm HI}$, $n_{\rm HeI}$ and $n_{\rm HeII}$ (see Eq. 1.24). It is relatively simple to calculate the energy injected in the IGM by H I photoionization as follows (Hui and Gnedin 1997),

$$E_{\rm J} \equiv \frac{\int\limits_{\nu_{\rm HI}}^{\infty} 4\pi\, J_\nu\, \sigma_{\rm HI}\, (h_p\nu - h_p\nu_{\rm HI})\, \frac{d\nu}{h_p\nu}}{\int\limits_{\nu_{\rm HI}}^{\infty} 4\pi\, J_\nu\, \sigma_{\rm HI}\, \frac{d\nu}{h_p\nu}} \tag{1.29}$$

where J_ν, $\sigma_{\rm HI}$, $\nu_{\rm HI}$ and $h_{\rm p}$ is the specific intensity of the surrounding radiation field, photoionization cross-section of H I, frequency of H I photoionization and Planck constant respectively. Using $J_\nu = J_0\, \nu^{-\alpha_\nu}$ and $\sigma_{\rm HI} = \sigma_0\, \nu^{-3}$, it is straightforward to show that,

$$E_{\rm J} = \frac{h_p\nu_{\rm HI}}{3} \tag{1.30}$$

The corresponding temperature of the IGM (assuming $\alpha_\nu \sim 1$) due to photoheating from H I photoionization is,

$$\begin{aligned} T_{\rm R} &= \frac{E_{\rm J}}{3k_B} = \frac{h_p\nu_{\rm HI}}{9k_B} \\ T_{\rm R} &= \frac{13.6 \times 1.602 \times 10^{-19}}{9 \times 1.38 \times 10^{-23}} \approx 17500 \text{ K}. \end{aligned} \tag{1.31}$$

For a given uncertainty in $\alpha_\nu = 1.0$ to 1.41, the IGM temperature varies from $T \sim 15000$ K to 20000 K.

Prior to H I reionization the temperature of the IGM is close to ~ 100 K (CMB temperature). Although the IGM is heated to ~ 20000 K during H I reionization, the temperature of the IGM at $z \sim 6$ (where H I and He I reionization is completed) is ~ 10000 K due to cooling from Hubble expansion and inverse Compton scattering (see previous section). The He II reionization is quite separated ($z \sim 3-4$) from H I, He I reionization as QSO population at $z \sim 3-4$ is sufficient to ionize He II ($E \sim 54.4$ eV). Note that prior to He II reionization

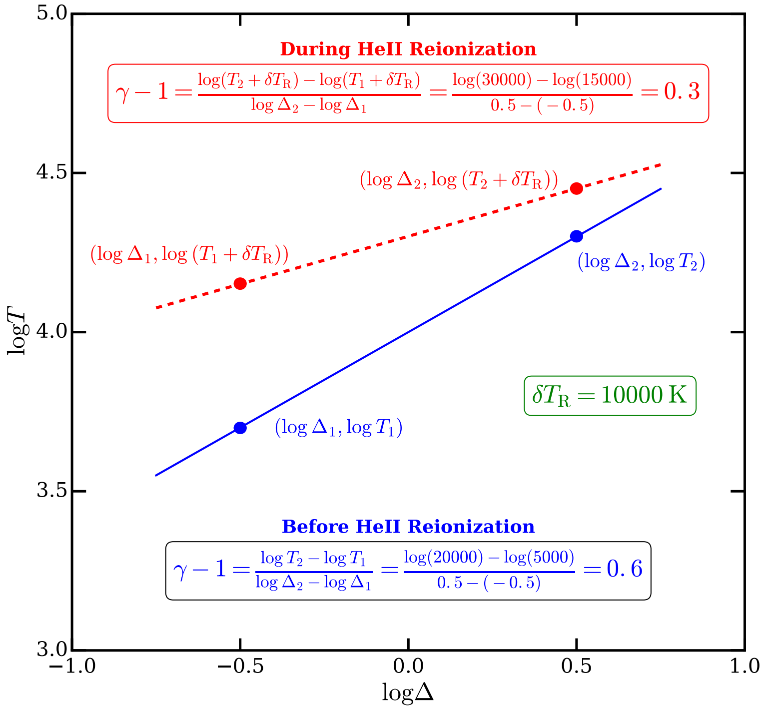

Figure 1.8: Schematic diagram showing the effect of residual photoheating during He II reionization on thermal history parameters T_0 and γ. Due to the density independent heating of the IGM during the reionization, the mean IGM temperature increases (by $\delta T_{\rm R} \sim 10000$ K) and $T-\Delta$ relation becomes flat ($\gamma \sim 1.3$).

the IGM temperature is ~ 10000 K and during He II reionization (similar to H I reionization) the IGM may be heated to ~ 20000 K. It is interesting to see the effect of such a increase in temperature ($\delta T_{\rm R} \sim 10000$ K) during He II reionization on the slope of $T-\Delta$ relation.

Let us assume that the photoheating during He II reionization is independent of density. In this scenario, in addition to rise in temperature of the IGM, the $T-\Delta$ relation becomes flatter ($\gamma \sim 1.3$ Puchwein et al. 2015). This is illustrated in Fig. 1.8 where blue solid line shows the $T-\Delta$ relation before He II reionization. Consider two gas particles on this line with ($\log\Delta_1$, $\log T_1$) and ($\log\Delta_2$, $\log T_2$). The slope of the $T-\Delta$ relation in this case is given by,

$$\gamma - 1 = \frac{\log T_2 - \log T_1}{\log\Delta_2 - \log\Delta_1} = \frac{\log 20000 - \log 5000}{0.5-(-0.5)} = 0.602 \tag{1.32}$$

During He II reionization the temperature of both the particle is increased by the same amount $\delta T_{\rm R} = 10000$ K[1]. These two particles move to new location $(\log \Delta_1, \log(T_1 + \delta T_{\rm R}))$ and $(\log \Delta_2, \log(T_2 + \delta T_{\rm R}))$ and attains new $T - \Delta$ relation shown by red dashed line. The slope of new $T - \Delta$ relation is given by,

$$\gamma - 1 = \frac{\log(T_2 + \delta T_{\rm R}) - \log(T_1 + \delta T_{\rm R})}{\log \Delta_2 - \log \Delta_1} = \frac{\log 30000 - \log 15000}{0.5 - (-0.5)} = 0.301 \quad (1.33)$$

Thus for a density independent photoheating during He II reionization the amplitude of $T - \Delta$ relation increases where as slope decreases (i.e. $T - \Delta$ relation becomes relatively flatter).

Since parameters such as H I photo-ionization rate ($\Gamma_{\rm HI}$) and thermal history parameters (T_0, γ) depends on astrophysical processes dominated during reionization, it is important to measure these parameters from observations to constrain the reionization history.

1.2.5 Fluctuating Gunn-Peterson Approximation (FGPA)

We can also qualitatively derive the analytical expression for Lyα optical depth under the fluctuating Gunn-Peterson approximation (FGPA, Gunn and Peterson 1965; Weinberg et al. 1998). Applying ionization equilibrium (Eq. 1.17) and using power law $T - \Delta$ (Eq. 1.28) in expression for Gunn-Peterson optical depth (Eq. 1.12) we get,

$$\tau_{GP} \approx \frac{\pi\, e^2\, f\, \lambda_\alpha\, \alpha_0\, \mu_{\rm e}\, \overline{n}^2_{\rm H,0}}{m_e\, c\, H_0}\; \frac{\Delta^{2-0.7(\gamma-1)}\, (1+z)^{4.5}\, T_0^{-0.7}}{\Gamma_{\rm HI}\, \Omega_{\rm m}^{0.5}} \quad (1.34)$$

$$\tau_{GP} \propto (1+z)^{4.5}\, \Omega_{\rm m}^{-0.5}\, \Gamma_{\rm HI}^{-1}\, T_0^{-0.7}\, \Delta^{2-0.7(\gamma-1)} \quad (1.35)$$

It is clear from the above expression that the Lyα optical depth strongly evolves with redshift, it depends on cosmological parameters such as hubble parameter $H(z)$, cosmic matter density parameter ($\Omega_{\rm m}$), H I photoionization rate ($\Gamma_{\rm HI}$) and thermal history parameters (T_0 and γ). Even for a small H I neutral fraction[2], the Gunn-Peterson optical depth (τ_{GP}) is so large that it produces the absorption

---

[1] For simplicity, we assumed that the density of the two particles does not change during reionization.

[2] Ratio of H I number density to total hydrogen number density

trough. Hence Lyα forest studies can place a good constraint on Γ_{HI}. Thus Lyα forest is useful to constrain the cosmological and astrophysical parameters. Usually constraining these parameters involves comparing different properties of the Lyα forest derived from observed spectra with those from the simulated ones (Becker and Bolton 2013; Becker et al. 2011; Bolton and Haehnelt 2007; Bolton et al. 2008; Calura et al. 2012; Cooke et al. 1997; Gurvich et al. 2017; Kollmeier et al. 2014; Lidz et al. 2010; McQuinn et al. 2011; Meiksin and White 2004; Pontzen et al. 2014; Rauch et al. 1997; Shull et al. 2015; Viel et al. 2017; Zaldarriaga et al. 2001). Hence an efficient and accurate method to simulate Lyα forest is essential for large parameter space exploration.

1.3 Simulations of Lyα forest

In this section we review the existing methods to simulate the Lyα forest in the literature. These methods differ from each other in the way they calculate the overdensity (Δ), peculiar velocity (v) and temperature field (T). Study of these methods is essential to get physical insight of the Lyα forest, to see the advantages and limitations of each method for further improvement.

1.3.1 Semi-analytic method

The semi-analytic methods model the non-linear evolution of the baryonic density and velocity field fluctuations that produce Lyα forest using one of the approximations given below. In all these cases, the baryonic density field is estimated from the dark matter density field by some approximation scheme. These methods are simple, computationally inexpensive and helpful in getting the physical picture of the origin of the Lyα forest.

1.3.1.1 Zel'dovich approximation

In cosmological models, pressureless fluid (dark matter) dominates the mass density of the universe. The Zel'dovich approximation is a formulation of linear perturbation theory applicable to pressureless fluid (Doroshkevich and Shandarin 1977; Hui et al. 1997; McGill 1990; Zel'dovich 1970). Since all the fluctuations were smaller at early times, it is reasonable to assume that at recent epochs only the growing mode $D_+(t)$ has significant amplitude. The displacement of a given mass of fluid element in Zel'dovich approximation is given by (Doroshkevich and Shandarin 1977),

$$\mathbf{x}(\mathbf{q}, t) = \mathbf{q} + D_+(t)\ \nabla_q\ \psi(\mathbf{q}) \tag{1.36}$$

where $\mathbf{q}$ is the coordinate of the initial position, $D_+(t)$ is the linear growth factor and $\nabla_q\ \psi(\mathbf{q})$ is time independent part determined by the initial conditions. The peculiar velocity ($\mathbf{v}_{\rm pec}$) and overdensity ($\Delta_{\rm DM}$) for the dark matter is given by,

$$\begin{aligned}\Delta_{\rm DM} = 1 + \delta_{\rm DM} &= \det{}^{-1}\left[\delta_{ij} + D_+(t)\ \frac{\partial^2\psi}{\partial q_i \partial q_j}\right] \\ \mathbf{v}_{\rm pec} = a\frac{d\mathbf{x}}{dt} &= a\ \dot{D}_+\ \nabla_q\ \psi\end{aligned} \tag{1.37}$$

where overdot represents differentiation with respect to proper time, the right hand side of the first expression is Jacobian of the $\mathbf{q} - \mathbf{x}$ transformation matrix.

Given the overdensity, peculiar velocity and temperature (can be obtained by imposing $T - \Delta$ relation given in Eq. 1.28 or solving Eq. 1.19), one can compute the Lyα optical depth. However, there are two main problems in this approach: first the Zel'dovich approximation predicts the rapid growth in the thickness of the pancake (known as orbit crossing) and second on large scale baryons follow dark matter but on small scale baryon density field is smoothed as compared to dark matter density field because of finite pressure of the gas (Gnedin and Hui 1998). To circumvent these problems, the initial power spectrum is smoothed on small scales such that (a) the fluctuations in large scale due to Zel'dovich approximation are not affected and (b) the effect of finite pressure of baryons on the scales of local Jeans length[1] is accounted for.

The realization of overdensity, peculiar velocity and temperature fields are then generated by following method, (i) First a Gaussian random field $\psi(\mathbf{q})$ is generated on the grids, (ii) the particles are displaced from their initial grid positions to the new position according to Eq. 1.36, (iii) the peculiar velocities are assigned to the particles using Eq. 1.37, (iv) given the evolution of density of fluid element, it is straightforward to integrate temperature and ionization

---

[1] Pressure smoothing is usually characterized by Jeans length.

evolution equation and (v) the Lyα optical depth can then be calculated from Δ, $v_{\rm pec}$ and T field

This method is efficient as there is no need to integrate the equations of motion. The Lyα forest spectra obtained from this method are similar to those from observations (Hui et al. 1997). However, this method fails to reproduce the non-linear density field accurately. In addition, the small scale structure in Lyα forest are washed out due to smoothing of initial power spectrum on those scales.

1.3.1.2 Lognormal approximation

Since the Lyα forest arise from the linear or quasi-linear density fluctuations, one can neglect highly non-linear scales as a first order approximation. The basic assumption in this approach is that the probability distribution function of the density of gas in the IGM is given by a lognormal random field (Bi 1993; Bi and Davidsen 1997; Bi et al. 1992; Choudhury et al. 2001; Gnedin and Hui 1996). The lognormal approximation (hereafter LNA) has several features: (i) this model is simple and has no more free parameters than the Gaussian random field, (ii) it has correct asymptotic behavior on large scales where the IGM evolve linearly and small scales such as intracluster gas, (iii) the baryon density distribution in LNA is found to be in good agreement with that from hydrodynamical simulations, (iv) the Lyα forest can be simulated in 1 dimension with high enough resolution and (v) LNA has numerous predictions to compare with observations. The basic steps for calculating density and velocity field in LNA are (Bi and Davidsen 1997; Choudhury et al. 2001),

1. First, the 3 dimensional linear density field power spectrum for baryons is obtained from the 3 dimensional linear dark matter power spectrum by smoothing over scales below the Jeans length.

2. Power spectrum of density and velocity perturbations in 1 dimension is obtained by (angle) averaging the k modes in the 3 dimensional baryon density field power spectrum.

3. To simulate the density and velocity fields in 1 dimension, two Gaussian random fields with unit power spectra are generated.

4. The 1 dimensional density and velocity fields in real space is obtained from the linear combination of two Gaussian random fields (Bi 1993; Bi et al. 1992; Choudhury et al. 2001). Whereas, the 1 dimensional temperature is obtained by imposing power-law $T-\Delta$ (see Eq. 1.28) with T_0 and γ as free parameters.

Despite its simplicity and ability to capture basic picture of Lyα forest, LNA does not model non-linear density regime accurately (Coles et al. 1993; Viel et al. 2002). Furthermore, LNA is more analogous to Zel'dovich approximation in one dimension and can not reproduce pancaking in 3 dimensions.

1.3.2 Cosmological N body simulation

For realistic picture of the cosmic density and velocity fields, one need to solve the equations of motions (i.e., continuity equation, Euler equation, Poisson equation and equation of state see Eq. 1.74 and Eq. 1.75 in Appendix 1.7.6) simultaneously for many particles. In general these equations are difficult to solve analytically. Thanks to the advancement in computation, one can solve these equations numerically and get a realistic picture of the large scale structures in the universe. In N body simulations, dark matter (or baryons if present) distribution function is sampled using massive pseudo-particles and solve for their evolution either on a mesh or using a hierarchical tree. Typically the mesh code uses either Particle-Mesh (PM) or Particle-Particle Particle-Mesh (P^{3}M) techniques. On the other hand, in Tree algorithm the particles are arranged into a hierarchy of groups. The gravitational forces exerted by distant groups on a given particle are approximated by the lowest multiple moments.

1.3.2.1 Dark matter only simulation (PM method)

The contribution of dark matter to the mass density of the universe is large as compared to baryons. Since dark matter is pressure-less fluid, the equations of motion Eq. 1.74 and Eq. 1.75 can be simplified (Croft et al. 1998; Muecket et al. 1996; Petitjean et al. 1995). However, the finite pressure exerted by baryons prevents further collapse on small scales. This approximation is valid as long as the shocks are not occurred. As Lyα forest probes mildly non-linear density regime, the significant fraction of gas responsible for Lyα forest may not have been gone through shocks. Hence to the first order, this approximation is reasonable. By solving the Eq. 1.74 for dark matter, the density and velocity for baryons can be obtained. The temperature for the baryons can be calculated by integrating the energy equation along the particle trajectory (Muecket et al. 1996; Petitjean et al. 1995). However, Croft et al. (1998) used power law $T-\Delta$ relation (see Eq. 1.28) motivated by analytic calculations of Hui and Gnedin (1997). The Lyα forest obtained from PM method is in reasonable agreement with that from observations. Although this method is accurate than semi-analytic method, it does not model the highly non-linear density field accurately. It also ignores the effect of pressure smoothing of the baryons on small scales (Croft et al. 1998; Muecket et al. 1996; Petitjean et al. 1995). Hereafter we simply refer this method as PM method of Lyα forest.

1.3.2.2 Smoothed Particle Hydrodynamics (SPH) simulation

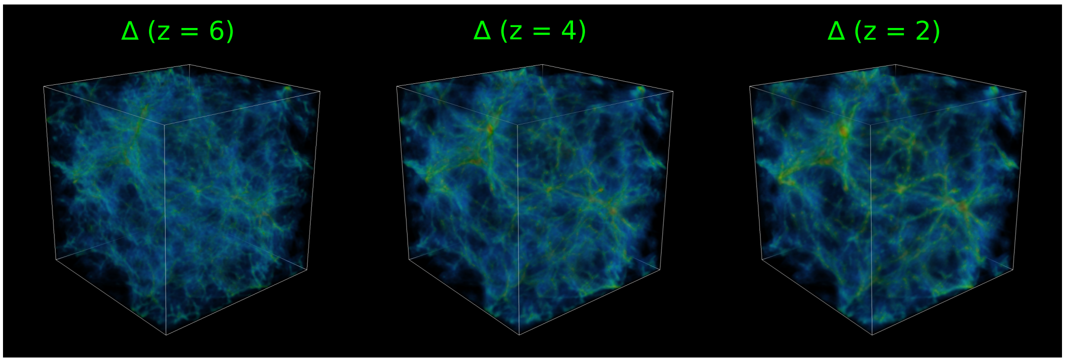

Figure 1.9: 3D evolution of the cosmic baryon density field at z = 6, 4, and 2 (from left to right respectively) in GADGET-3 simulation. The high density points are shown by red color whereas low density points occupying significant volume of the box are shown by blue points. The formation of high density, non-linear structures is clearly visible at low redshift $z = 2$. When a sightline passes through the filaments, absorption seen in Lyα forest is produced.

In SPH formulation the equations of motions for dark matter and baryons are solved simultaneously. SPH is based on Lagrangian description of the fluid flow in which observer follows an individual fluid element as it moves through space and time (Gingold and Monaghan 1977; Hernquist 1993; Hernquist and Katz 1989; Lucy 1977; Monaghan 1992; Springel 2005; Springel et al. 2001). Ideally a real fluid is represented by infinite number of such fluid elements with infinitesimal size. Let f be any continuous field quantity (e.g. density, velocity or temperature) to be represented in SPH formulation. The value of f at any point $\mathbf{r}$ is given by (Monaghan 1992),

$$f(\mathbf{r}) = \int_{\mathcal{V}'} f(\mathbf{r}')\, \delta^{3\mathrm{D}}(\mathbf{r} - \mathbf{r}')\, d\mathbf{r}' \tag{1.38}$$

where $\delta^{3\mathrm{D}}(\mathbf{r} - \mathbf{r}')$ is 3 dimensional Dirac delta function and integration is over volume $\mathcal{V}'$. Representing a fluid with infinite number of fluid element is not possible computationally. Hence the standard procedure is to represent the fluid with N fluid element (called as particles) with finite size. The local averages are performed over the volumes of the non-zero extent. In this formulation the estimate of the quantity f at any point $\mathbf{r}$ is given by (Monaghan 1992),

$$\langle f(\mathbf{r}) \rangle = \sum_{i=1}^{N} f(\mathbf{r_i})\, W(\mathbf{r} - \mathbf{r_i}, h_i)\, \frac{m_i}{\rho_i} \tag{1.39}$$

where h_i, m_i, ρ_i is smoothing length, mass and density of the SPH particle i respectively. $W(\mathbf{r} - \mathbf{r_i}, h_i)$ is known as smoothing kernel. The smoothing length h describes the extent of the averaging volumes. The smoothing kernel satisfies the two important properties (Monaghan 1992).

$$\begin{aligned} &\int_{\mathcal{V}'} W(\mathbf{r} - \mathbf{r}', h)\, d\mathbf{r}' = 1 \\ &\lim_{h \to 0} W(\mathbf{r} - \mathbf{r}', h) = \delta^{3\mathrm{D}}(\mathbf{r} - \mathbf{r}') \end{aligned} \tag{1.40}$$

It should be also noted that $\langle f(\mathbf{r}) \rangle \to f(\mathbf{r})$ when $h \to 0$. These properties also ensures the conservation of mass in SPH formulation. The smoothing kernel is typically chosen to be spherically symmetric spline kernel of the form (Springel 2005; Springel et al. 2001),

$$W(r,h) = \frac{8}{\pi\, h^3} \begin{cases} 1 - 6\left(\frac{r}{h}\right)^2 + 6\left(\frac{r}{h}\right)^3, & 0 \le \frac{r}{h} \le \frac{1}{2} \\ 2\left(1 - \frac{r}{h}\right)^3, & \frac{1}{2} \le \frac{r}{h} \le 1 \\ 0, & \frac{r}{h} > 1 \end{cases} \tag{1.41}$$

One of the most important quantity in any SPH formulation is the density estimate which is obtained from Eq. 1.39 as,

$$\langle \rho(\mathbf{r}) \rangle = \sum_{i=1}^{N} \rho_i \, W(\mathbf{r} - \mathbf{r_i}, h_i) \, \frac{m_i}{\rho_i} = \sum_{i=1}^{N} m_i \, W(\mathbf{r} - \mathbf{r_i}, h_i) \tag{1.42}$$

The equations of motion Eq. 1.74 involves the operation of gradient (∇) and time derivative ($\partial/\partial t$) operator on the quantity $f(\mathbf{r})$ (Monaghan 1992). The SPH formulation also provides a natural way of expressing the gradient of the local fluid properties as follows,

$$\langle \nabla f(\mathbf{r}) \rangle = \sum_{i=1}^{N} f(\mathbf{r_i}) \, \nabla W(\mathbf{r} - \mathbf{r_i}, h_i) \, \frac{m_i}{\rho_i} \tag{1.43}$$

Higher order gradients are calculated in similar manner.

Using these expressions, equations of motion Eq. 1.74 and Eq. 1.75 can be written in terms of smoothing kernel. These equations, in principle, can be solved for N fluid elements simultaneously in the computers. Springel (2005); Springel et al. (2001) developed a code "GAlaxies with Dark matter and Gas intEracT" (GADGET) for cosmological N body and SPH simulations on massively parallel computers with distributed memory. In GADGET-2[1], short-range gravitational forces are calculated using hierarchical tree algorithm whereas long-range gravitational forces are computed using FFT based particle-mesh (PM) method. The long-range and short-range gravitational forces are integrated with different time-steps. In addition, SPH formulation incorporates fully adaptive smoothing lengths and a novel entropy conserving technique (Springel and Hernquist 2002).

1.3.2.3 Lyα forest in HPM method

In the SPH simulations like GADGET-2 (Springel 2005) described above, one solves the equation of motion Eq. 1.74 and Eq. 1.75 for equal number of baryon and dark matter particles simultaneously. The baryon density and velocity fields obtained using GADGET-2 are accurate than semi-analytic method and dark matter only simulation. However, GADGET-2 does not incorporate radiative heating and cooling term (last two terms on right hand side of Eq. 1.19). As a result,

---

[1] Updated version of GADGET `http://wwwmpa.mpa-garching.mpg.de/gadget/`

$T - \Delta$ (see Eq. 1.28) for low density IGM is not power law. However one can still simulate Lyα forest by assuming power law $T - \Delta$ relation (Eq. 1.28) for a given gas density with T_0 and γ as free parameters (Choudhury et al. 2001; Gnedin and Hui 1998; Padmanabhan et al. 2014, 2015; Viel and Haehnelt 2006). This method although more accurate than semi-analytic method, does not capture smoothing effect on the density field due to finite pressure of the gas. In this work, we simply refer this method as hydrodynamic particle mesh or HPM method.

1.3.2.4 Lyα forest in full hydrodynamic simulation

Current state-of-the-art simulations like GADGET-3 (an updated version of GADGET-2), ILLUSTRIS[1] (Nelson et al. 2015; Vogelsberger et al. 2014), EAGLE[2] (Schaye et al. 2015) and GASOLINE[3] (Wadsley et al. 2004) etc. solves the ionization evolution equation (Eq. 1.22), calculates the temperature (Eq. 1.19) and evolves the density, velocity field (Eq. 1.74 and Eq. 1.75) self-consistently. Thus this method captures the effect of pressure smoothing on small scales in the baryon density and velocity fields.

Fig. 1.9 shows the 3D evolution of the cosmic baryon density field at $z = 6$, 4, and 2 (from left to right respectively) in GADGET-3 simulation. The high density points are shown by red color whereas low density points occupying significant volume of the box are shown by blue points. At low $z = 2$, one can clearly see the formation of sheets, filaments and nodes. The Lyα forest seen in the QSO absorption spectra is produced when a sightline passes through the filaments in cosmic density field. Fig. 1.10 shows a slice of the overdensity ($\log\Delta$, left panel), velocity (v_x along x axis, middle panel) and temperature ($\log T$, right panel) from GADGET-3 simulation. By shooting random sightlines through this simulation box $\log\Delta$, v_x and $\log T$ fields are generated. The Lyα optical depth is then computed from these fields. In addition to the basic physics, these simulations also incorporates complex astrophysical processes such as AGN, stellar feedback, shocks, galactic winds, metals and micro-turbulence etc. These simulations are

---

[1] `http://www.illustris-project.org/`

[2] `http://icc.dur.ac.uk/Eagle/`

[3] `http://imp.mcmaster.ca/software/`

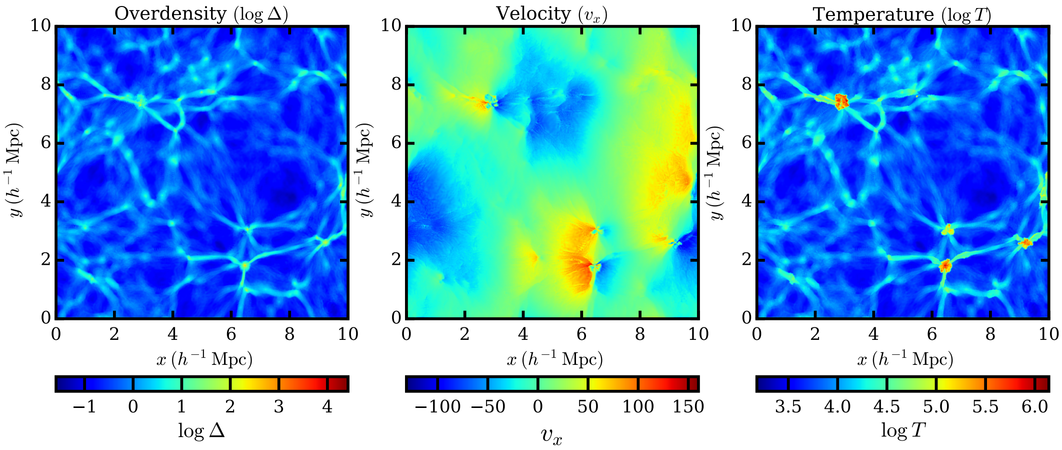

Figure 1.10: Slices of width ~ 10 ckpc from a GADGET-3 simulation box at $z = 2.5$. Left, middle and right panel show baryon overdensity ($\log \Delta$), baryon velocity component (v_x) along x axis and baryon temperature ($\log T$) field respectively. The colour scheme represents density of points in logarithmic unit. Random sightlines are shoot through simulation box and $\log \Delta$, v_x and $\log T$ along these sightlines are extracted. The Lyα forest is generated along these sightlines.

more accurate in modeling Lyα forest as compared to previously described methods. However, the accuracy of the method comes at the expense of computational time consumption.

SPH method is versatile, relatively simple to implement in numerical code, robust and transparent. By construction, SPH has conservation properties for energy, linear momentum and angular momentum. The latter is not automatically guaranteed in codes based on Eulerian approach.

Despite offering above mentioned advantages, SPH has limited accuracy in multi-dimensional flows. This is because local kernel interpolants are approximated through discrete sums over a small set of nearest neighbors. SPH can achieve high resolution in dense regions. However, in low-density regions where small number of SPH particles samples the density field the resolution is poor. SPH is also not suitable to model fluid instabilities across discontinuities such as Kelvin-Helmholtz and Rayleigh-Taylor instabilities (Agertz et al. 2007). In SPH particles may stream through each other. This can cause oscillations of SPH particles around converging flows such as shock fronts. To circumvent this problem artificial viscosity is usually introduced in SPH. This artificial viscosity is operating at some level outside of shocks which limits Reynolds numbers that can be reached in SPH.

1.3.2.5 Adaptive mesh refinement (AMR)

AMR is based on Eulerian description of the fluid flow in which observer is fixed at some location and tracks the fluid properties such as density, temperature flowing through that location (Berger and Colella 1989). In AMR, first the fluid properties are characterized in coarse grid. As the fluid evolves, the regions requiring more resolution are identified by some parameter that characterizes the solution e.g. local truncation error. The finer grid is superimposed only on these regions. Finer and finer sub grids are added until the local truncation error has dropped below the desired limit. It is important to note that grid spacing is fixed for the initial coarse grid whereas it is determined locally for the sub grids (Bryan et al. 1995). Currently there are several codes that solves equations of motion using AMR method for example ENZO[1] (O'Shea et al. 2004), RAMSES[2] (Teyssier 2002), NYX[3] (Almgren et al. 2013), HYDRA(Couchman et al. 1995) and ART(Knebe et al. 2001).

The main advantage of the AMR is that they have sufficient high resolution in low and high density regions. The fluid instabilities (like Kelvin-Helmholtz and Rayleigh-Taylor instabilities) can be better treated using an Eulerian formulation (AMR) of hydrodynamics. The artificial viscosity is not incorporated in this method. Unlike SPH, in AMR, angular momentum is not conserved by construction. Because finer sub grids are created while other cells are deleted to achieve resolution in area of interest, AMR codes are relatively complex for numerical implementation. AMR is not suitable for problems involving turbulent flows where fine resolution is required everywhere.

To summarize, in this section we described six different methods in the literature to simulate the Lyα forest namely: (i) Zel'dovich approximation, (ii) Lognormal approximation, (iii) Dark matter only simulation, (iv) GADGET-2 +

---

[1] `http://enzo-project.org/`

[2] `http://www.ics.uzh.ch/~teyssier/ramses/RAMSES.html`

[3] `https://ccse.lbl.gov/Research/NYX/`

$T-\Delta$ relation, (v) GADGET-3 with internal temperature evolution and (vi) AMR codes where temperature is evolved internally. The Lyα forest modeled from each of the above method has its own advantages and disadvantages. In the next section, we show the comparison of Lyα forest from different methods.

1.4 Comparison of Lyα forest from different simulations

With increasing number of observed high resolution QSO absorption spectra, the statistical errors on the measurement of Lyα flux statistics have reached the per cent level accuracy (McDonald et al. 2005; Viel et al. 2004b). However, the total error budget in the cosmological and astrophysical parameters constrained from Lyα forest are began to dominate by systematic uncertainties. Apart from observational effects, one of the main source of systematic uncertainty is in the modeling of the Lyα forest itself.

The semi-analytic methods like Zel'dovich approximation, Lognormal approximation provided an early argument that the Lyα forest itself may be produced from density fluctuations in a diffuse IGM. These methods were able to reproduce almost all the observed properties of the Lyα forest (Bi and Davidsen 1997; Bi et al. 1992; Choudhury et al. 2001; Hui et al. 1997). Soon the idea was confirmed by hydrodynamical simulations. Despite some differences in simulations, the general consensus was that the Lyα systems are produced by gravitationally induced density fluctuations in the matter density field (Cen et al. 1994; Hernquist et al. 1996; Miralda-Escudé et al. 1996; Zhang et al. 1995).

Since full hydrodynamic simulations are expensive, alternate approximate methods like PM and HPM were introduced. Typically the accuracy of these methods were tested by comparing Lyα flux statistics (e.g. flux probability distribution function, PDF,and flux power spectrum, PS) from full hydrodynamic simulations with those from these methods. Zhan et al. (2005) have compared the PM, HPM with their full hydrodynamic method and found that the flux PS in these methods differ by ~ 20 percent in both cases at $1 \leq z \leq 3$. The differences are larger at the lower redshift $z = 0$ due to fraction of baryons responsible for producing Lyα forest are not modeled correctly in their simulations (Zhan et al. 2005). Similarly Viel and Haehnelt (2006) found that the differences between the flux PS obtained from HPM (using GADGET-2) simulations and full hydrodynamical simulations are generally large ($20 - 30$ percent), and are not due to resolution effect of the HPM simulation. On the other hand flux PDF in two methods differed by $5 - 20$ % which was also consistent with Gnedin and Hui (1998).

Regan et al. (2007) compared simulations of the Lyα forest performed with two hydrodynamical codes ENZO (mesh based code) and GADGET-3 (SPH code, GADGET-2 modified for Lyα forest). The dark matter power spectrum in two methods show differences at the level of $1 - 3$ percent. Using the same algorithm for ionization and temperature evolution, the differences in the Lyα flux PDF and PS were ≤ 10 and ~ 5 percent respectively. Thus Lyα flux statistics (which are easy to obtained from observations) from PM and HPM differ by ~ 20 and 10 percent respectively when compared with full hydrodynamic simulations. On the other hand current state-of-the-art simulations using mesh based and SPH approach differ by 5 percent.

1.5 Motivation

As discussed in section 1.2, models of the Lyα forest depend on the cosmological and astrophysical parameters and are frequently used to constrain these parameters from observations (Becker and Bolton 2013; Becker et al. 2011; Bolton and Haehnelt 2007; Bolton et al. 2008; Calura et al. 2012; Cooke et al. 1997; Gurvich et al. 2017; Kollmeier et al. 2014; Lidz et al. 2010; McQuinn et al. 2011; Meiksin and White 2004; Pontzen et al. 2014; Rauch et al. 1997; Shull et al. 2015; Viel et al. 2017; Zaldarriaga et al. 2001). The basic idea behind constraining any astrophysical or cosmological parameter is (i) to identify a Lyα flux statistics sensitive to the parameter, (ii) derive the statistics from Lyα forest observations, (iii) derive the same statistics by modeling the Lyα forest in simulation with the parameter of interest as a free parameter, (iv) compute χ^2 between the statistics derived from observation and simulation and (v) obtain the best fit value and the associated statistical error on the quantity from χ^2 parabola.

Table 1.2: Different type of simulations in the literature

Simulations	Advantages	Disadvantages
Semi-analytical methods (Bi and Davidsen 1997; Bi et al. 1992; Choudhury et al. 2001; McGill 1990; Zel'dovich 1970)		
Zel'dovich approximation Lognormal approximation	Fast, captures basic picture of Lyα forest	quasi and non-linear density fields not modeled accurately
Simulations assuming Equation of State (Gnedin and Hui 1998; Muecket et al. 1996; Petitjean et al. 1995; Springel 2005)		
Dark matter only DM + Baryons (GADGET-2)	density and velocity field accurate than semi-analytical	does not include radiative cooling (and heating) and pressure smoothing
State-of-the-art simulation (Bryan et al. 2014; O'Shea et al. 2004; Springel 2005)		
SPH Simulation (GADGET-3) AMR Simulation (ENZO)	includes radiative processes, shocks, feedback processes, turbulence	Computationally expensive: Not efficient for large parameter space exploration

Often the cosmological and astrophysical parameters of interest are degenerate with other parameters for a given Lyα flux statistics e.g. thermal history parameters T_0 and γ are know to be degenerate with H I photoionization rate $\Gamma_{\rm HI}$. Hence one needs to probe the large parameter space to account for the degeneracy. At the same time, uncertainties in the Lyα forest simulations contribute significantly to the total error budget of the parameter under consideration. Hence the models of the Lyα forest, from which flux statistics are derived, should be sufficiently accurate to minimize the systematic uncertainty in the total error budget.

Table 1.2 summarizes the advantages and disadvantages of the various existing methods to simulate Lyα forest discussed in section 1.3 . The semi-analytical methods are fast and captures basic physics of Lyα forest but are not sufficiently accurate. On the other hand state of the art simulations like GADGET-3, ENZO incorporates complex astrophysical processes and are more accurate till date but are computationally expensive to probe large parameter space. Whereas, Lyα forest using PM and HPM method has intermediate accuracy and speed.
Our aim in this thesis is:

- To develop an efficient, flexible at the same time sufficiently accurate method to simulate the Lyα forest in the post-processing step of GADGET-2 simulation (analogous to HPM method).
- To account for the physical effects such as radiative heating and cooling processes, ionization (Eq. 1.22) and temperature evolution (Eq. 1.19), pressure smoothing of the baryons etc.

- To show the consistency of the method with full hydrodynamical simulation like GADGET-3 and other simulations in the literature
- To constrain the astrophysical parameter from Lyα forest observations with appropriate total error budget accounting for the degeneracy among different parameters.

1.6 Outline

The main scientific contribution of this thesis is efficient modeling of the Lyα forest in the post-processing step of cosmological N-body hydrodynamic simulation GADGET-2 and application of these models to constrain astrophysical parameters from observations. In particular, we develop an efficient post-processing tool for GADGET-2 simulations that accounts for the effect of thermal, ionization evolution and pressure smoothing of gas. Using these tools, we model various observed properties of the Lyα forest and compare with observations to constrain astrophysical parameters with appropriate errorbars.

The thesis is organized in 6 Chapters. In Chapter-2 and 3, we develop necessary tools to simulate and analyze Lyα forest. In Chapter-4, we show the consistency of these tools with other simulations / techniques in the literature. In Chapter-5, we apply these tools and constrain the H I photoionization rate ($\Gamma_{\rm HI}$) at $z \lesssim 0.45$ using Lyα forest observations obtained using Cosmic Origins Spectrograph on board Hubble Space Telescope (HST-COS). Finally we summarize the outcomes of the thesis in Chapter-6 with possible future work. We now briefly summarize each chapter as below.

- **Chapter - 2:** First we describe the "Code for Ionization and Temperature Evolution" (CITE) to model the thermal and ionization evolution of particles in the post-processing step of GADGET-2. At moderate to low resolution, the effect of pressure smoothing is not important. However for high resolution simulation, we capture the effect of pressure smoothing by running GADGET-2 at an elevated temperature floor of $\sim 10^4$ K, post-processed with CITE and using an appropriate smoothing kernel. We also discuss our module for "Generating Lyman Alpha forest Spectra in Simulation" (GLASS) along random line of sight. We show that the line of sight density, velocity, temperature field and Lyα transmitted flux is remarkably similar to those obtained from self-consistent simulations like GADGET-3.

- **Chapter - 3:** The statistics like H I column density distribution function (CDDF), linewidth (b) distribution require decomposition of Lyα forest spectra into multi-component Voigt profiles. The manual Voigt profile fitting of the large number of simulated Lyα forest is laborious and time consuming. Hence we have developed a parallel code called "VoIgt profile Parameter Estimation Routine" (VIPER) for automatically fitting the H I Lyα forest. We validate the VIPER with other techniques in the literature by comparing CDDF, b parameter distribution and b versus $\log \mathrm{N}_{\mathrm{HI}}$ lower envelope obtained using spectra of 82 QSO sightlines from HST-COS at $z \leq 0.45$.

- **Chapter - 4:** This chapter focuses on validity of our method of modeling Lyα forest. For low resolution simulation, that is used at low-z ($z < 0.5$) for comparison with HST-COS observations, we show the consistency of our method with other simulations in the past by comparing three metrics. On the other hand for high resolution simulation (at $2 \leq z \leq 4$), we validate our method by comparing 8 different statistics of Lyα forest with self-consistent simulation GADGET-3.

- **Chapter - 5:** This chapter is an application of tools developed in Chapter-2 and 3. We constrain the H I photoionization rate Γ_{HI} and its evolution in 4 different redshift bins $\bar{z} = 0.1, 0.2, 0.3$ and 0.4 using Lyα forest data along 82 QSO sightlines from HST-COS. In particular, we compare Lyα forest from observations with that from simulations using 3 statistics (i) Flux probability distribution function, (ii) flux power spectrum and (iii) CDDF. The final uncertainty in Γ_{HI} contains not only the statistical errors but also those arising from possible degeneracy with the thermal history of the IGM and cosmological parameters and uncertainties in fitting the QSO continuum.

- **Chapter - 6:** Finally we summarize the thesis in Chapter-6 and describe the possible improvements and application of these tools in future work.

1.7 Appendices

1.7.1 Basic Cosmology

In this section we define the important cosmological terms that are frequently used in the theoretical modeling of the Lyα forest. We refer reader to the Mo et al. (2010); Narlikar (2002); Padmanabhan (2000, 2010); Peebles (1993); Rindler (2001) references for the details on this section.

Cosmological principle: Modern cosmology is based on Einstein's general theory of relativity (GR) and cosmological principle which states that *on sufficiently large scale, the Universe is spatially homogeneous and isotropic.* Spatial homogeneity refers to translation symmetry. This means there is no preferred location in the universe i.e. the universe would appear same for two observer at two different locations in the universe. On the other hand principle of isotropy refers to rotational symmetry with no preferred direction in universe i.e. the universe appears same in all direction.

Proper length interval in GR: GR treats space and time on equal footing and combines them in four dimensional manifold called space-time. Each point in four dimensional space-time hyper surface is represented by three spatial coordinate and one time coordinate and is known as an event. The distance between any two events on this hyper surface is the length interval ds given by,

$$ds^2 = c^2\,dt^2 - dl^2 = c^2\,dt^2 - a^2(t)\left[\frac{dr^2}{1-Kr^2} + r^2\,d\Omega^2\right] \tag{1.44}$$

where dl is the proper spatial separation between the two events at world time t, K is curvature, r is the polar radius, $d\Omega$ is differential solid angle and c is speed of light. This is known as Friedmann-Lemaître-Robertson-Walker (FLRW) metric. Note that for $K = 0$, the above equation reduces to equation for flat space-time.

Comoving distance, Hubble parameter and Hubble time: The proper distance (l) between an observer at origin $r = 0$ and the observer at (r_1, Ω) can be written as,

$$l = a(t) \int_0^{r_1} \frac{dr}{\sqrt{1 - Kr^2}} = a(t)\ \chi(r_1) \tag{1.45}$$

where $a(t)$, χ is scale factor and comoving distance between the two observers. The redshift z is defined as $z \equiv 1/a(t) - 1$. The Hubble parameter, $H(t)$, at a cosmic time is defined as rate of change of proper length (l) expressed in units of l.

$$H(t) = \frac{1}{l}\,\frac{dl}{dt} = \frac{\dot{a}(t)}{a(t)} \tag{1.46}$$

The present time Hubble parameter is called as Hubble constant (H_0) and is usually expressed as,

$$\begin{aligned} h &\equiv \frac{H_0}{100\ \mathrm{km\,s^{-1}}\ \mathrm{Mpc}} \\ H_0 &= 3.241h \times 10^{-18}\ \mathrm{s}^{-1} \end{aligned} \tag{1.47}$$

It is usually convenient to express the densities of the various component of the Universe in terms of critical density (ρ_{crit}) defined as,

$$\begin{aligned} \rho_{\mathrm{crit}} &\equiv \frac{3H^2(t)}{8\pi G} \\ \rho_{\mathrm{crit},0} &= \frac{3H_0^2}{8\pi G} = \frac{3 \times (3.241h \times 10^{-18})^2}{8\pi \times 6.67259 \times 10^{-8}} = 1.88h^2 \times 10^{-29}\ \mathrm{g\ cm^{-3}} \end{aligned} \tag{1.48}$$

where $\rho_{\mathrm{crit},0}$ is present day critical density. The cosmic density parameters at a given epoch for matter, radiation, dark energy and curvature respectively is defined as,

$$\Omega_{\mathrm{m}} \equiv \frac{\rho_{\mathrm{m}}}{\rho_{\mathrm{crit}}}\ ,\ \ \Omega_{\mathrm{r}} \equiv \frac{\rho_{\mathrm{r}}}{\rho_{\mathrm{crit}}}\ ,\ \ \Omega_{\Lambda} \equiv \frac{\rho_{\Lambda}}{\rho_{\mathrm{crit}}}\ ,\ \ \Omega_{\mathrm{K}} \equiv \frac{Kc^2}{H^2(t)} = 1 - \Omega_{\mathrm{m}} - \Omega_{\mathrm{r}} - \Omega_{\Lambda} \tag{1.49}$$

where ρ_{m}, ρ_{r}, ρ_{Λ} are densities of matter, radiation and dark energy respectively at epoch. The Hubble parameter at redshift z can be written as,

$$H(z) = H_0\ E(z) = H_0\ [\Omega_{\mathrm{K},0}\ (1+z)^2 + \Omega_{\mathrm{m},0}\ (1+z)^3 + \Omega_{\mathrm{r},0}\ (1+z)^4 + \Omega_{\Lambda,0}]^{1/2} \tag{1.50}$$

where $\Omega_{\mathrm{X},0}$ is present day cosmic density parameter for component X.

Table 1.3: Cosmological parameters used in this thesis

Parameter	Symbol	Values
Matter density parameter	Ω_m	0.31
Baryon density parameter	Ω_b	0.0486
Dark energy density parameter	Ω_Λ	0.69
hubble parameter	h	0.674
Primordial power spectrum index	n_s	0.96
Helium fraction by mass	Y	0.24
RMS density fluctuations in $8h^{-1}$ Mpc	σ_8	0.83

Cosmological parameters: The modeling of the Lyα forest crucially depends on the type of cosmology and values of the cosmological parameters. These cosmological parameters are typically constrained from the sensitive observations of temperature and polarization anisotropies of the CMB (Hinshaw et al. 2013; Komatsu et al. 2011; Mather et al. 1994; Planck Collaboration et al. 2014, 2016). Throughout this work we use the flat ΛCDM cosmology consistent with latest constraints by Planck Collaboration et al. (2016).

1.7.2 The ultra-violet (UV) background

The UV background at any epoch is a diffuse isotropic radiation in the universe. The UVB is defined over a wavelength range 0.1 to 1000 μm is mainly contributed by QSOs and galaxies. The IGM at $z < 6$ is maintained at highly ionized state by the UVB. The accurate characterization (amplitude and shape) UVB is important for modeling the thermal and ionization state of the IGM (Becker and Bolton 2013; Becker et al. 2011; Boera et al. 2014; Lidz et al. 2010), measuring the baryon content in the IGM (Shull et al. 2012), and deriving metal abundances in the IGM using ionization corrections (Carswell et al. 2002; Peeples et al. 2014; Shull et al. 2014; Songaila 2001).

The number density of UVB photons at frequency ν_0 and redshift z_0 is given by,

$$n(\nu_0, z_0) = \frac{4\pi\, J_{\nu_0}(z_0)}{h_p\, c\, \nu_0} \tag{1.51}$$

where h_p, c and $J_{\nu_0}(z_0)$ is Planck constant, speed of light and the specific intensity of the UVB (in units of erg cm^{-2} s^{-1} Hz^{-1} sr^{-1}) at frequency ν_0 and redshift z_0 respectively. In UVB calculations, it is usually assumed that the UVB is contributed mainly by QSOs and galaxies at all wavelengths. The specific intensity of the UVB for an observer at redshift z_0 and frequency ν_0 can be obtained by solving the radiative transfer equation,

$$J_{\nu_0}(z_0) = \frac{1}{4\pi} \int_{z_0}^{\infty} dz \, \frac{dl}{dz} \, \frac{(1+z_0)^3}{(1+z)^3} \, \epsilon_\nu(z) \, \exp^{-\tau_{\rm eff}(\nu_0, z_0, z)} \tag{1.52}$$

where dl/dz is FLRW line element, $\nu = \nu_0(1+z)/(1+z_0)$ frequency of photon originated at redshift z, $\tau_{\rm eff}(\nu_0, z_0, z)$ is effective IGM optical depth encountered by photon (of frequency ν) emitted at redshift z while traveling to redshift z_0 and $\epsilon_\nu(z)$ is space averaged specific volume emissivity in units of erg s^{-1} Hz^{-1} Mpc^{-3} and is usually obtained from observed luminosity functions of QSOs and galaxies at a redshift z and frequency ν.

The photo-ionization rate (in units of s^{-1}) for specie X $\equiv$ (H I, He I, He II) is given by,

$$\Gamma_{\gamma \rm X} = \int_{\nu_{\rm X}}^{\infty} 4\pi \, J_\nu \, \sigma_{\rm X} \, \frac{d\nu}{h_p \nu} \tag{1.53}$$

where J_ν is specific intensity given in Eq. 1.52, $\nu_{\rm X}$ is frequency above which a photon can ionize the specie X and $\sigma_{\rm X}$ is cross-section for this process.

The photo-heating rate (in units of erg s^{-1}) for specie X $\equiv$ (H I, He I, He II) is given by,

$$\epsilon_{\gamma \rm X} = \int_{\nu_{\rm X}}^{\infty} 4\pi \, J_\nu \, \sigma_{\rm X} \, \frac{h_p\nu - h_p\nu_{\rm X}}{h_p \nu} \, d\nu \tag{1.54}$$

where J_ν, $\sigma_{\rm X}$ and $\nu_{\rm X}$ is the is specific intensity given in Eq. 1.52, photoionization cross-section of specie X, frequency above which specie X can be photoionized respectively. In this thesis, we have used two UVB models in the literature (Haardt and Madau 2012) and (Khaire and Srianand 2015b). Left panel in Fig. 1.11 shows the H I, He I and He II photoionization rate from (Khaire and Srianand 2015b) and (Haardt and Madau 2012) UVB models. Whereas the right panel in Fig. 1.11 shows the comparison of H I, He I and He II photoheating rates from (Khaire and Srianand 2015b) and (Haardt and Madau 2012) UVB models.

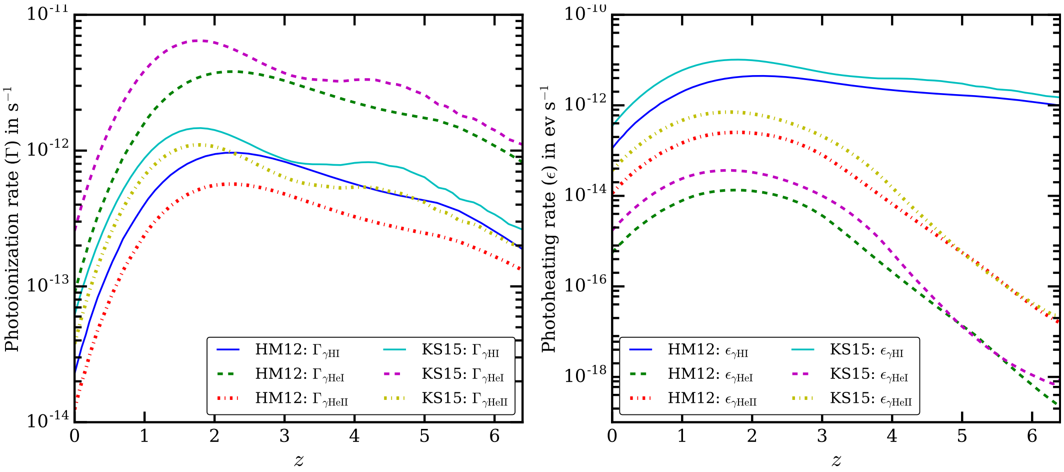

Figure 1.11: Left hand panel shows the photoionization rate for H I (solid curve), He I (dashed curve) and He II (dot dashed curve) from HM12 and KS15 UVB models. Right hand panel shows the photoheating rate for H I (solid curve), He I (dashed curve) and He II (dot dashed curve) in HM12 and KS15 UVB models.

1.7.3 Time scales

The low density gas in IGM is affected by various radiative processes such as photo-ionization, recombination etc. Furthermore the filaments and sheet like structures are formed by collapse of the objects under gravitational force. Typically the properties of the IGM are affected by all such processes. A simple way to judge the importance of these processes is to compare the time scales associated with them. Usually the processes with time scales greater than age of the Universe can be safely ignored which simplifies the analytic or numerical calculations. In this section we discuss the time scales relevant for the low density IGM studies in the cosmological context.

Hubble time: In the literature, it is customary to compare time scales of different physical processes with Hubble time scale. The Hubble time scale at redshift z is defined as the inverse of the Hubble parameter $H(z)$ at z.

$$t_H(z) = \frac{1}{H(z)} \tag{1.55}$$

Hubble time represents the time elapsed between big-bang singularity and epoch at redshift z *if expansion rate of the universe is constant* (i.e. Hubble parameter does not evolve with redshift). The Hubble parameter is not constant over the history of the universe. Hence Hubble time at redshift z is *not* same as the age of the universe at z. However, for an order of magnitude estimation the Hubble time is a good approximation to age of the Universe. For example the Hubble time at $z = 0$ is ~ 14.5 billion years (4.57×10^{17} s) whereas the age of the Universe at same redshift is ~ 13.77 (4.28×10^{17} s) billion years.

Dynamical time scale: Consider a spherical gas cloud of mass M and radius R which is supported by pressure against gravity. The dynamical time scale ($t_{\rm dyn}$) for this gas cloud is defined as the time required to collapse this object under the influence of gravity if pressure support is removed suddenly.

$$t_{\rm dyn} = \sqrt{\frac{3\pi}{16\, G\, \overline{\rho}}} \quad \text{with} \quad \overline{\rho} = \frac{3M}{4\pi R^3} \tag{1.56}$$

The dynamical time scale is related to free-fall time scale by $t_{\rm ff} = t_{\rm dyn}/\sqrt{2}$.

Sound-crossing time scale: Sound-crossing time scale for a gas cloud of size L at temperature T is defined as the time required for the pressure wave to propagate the distance equal to size of the cloud.

$$t_{\rm sc} \sim \frac{L}{c_s} \quad \text{with} \quad c_s = \left(\frac{5k_BT}{3m_p}\right)^{1/2} \tag{1.57}$$

where c_s, k_B, m_p is sound speed, Boltzmann constant, mass of proton. In the expression of sound speed, we have assumed the cloud is composed of non-relativistic monoatomic gas.

Photoionization time scales: The photoionization time scale is defined as the inverse of photoionization rate (see Eq. 1.53). Using $J_\nu = J_0\, \nu^{-\alpha}$ ($\alpha \sim 1$) and $\sigma_{\rm HI} = \sigma_0\, \nu^{-3}$, the photoionization time scale ($t_{\rm ph}$) for H I is given by,

$$\begin{aligned} t_{\rm ph} &= \Gamma_{\rm HI}^{-1} = \left(\frac{4\pi\, J_0\, \sigma_0}{h_p\, \nu_{\rm HI}^{3+\alpha}}\, \frac{1}{3+\alpha}\right)^{-1} \\ t_{\rm ph} &= 3.17 \times 10^{11}\ \text{s} \ \approx\ 10^{12}\ \text{s} \end{aligned} \tag{1.58}$$

Recombination time scales: For an optically thin gas the recombination time scales is defined as,

$$t_{\rm rec} = \frac{1}{\alpha^{\rm A}(T)\, n_{\rm e}} \tag{1.59}$$

where $\alpha^{\rm A}(T)$, $n_{\rm e}$ is case A recombination rate and number density of electrons respectively. Writing above equation in terms of mean total hydrogen density $\overline{n}_{\rm H}$ and putting numbers,

$$\begin{aligned} t_{\rm rec} &= \frac{1}{\alpha^{\rm A}(T)\, n_{\rm H}} \frac{\overline{n}_{\rm H}}{n_{\rm e}} \\ t_{\rm rec} &= 4.33 \times 10^{16}\ {\rm s}\ \frac{\overline{n}_{\rm H}}{n_{\rm e}} \end{aligned} \tag{1.60}$$

But we know that from Gunn-Peterson effect (see section 1.0.3) that the Universe is highly ionized at $z < 6$ such that neutral fraction is $\sim 10^{-4} - 10^{-5}$. This implies the recombination time scale $t_{\rm rec} \approx 10^{12}$ s.

1.7.4 Radiative cooling rates

In this section we have listed the various radiative cooling rates used in this thesis.

1.7.5 Net heating time

The net heating time for gas at a temperature T is given by,

$$\begin{aligned} t_{\rm heat} &= \frac{3\, k_{\rm B}\, T}{2\mu} \frac{m_{\rm H}}{\rho\, (1-Y)^2\, (\mathscr{H} - \mathscr{C})} \\ t_{\rm heat} &= \frac{3\, k_{\rm B}\, T}{2\mu} \frac{m_{\rm H}}{\overline{\rho}_0\, (1+z)^3\, \Delta\, (1-Y)^2\, (\mathscr{H} - \mathscr{C})} \\ t_{\rm heat} &= \frac{K_1\, T}{\Delta\, \mathscr{L}} \end{aligned} \tag{1.61}$$

where $k_{\rm B}$, μ, ρ, $\overline{\rho}_0$ and K_1 is Boltzmann constant, mean molecular weight, density of the gas at redshift z, mean density at redshift $z = 0$ and K_1 is factor that depends on physical constants and cosmological parameters respectively. Using the Eq. 1.21 and Eq. 1.24, we may calculate $\mathscr{L}$ for a given density, temperature and UVB model. We can however, simplify the cooling rates equation considerably for low-density IGM ($\Delta < 10$). In low density regime, the net heating rate, $\mathscr{L}$,

Table 1.4: Cooling rates for various radiative processes

Collisional ionization cooling

$c_1 = 2.54 \times 10^{-21}\, T^{1/2}\, \mathrm{e}^{-157809.1/T}\, (1 + T_5^{1/2})^{-1}\, e\,\mathrm{H\,I}$

$c_2 = 1.88 \times 10^{-21}\, T^{1/2}\, \mathrm{e}^{-285335.4/T}\, (1 + T_5^{1/2})^{-1}\, e\,\mathrm{He\,I}$

$c_3 = 9.90 \times 10^{-22}\, T^{1/2}\, \mathrm{e}^{-631515.0/T}\, (1 + T_5^{1/2})^{-1}\, e\,\mathrm{He\,II}$

Recombination cooling

$c_4 = 8.70 \times 10^{-27}\, T^{1/2}\, T_3^{-0.2}\, (1 + T_6^{0.7})^{-1}\, e\,\mathrm{H\,II}$

$c_5 = 1.55 \times 10^{-26}\, T^{0.3647}\, e\,\mathrm{He\,II}$

$c_6 = 3.48 \times 10^{-26}\, T^{1/2}\, T_3^{-0.2}\, (1 + T_6^{0.7})^{-1}\, e\,\mathrm{He\,III}$

Dielectric recombination cooling

$c_7 = 1.24 \times 10^{-13}\, T^{-1.5} \mathrm{e}^{-470000/T}\, (1 + 0.3\mathrm{e}^{-94000/T})\, e\,\mathrm{He\,II}$

Collisional excitation cooling

$c_8 = 7.50 \times 10^{-19}\, \mathrm{e}^{-118348/T}\, (1 + T_5^{1/2})^{-1}\, e\,\mathrm{H\,I}$

$c_9 = 5.54 \times 10^{-17}\, \mathrm{e}^{-473638/T}\, (1 + T_5^{1/2})^{-1}\, T^{-0.397}\, e\,\mathrm{He\,II}$

Bremsstrahlung

$c_{10} = 1.42 \times 10^{-27}\, g_f\, T^{1/2}\, e\, (\mathrm{H\,II} + \mathrm{He\,II} + 4\mathrm{He\,III})$

$g_f = 1.1 + 0.34\, \mathrm{e}^{-[5.5 - \log_{10}(T)]^2/3}$

Inverse Compton cooling

$c_{11} = 5.406 \times 10^{-36}\, [T - 2.7(1+z)]\, (1+z)^4 e/n_H$

z is redshift of interest

$T_n \equiv T/10^n$, T is temperature in K

n_H is number density of total hydrogen

$\mathrm{X} = n_\mathrm{X}/n_\mathrm{H}$ where $\mathrm{X} \equiv [\mathrm{H\,I}, \mathrm{H\,II}, \mathrm{He\,I}, \mathrm{He\,II}, \mathrm{He\,III}, \mathrm{e}]$

n_X is number density of specie X

Table 1.5: Recombination and collisional ionization rates in s^{-1}

Recombination rates

$$\alpha_{\mathrm{HII}} = 6.30 \times 10^{-11}\, T^{-1/2}\, T_3^{-0.2}/(1 + T_6^{0.7})$$

$$\alpha_{\mathrm{HeII}} = 1.50 \times 10^{-10}\, T^{-0.6353} + \alpha_{\mathrm{HeII}}^{(D)}$$

$$\alpha_{\mathrm{HeIII}} = 3.36 \times 10^{-10}\, T^{-1/2}\, T_3^{-0.2}/(1 + T_6^{0.7})$$

Dielectric recombination rates

$$\alpha_{\mathrm{HeII}}^{(D)} = 1.90 \times 10^{-3}\, T^{-1.5}\, \mathrm{e}^{-470000/T}(1 + 0.3\mathrm{e}^{-94000/T})$$

Collisional ionization rates

$$\Gamma_{e\mathrm{HI}} = 1.17 \times 10^{-10}\, T^{1/2}\, \mathrm{e}^{-157809.1/T}\, (1 + T_5^{1/2})^{-1}$$

$$\Gamma_{e\mathrm{HeI}} = 4.76 \times 10^{-11}\, T^{1/2}\, \mathrm{e}^{-285335.4/T}\, (1 + T_5^{1/2})^{-1}$$

$$\Gamma_{e\mathrm{HeII}} = 1.14 \times 10^{-11}\, T^{1/2}\, \mathrm{e}^{-631515.0/T}\, (1 + T_5^{1/2})^{-1}$$

is dominated by H I, He II photo-heating and small but non-negligible inverse Compton cooling. We can write the photoheating component of $\mathscr{L}$ as $\mathscr{L}_\epsilon$ and is given by,

$$\mathscr{L}_\epsilon = \mathrm{HI}\, \frac{\epsilon_{\gamma\mathrm{HI}}}{n_{\mathrm{H}}} + \mathrm{HeII}\, \frac{\epsilon_{\gamma\mathrm{HeII}}}{n_{\mathrm{H}}} \tag{1.62}$$

Since H I and H II are likely to be highly photoionized, we may assume photoionization equilibrium and write

$$\mathrm{HI} = \frac{n_{\mathrm{HI}}}{\mathrm{n_H}} = \frac{\alpha_{\mathrm{HII}}\, n_{\mathrm{e}}}{\Gamma_{\gamma\mathrm{HI}}} \quad \text{and} \quad \mathrm{HeII} = \frac{n_{\mathrm{HeII}}}{\mathrm{n_H}} = \frac{\alpha_{\mathrm{HeIII}}\, n_{\mathrm{e}}\, y}{\Gamma_{\gamma\mathrm{HeII}}} \tag{1.63}$$

At $T < 10^5$ K, the recombination rate coefficients can be well approximated by $T^{-\beta}$ (typically $\beta = 0.7$) such that,

$$\alpha_{\mathrm{HII}} = \alpha'_{\mathrm{HII}}\, T^{-\beta} \quad \text{and} \quad \alpha_{\mathrm{HeIII}} = \alpha'_{\mathrm{HeIII}}\, T^{-\beta} \tag{1.64}$$

Substituting Eq. 1.63 and 1.64 into Eq. 1.62 and using $\mathrm{e} = n_{\mathrm{e}}/n_{\mathrm{H}} = (1 + 2y)$,

$$\begin{aligned} \mathscr{L}_\epsilon &= \left[\alpha'_{\mathrm{HII}}\, \mathrm{e}\, \frac{\epsilon_{\gamma\mathrm{HI}}}{\Gamma_{\gamma\mathrm{HI}}} + \alpha'_{\mathrm{HeIII}}\, \mathrm{e}\, y\, \frac{\epsilon_{\gamma\mathrm{HeII}}}{\Gamma_{\gamma\mathrm{HeII}}} \right] T^{-\beta} \\ \mathscr{L}_\epsilon &= K_2\, T^{-\beta} \end{aligned} \tag{1.65}$$

Note that photoionization rate (Γ_γ) and photoheating rate (ϵ_γ) evolves with redshift z and can be obtained from UVB model (see Fig. 1.11). As mentioned earlier, there is also small contribution from inverse Compton cooling, $\mathscr{L}_{\rm cc}$, (coefficient c_{11} in Table 1.4 for $T >> T_{\rm CMB}$) which may be written as,

$$\begin{aligned}
\mathscr{L}_{\rm cc} &= \frac{5.406\times10^{-36}\,T\,(1+z)^4\,{\rm e}}{n_{\rm H}} = \frac{5.406\times10^{-36}\,T\,(1+z)^4\,{\rm e}}{\overline{n}_{\rm H,0}\,\Delta\,(1+z)^3}\\
\mathscr{L}_{\rm cc} &= \frac{5.406\times10^{-36}\,(1+z)^3\,{\rm e}}{\overline{n}_{\rm H,0}}\;\frac{T}{\Delta}\\
\mathscr{L}_{\rm cc} &= K_3\,\frac{T}{\Delta}
\end{aligned} \tag{1.66}$$

where $\overline{n}_{\rm H,0}$ mean number density of total hydrogen at redshift $z=0$. The total net heating rate is given by,

$$\begin{aligned}
\mathscr{L} &= \mathscr{L}_\epsilon - \mathscr{L}_{\rm cc}\\
\mathscr{L} &= K_2\,T^{-\beta} - K_3\,T\,\Delta^{-1}
\end{aligned} \tag{1.67}$$

The net heating time is given by Eq. 1.61 and Eq. 1.67 as,

$$t_{\rm heat} = \frac{K_1\,T}{\Delta\,(K_2\,T^{-\beta} - K_3\,T\,\Delta^{-1})} \tag{1.68}$$

1.7.6 Fluid Equations in expanding universe

In the hierarchical structure formation model, the small initial density fluctuations (generated by some mechanism e.g. quantum fluctuations) grow via gravitational instabilities. These perturbations grow with time and eventually form the large scale structures we see today. Since the formation of structures takes place at scales much smaller than Hubble scale, relativistic effects can be ignored and Newtonian approach is applicable to extract essential physics. We can treat baryons and dark matters as fluids. The evolution of fluid element with density (ρ), velocity ($\mathbf{V}$) and pressure (P) under the influence of gravitational potential (ϕ) is governed by following set of equations,

$$\begin{aligned}
\text{Continuity Equation:}\quad & \dot{\rho}(t,\mathbf{r}) + \nabla_r\,\cdot\,[\rho(t,\mathbf{r})\;\mathbf{V}(t,\mathbf{r})] = \;0\\
\text{Euler Equation:}\quad & \dot{\mathbf{V}}(t,\mathbf{r}) + \Big[\mathbf{V}(t,\mathbf{r})\cdot\nabla_r\Big]\,\mathbf{V}(t,\mathbf{r}) \;= -\nabla_r\phi(t,\mathbf{r}) - \frac{\nabla_r P(t,\mathbf{r})}{\rho(t,\mathbf{r})}\\
\text{Poisson Equation:}\quad & \nabla_r^2\phi(t,\mathbf{r}) = 4\pi G\rho(t,\mathbf{r})
\end{aligned} \tag{1.69}$$

where overdot represent partial derivative with respect to time and ∇_r is spatial gradient operator with respect to proper co-ordinate $\mathbf{r}$. These equations can be rewritten in terms of comoving co-ordinate by following transformations,

$$\begin{aligned} \mathbf{r} &= a(t)\,\mathbf{x} \\ \frac{d\mathbf{r}}{dt} &= \frac{da(t)}{dt}\,\mathbf{x} + a(t)\,\frac{d\mathbf{x}}{dt} \\ \mathbf{V} &= \frac{\dot{a}}{a}\,\mathbf{r} + \mathbf{v} \end{aligned} \tag{1.70}$$

where $\mathbf{v}$ is peculiar velocity with respect to fundamental observer. To switch from frame of $(\mathbf{r}, t)$ to $(\mathbf{x}, t)$ one has to do the following transformation,

$$\nabla_r = \frac{1}{a}\,\nabla_x \quad \text{and} \quad \frac{\partial}{\partial t} \to \frac{\partial}{\partial t} - \frac{\dot{a}}{a}\mathbf{x}\cdot\nabla_x \tag{1.71}$$

where ∇_x is spatial gradient operator with respect to comoving co-ordinate $\mathbf{x}$. The resultant equation is simplified if written in terms of perturbed quantity as follows,

$$\begin{aligned} \text{Density contrast:} \quad & \delta(t,\mathbf{x}) \equiv \Delta(t,\mathbf{x}) - 1 = \frac{\rho(t,\mathbf{x})}{\overline{\rho}(t)} - 1 \\ \text{Peculiar velocity field:} \quad & \mathbf{v}(\mathbf{t},\mathbf{x}) = \mathbf{V}(\mathbf{t},\mathbf{x}) - \frac{\dot{a}}{a}\,\mathbf{r} \\ \text{Perturbed pressure:} \quad & p(t,\mathbf{x}) = P(t,\mathbf{x}) - \overline{P}(t) \\ \text{Perturbed gravitational potential:} \quad & \phi(t,\mathbf{x}) = \Phi(t,\mathbf{x}) - \overline{\Phi}(t) \end{aligned} \tag{1.72}$$

where symbols with bar represents average quantities independent of spatial co-ordinates. The perturbed fluid equation after subtracting out zeroth order unperturbed part are given as,

$$\begin{aligned} &\text{Perturbed continuity Equation:} \\ &\dot{\delta} + \frac{1}{a}\,\nabla_x\cdot[(1+\delta)\mathbf{v}] = 0 \\ &\text{Perturbed Euler Equation:} \\ &\dot{\mathbf{v}} + \frac{\dot{a}}{a}\mathbf{v} + \frac{1}{a}\Big(\mathbf{v}\cdot\nabla_x\Big)\mathbf{v} = -\frac{1}{a}\,\nabla_x\left(\phi + \frac{\ddot{a}\,a\,\mathbf{x}^2}{2}\right) - \frac{\nabla_x\,p}{\overline{\rho}\,(1+\delta)\,a} \\ &\text{Perturbed Poisson Equation:} \\ &\nabla_x^2\phi = 4\,\pi\,G\,\overline{\rho}\,\delta\,a^2 \end{aligned} \tag{1.73}$$

We need to solve the above set of equations for dark matter and baryons separately to study the evolution structure formation in details. Unlike baryons, for collisionless dark matter the pressure term in above expression is zero ($p = 0$).

$$\begin{aligned}
\dot{\delta}_{\mathrm{DM}} + \frac{1}{a}\,\nabla_x \cdot [(1+\delta_{\mathrm{DM}})\mathbf{v}_{\mathrm{DM}}] = {}& 0 \\
\dot{\mathbf{v}}_{\mathrm{DM}} + \frac{\dot{a}}{a}\mathbf{v}_{\mathrm{DM}} + \frac{1}{a}\Big(\mathbf{v}_{\mathrm{DM}} \cdot \nabla_x\Big)\mathbf{v}_{\mathrm{DM}} = {}& -\frac{1}{a}\,\nabla_x\left(\phi + \frac{\ddot{a}\,a\,\mathbf{x}^2}{2}\right) \\
\dot{\delta}_B + \frac{1}{a}\,\nabla_x \cdot [(1+\delta_B)\mathbf{v}_B] = {}& 0 \\
\dot{\mathbf{v}}_B + \frac{\dot{a}}{a}\mathbf{v}_B + \frac{1}{a}\Big(\mathbf{v}_B \cdot \nabla_x\Big)\mathbf{v}_B = {}& -\frac{1}{a}\,\nabla_x\left(\phi + \frac{\ddot{a}\,a\,\mathbf{x}^2}{2}\right) - \frac{\nabla_x\,p_B}{\overline{\rho}_B\,(1+\delta_B)\,a} \\
\nabla_x^2\phi = 4\,\pi\,G\,a^2\,(\overline{\rho}_{\mathrm{DM}}\,\delta_{\mathrm{DM}} + \overline{\rho}_{\mathrm{B}}\,\delta_{\mathrm{B}})&
\end{aligned} \tag{1.74}$$

The above set of equations are usually supplemented with “effective equation of state” that characterizes the relation between density and pressure of baryons. Usually, it is expressed in polytropic form with index γ,

$$p_B = \frac{\rho_B\,k_{\mathrm{B}}\,T}{\mu\,m_p} = \frac{\overline{\rho}_B\,k_{\mathrm{B}}\,T_0}{\mu\,m_p}\,(1+\delta_B)^\gamma \tag{1.75}$$

where $\overline{\rho}_B$, k_{B}, T_0, m_p, $\mu = 4/(8-5Y)$ is average baryon density (assuming fully ionized H and He), Boltzmann constant, temperature of baryons at mean cosmic density, mass of proton and mean molecular weight for gas composed of fully ionized hydrogen and helium respectively. Y is helium fraction by mass. The Eq. 1.74 with Eq. 1.75 completely describes the growth of structures in the universe. In the limited cases and under some approximations, the above set of equations can be solved analytically. While analytic methods are faster and provide physical insights, they may not capture the non-linear physics correctly. In this case one has to resort to cosmological N-body hydrodynamic simulations. As described in section 1.3, these equations of motion are solved in N-body simulations using Lagrangian or Eulerian formulation of fluid elements and provides a method to study physical phenomenon in non-linear regime.

2 Code for Ionization and Temperature Evolution

A major part of this chapter is based on Gaikwad et al. (2017a,b).

2.1 Introduction

The observed properties of the Lyα forest are sensitive to fluctuations in the cosmic density fields, velocity fields and physical conditions like the temperature, turbulence and ionizing radiation prevailing in the intergalactic medium (IGM; Cen et al. 1994; Hernquist et al. 1996; Miralda-Escudé et al. 1996; Zhang et al. 1995). As a result, Lyα forest has been used in the literature to constrain cosmological parameters such as $\Omega_{\rm m}$, $\Omega_{\rm b}$, σ_8, $n_{\rm s}$ (see, e.g., McDonald et al. 2005; Viel et al. 2004a,b), and astrophysical parameters such as mean IGM temperature T_0, slope of the equation of state γ (Becker et al. 2011; Boera et al. 2014; Lidz et al. 2010; Schaye et al. 1999, 2000; Zaldarriaga et al. 2001) and H I photoionization rate ($\Gamma_{\rm HI}$, Becker and Bolton 2013; Gurvich et al. 2017; Kollmeier et al. 2014; Rauch et al. 1997; Shull et al. 2015; Viel et al. 2017). Usually constraining these parameters involves comparing different properties of the Lyα forest derived from observed spectra with those from the simulated ones. Since many of the parameters, particularly those related to the thermal state of the IGM, are poorly understood, obtaining robust constraints would require exploring a sufficiently wide range of parameter values. It is thus useful to develop newer methods of simulating the IGM that are efficient, flexible and at the same time sufficiently accurate.

In this chapter, we present our method which is based on performing smooth particle hydrodynamical (SPH) simulations to generate the density and velocity distributions of baryons and then post-processing the outputs to solve for the IGM temperature in presence of a UVB. For this purpose, we have developed a module, called the "Code for Ionization and Temperature Evolution" (CITE), to evolve the IGM temperature from initial (higher) redshift to the final (lower) redshift of our interest. While CITE works well in the low resolution simulation which are suitable for the Lyα forest at $z < 0.5$, the dynamical evolution of SPH particles at finite pressure is an important effect when we consider high resolution simulations that are suitable for the Lyα forest at $2 \leq z \leq 4$. We show that for high resolution simulation ($2 \leq z \leq 4$), we can account for the effect of pressure smoothing by running a GADGET-2 simulation at elevated temperature floor and using local Jeans scale. The advantage of this method is that it is computationally less expensive and sufficiently flexible to account for variations in the thermal history. Our analysis allows us to study the degeneracy between various astrophysical parameters related to the ionization and thermal history.

This chapter is organized as follows. In Section 2.2, we describe our method of modeling low-z ($z \leq 0.5$) Lyα forest. Details of the simulations, along with our method of calculating the temperature evolution using CITE and generating Lyα forest using GLASS is discussed in Section 2.2.1 , 2.2.3 and 2.2.4 respectively. We discuss the method of simulating Lyα forest at high-z ($2 \leq z \leq 4$) in Section 2.3. The method to account for the effect of dynamic pressure by smoothing (in 3 dimensions) the density and velocity fields over a local Jeans scale is described in Section 2.3.2.2. Finally, we summarize our method in Section 2.4.

Throughout this thesis, we use flat ΛCDM cosmology with parameters given in Table 1.3 which is consistent with Planck Collaboration et al. (2016). The H I photoionization rate ($\Gamma_{\rm HI}$) expressed in units of 10^{-12} s^{-1} is denoted as Γ_{12}. Unless mentioned all the distances are expressed in comoving co-ordinates.

2.2 Low-z Lyα forest

The Lyα forest arises from relatively low density and low temperature diffuse medium which accounts for 90 per cent of the baryons at $2 \leq z \leq 4$

(Bolton and Becker 2009; Cen et al. 1994; Faucher-Giguère et al. 2008a; Hernquist et al. 1996; Miralda-Escudé et al. 1996; Paschos and Norman 2005; Rauch et al. 1997; Zhang et al. 1995). On the other hand at low-z ($z \leq 1.6$), only a small fraction ($\sim 30 - 40$ per cent) of the baryons are in diffuse medium responsible for Lyα forest (Davé et al. 2010; Davé and Tripp 2001; Shull et al. 2015; Smith et al. 2011; Tepper-García et al. 2012; Theuns et al. 1998a). It turns out that a significant fraction ($\sim 30 - 50$ per cent) of the baryons are in a phase known as the warm hot intergalactic medium (WHIM) (Cen and Fang 2006; Cen and Ostriker 1999, 2006; Davé et al. 2001, 2010; Lehner et al. 2007; Shull et al. 2012; Smith et al. 2011) and they are difficult to detect in either emission or absorption in the UV/optical bands. Hence to model the Lyα forest at low-z, one needs simulation incorporating accurate thermal and ionization history.

2.2.1 Simulation

We generate the cosmological density and velocity fields using the smoothed particle hydrodynamic code GADGET-2[1] (Springel 2005). The initial conditions for the simulations are generated at a redshift $z = 99$ using the publicly available code 2LPT[2] (Scoccimarro et al. 2012). We use $1/30^{th}$ of the mean inter-particle distance as our gravitational softening length. The simulation outputs are stored at a redshift interval of 0.1 between $z = 2.1$ and $z = 0$. We use 2 simulation boxes containing 512^3 dark matter and an equal number of gas particles in a cubical box. Both simulation boxes are $50h^{-1}$ cMpc in size with different initial conditions. We use these boxes to study cosmic variance. As shown by Smith et al. (2011), the phase distribution of baryons in simulation box at low-z is converged if the box size $50h^{-1}$ cMpc or above. The simulations used in this work do not include AGN feedback, outflows in the form of galactic wind or micro-turbulence. Note that the gas heating due to hydrodynamical processes arising from structure formation is incorporated in GADGET-2. However radiative heating and cooling processes are not incorporated in GADGET-2 and we include them in post processing step (see section 2.2.3). We use GADGET-2 with post-processing instead of

---

[1] http://www.mpa.mpa-garching.mpg.de/gadget/

[2] http://cosmo.nyu.edu/roman/2LPT/

GADGET-3 in order to probe the wide range of parameter space. Later in Chapter 4, we compare our results with those obtained using different simulations in the literature.

2.2.2 Density, velocity and temperature

Each GADGET-2 output snapshot contains the position, velocity, internal energy per unit mass and smoothing length l of each smooth particle hydrodynamic (SPH) particle. We evaluate the above quantities on a uniform grid in the box using the smoothing kernel (Springel 2005),

$$W(r,l) = \frac{8}{\pi\, l^3} \begin{cases} 1 - 6\left(\frac{r}{l}\right)^2 + 6\left(\frac{r}{l}\right)^3, & 0 \le \frac{r}{l} \le \frac{1}{2} \\ 2\left(1 - \frac{r}{l}\right)^3, & \frac{1}{2} \le \frac{r}{l} \le 1 \\ 0, & \frac{r}{l} > 1 \end{cases} \tag{2.1}$$

where r is the distance between the grid point and the particle position. The density at the i^{th} grid is simply the sum of density contribution from all particle weighted by the smoothing kernel

$$\rho_i = \sum_{j=1}^{N} m_j \, W(|\mathbf{r}_{ij}|, l_j), \tag{2.2}$$

where N is the total number of particles, $|\mathbf{r}_{ij}|$ is the distance between the i^{th} grid point and j^{th} particle. m_j and l_j are the mass and smoothing length of the j^{th} particle respectively. The overdensity at the grid point i is given by

$$\Delta_i = \frac{\rho_i}{\overline{\rho}}, \quad \text{with} \quad \overline{\rho} = \frac{1}{N}\sum_{i=1}^{N} \rho_i \tag{2.3}$$

where $\overline{\rho}$ is the average density. In this work the symbol Δ is used for baryon overdensity. The density weighted estimate of any quantity f at the i^{th} grid point is given by

$$f_i = \sum_{j=1}^{N} f_j \, \frac{m_j}{\rho_j} \, W(|\mathbf{r}_{ij}|, l_j), \tag{2.4}$$

where f_j is the value of the quantity for the j^{th} particle. The quantity f could be one of the velocity components (v_x, v_y, v_z) or the internal energy per unit mass

u. The temperature of the gas as computed by GADGET-2, which we denote as T_g, can be obtained from the internal energy using the relation

$$T_g = \frac{2m_p}{3k_B}\, u, \tag{2.5}$$

here we have assumed a monoatomic gas composition with the ratio of specific heats given by 5/3. In the above expression, m_p is the mass of a proton and k_B is the Boltzmann's constant.

Fig. 2.1 shows a two-dimensional slice from our simulation box at $z = 0.3$. The left-hand panel and middle panel shows, respectively, the overdensity (Δ) and the temperature (T_g) of baryons as obtained from the GADGET-2 output snapshots. By comparing the left-hand and middle panels, one can see that the temperature distribution broadly traces the density distribution. This is related to the fact that the high density regions are heated because of hydrodynamical processes. Hereafter we will refer to this heating simply as shock heating. The temperature can be as high as $\sim 10^7$ K in the vicinity of collapsed objects. The voids, on the other hand, remain extremely cool at temperatures ~ 10 K.

However, the above scenario does not capture all the relevant physics, in particular the photoheating of the low-density gas by UVB to higher temperatures and various cooling processes. It is indeed found from other simulations, that take into account the additional heating and cooling processes (Hui and Gnedin 1997; McDonald et al. 2001), that the temperature and density follow a reasonably tight relation (which we will refer to as the $T - \Delta$ relation) for mildly non-linear densities, i.e. $\Delta \leq 10$. As GADGET-2 does not include processes like the photoheating and radiative cooling, we find that the resulting temperature (T_g) and density relation does not show any power-law correlation.

This shortcoming has been addressed in subsequent versions of the GADGET, e.g., GADGET-3 (as discussed in Bolton et al. 2006) where one can perform the simulation in presence of a UVB. In this work, however, we follow a slightly different approach to account for the effects of photoionizing UVB and radiative cooling. Our method involves post-processing the GADGET-2 output to calculate the temperatures. We will show that this method produces results which are consistent with other works. The advantage of our method, however, is that we are able to explore the parameter space more efficiently without having to perform the full SPH simulation multiple times. In the following section we outline our method to evolve the IGM temperature in the post-processing step using 'Code for Ionization and Temperature Evolution' (CITE).

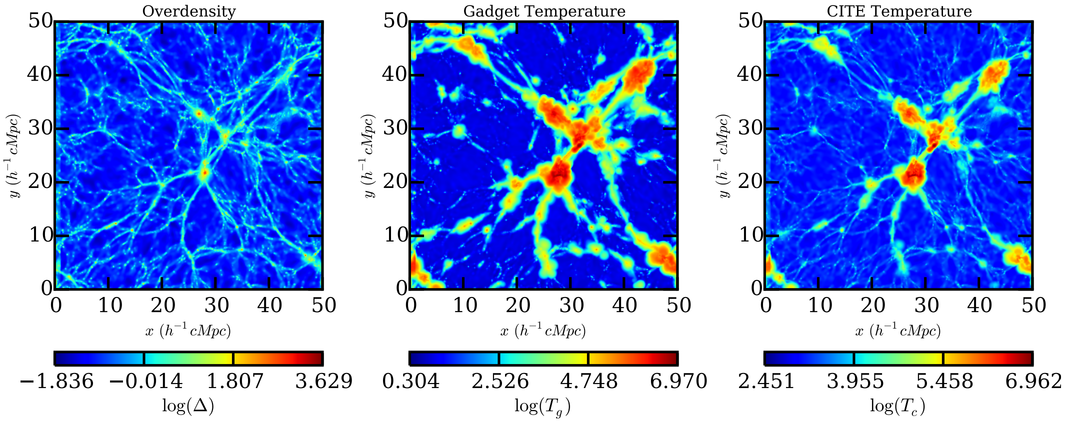

Figure 2.1: Two-dimensional slices of width 0.1 h^{-1} cMpc obtained from the GADGET-2 output snapshot at $z = 0.3$. *Left-hand panel:* the distribution of baryon overdensity Δ. Color scheme is such that red and blue color represent highest density and lowest density regions respectively. *Middle panel:* the gas temperature T_g from GADGET-2 (see section 2.2.2). *Right-hand panel:* the gas temperature T predicted after evolving the temperature from $z_1 = 2.1$ (initially at z_1, $T_0 = 15000$ K and $\gamma = 1.3$) using our post-processing module CITE (see Section 2.2.3). The highly overdense regions are at higher temperatures because of the shock heating resulting from the structure formation. The color scheme in middle and right-hand panel is such that red and blue color corresponds to highest temperature and lowest temperature regions respectively.

2.2.3 Code for Ionization and Temperature Evolution (CITE)

The temperature evolution equation for an overdense region in the IGM is given by (Hui and Gnedin 1997),

$$\frac{dT}{dt} = \left(-2HT + \frac{2T}{3\Delta}\frac{d\Delta}{dt} + \frac{dT_{shock}}{dt}\right) + \frac{T}{\sum_i X_i}\frac{d\sum_i X_i}{dt} + \frac{2}{3\,k_B\,n_b}\frac{dQ}{dt}. \qquad (2.6)$$

In the above equation the first three terms on right hand side (i.e., those in the large parenthesis) represent, respectively, the rate of cooling due to Hubble expansion, adiabatic heating and/or cooling arising from the evolution of the densities of gas particles and the change in temperature because of shocks which can be an important source of heating at low redshifts (Davé et al. 2001; Davé and Tripp 2001). These three mechanisms are taken into account in the default run of the GADGET-2. The fourth term on the right hand side represents the change in internal energy per particle arising from the change in the number of particles. The last term accounts for other heating and cooling processes, e.g., photo-heating and radiative cooling. The radiative cooling processes can, in principle, include cooling from recombinations, collisional ionization, collisional excitation, inverse Compton scattering and free-free emission.

As we discussed earlier, the default run of GADGET-2 results in temperatures that are too low at low densities, and does not show the tight $T-\Delta$ correlation at low to moderate overdensities. Therefore it is important to incorporate the effects of photoheating arising from the UVB to rectify the two problems. The method we follow to account for the photoheating and radiative cooling is as follows:

1. We start with the output snapshots at a moderately high redshift, in our case it is taken to be $z_1 = 2.1$. This is an optimum redshift for our purpose as the He II reionization is likely to be completed by then (Khaire and Srianand 2013; Kriss et al. 2001; Shull 2004; Theuns et al. 2002a; Worseck et al. 2014; Zheng et al. 2004) and thus the ionizing radiation can be taken to be uniform. If *all* the gas particles at z_1 follow a power-law equation of state, then the temperature would be given by[1]

$$\begin{aligned} T_1 \equiv T(z_1) &= T_0\ \Delta^{\gamma-1} \ \text{ for } \ \Delta < 10 \\ &= T_0\ 10^{\gamma-1} \ \text{ for } \ \Delta \geq 10, \end{aligned} \tag{2.7}$$

where T_0 and γ are free parameters. We have assumed that the high density gas with $\Delta \geq 10$ is able to cool via atomic processes and hence have temperatures smaller than what is implied by the power-law (Theuns et al. 1998b). Note that the temperature obtained using Eq. 2.7, in general, will be different from that obtained from the GADGET-2 output which we denote as $T_{g,1} \equiv T_g(z_1)$.

---

[1] We varied the Δ cutoff in Eq. 2.7 from 10 to 5 and 15 and found that the resulting $T-\Delta$ relations at low-z are not sensitive to our choice of this cutoff.

To obtain the *actual* temperature of a gas particle, we use the following argument: If $T_{g,1} > T_1$ for that particle, then it may have been shock heated in a recent time step, and hence must have moved away from the $T - \Delta$ relation. In that case the particle temperature is taken to be $T_{g,1}$. Otherwise we assume the particle temperature to be following the equation of state and assign it as T_1.

We calculate the initial fractions of different ionized species (i.e., fraction of all ionization states of hydrogen and helium and hence the fraction of free electrons) by assuming ionization equilibrium to hold at the initial redshift $z_1 = 2.1$, which is a reasonable approximation for optically thin gas in the post-He II reionization era (Becker et al. 2011; Bolton et al. 2008). At this redshift we used $\Gamma_{\rm HI}$ consistent with QSO dominated ($f_{\rm esc} = 0$) Khaire and Srianand (2015a) (hereafter KS15) UVB.

2. Given the initial temperature and the ionization fractions, it is straightforward to calculate the last two terms on the right hand side of Eq. 2.6. The expressions for the heating and cooling rates used are taken from Theuns et al. (1998b) (for similar expressions see Katz et al. 1996; Sutherland and Dopita 1993; Weinberg et al. 1997; Wiersma et al. 2009). To estimate the photo-heating rate at any given z, we used QSO dominated (i.e. $f_{\rm esc} = 0$) KS15 UVB model.

3. To obtain the particle temperature at the next redshift $z_2 = z_1 - \Delta z$, we first compare the GADGET-2 temperatures $T_{g,1}$ and $T_{g,2}$ at the two redshifts and thereby check whether the gas particle is shock heated within that time interval. If for a particle $T_{g,2} < T_{g,1}$ then the particle is not shock heated. In this case we solve the Eq. 2.6 by neglecting third term (i.e., the one corresponding to the shock heating) on the right hand side. However if the particle is shock heated[1] $T_{g,2} > T_{g,1}$ then we solve the same equation taking into account all the terms.

---

[1] We self consistently check if the particle is shock heated by solving Eq. 2.6 for GADGET-2 temperatures i.e., by using only the terms in the parenthesis on right hand side of Eq. 2.6.

4. For redshift z_2 we solve *non-equilibrium* ionization evolution equation to calculate various ionization fractions. In addition to the photoionization, we also include the collisional ionization in the non-equilibrium ionization evolution equation.

5. We then repeat the step 2 to 4 for subsequent redshifts and evolve the temperature of all the gas particles to our desired redshift.

Since the differential Eq. 2.6 is “stiff”, it tends to be numerically unstable if the time-step between two snapshots is too large. To circumvent such difficulties, we divide the time-step between two neighboring redshifts into 100 smaller steps. We linearly interpolate the GADGET-2 temperature and densities for these intermediate time-steps. We have checked the effect of varying number of time-steps on $T - \Delta$ relation and found that the results converge as long as the number of intermediate steps is 50 or more. We incorporated the above method in a module called Code for Ionization and Temperature Evolution ‘CITE’.

The *middle* and *right-hand* panels in Fig. 2.1 show the comparison between GADGET-2 temperature (T_g) and temperature from CITE (T_c). Note that the temperature scales in middle and right-hand panels are different. The highest temperature due to hydrodynamical processes is nearly same for T_g and T_c. But the lower temperature scale is considerably different due to additional processes incorporated through CITE. On an average, T_c also traces the density field shown in *left* hand panel of Fig. 2.1.

Fig. 2.2 shows the resulting distribution at $z = 0.3$ in the $T - \Delta$ plane, which is often called as the “phase diagram”. Note that the temperature and density plotted in the figure are volume-averaged, i.e., they are calculated by using the SPH kernel (see Eq. 2.4). For the plot we shoot 20000 random lines of sight through simulation box and calculate temperature and density on the grid points. The initial values at $z_1 = 2.1$ are chosen to be $T_0 = 15000$ K and $\gamma = 1.3$ (we refer to these values as initial T_0 and γ), the corresponding equation of state is shown by the red dashed line. The color coding represents density of points in logarithmic scale (i.e., the red color represents highest density of points). We can see from this figure that our simulation using CITE is able to produce the equation of state at low and moderate overdensities. In fact most of the grid points follow a power-law $T - \Delta$ relation (black dashed line) described by $T_0 = 4902$ K and $\gamma = 1.53$ (at $z = 0.3$) which is consistent with results from other low-z hydrodynamical simulations (Davé et al. 2010; Davé and Tripp 2001; Shull et al. 2015, 2012; Smith et al. 2011; Tepper-García et al. 2012). It is interesting to note that a significant fraction of points are at very high temperatures $T > 10^5$ K forming WHIM. We defer a detailed comparison of our simulations with those available in the literature to Chapter 4.

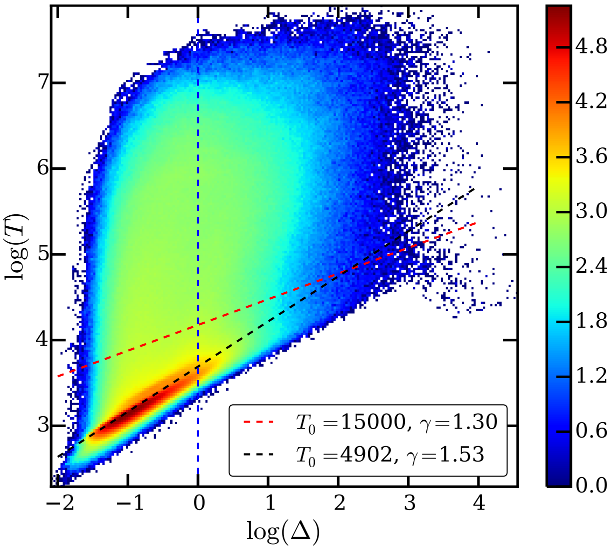

Figure 2.2: Distribution of grid points in the $T - \Delta$ plane at $z = 0.3$ when the temperatures are estimated using CITE (see Section 2.2.3 for details of CITE). The color scale indicates the density of points are shown (in logarithmic scale). At the initial redshift $z_1 = 2.1$ the values of the free parameters are chosen as $T_0 = 15000$K and $\gamma = 1.3$ (model $T15 - \gamma 1.3$ in Table 4.2), to define the effective equation of state of the IGM shown by the red dashed line. The final equation of state at $z = 0.3$ is best described by parameters $T_0 = 4902$ K and $\gamma = 1.53$ (black dashed line).

2.2.4 Generating Lyman Alpha forest Spectra in Simulation (GLASS)

We have developed a module for "Generating Ly-Alpha forest Spectra in Simulation" (GLASS) by shooting random lines of sight and storing the gas overdensities

Δ, component of velocities and the CITE temperatures T on grid points along the line of sight. Assuming that there is no significant evolution in the gas properties within the redshift bin, we splice together the lines of sight in such a way that it covers a redshift path identical to the observed spectra. The spectra of the Lyα transmitted flux (F) are generated using the procedure given in Choudhury et al. (2001) and Padmanabhan et al. (2015). There are essentially four steps in simulating the spectra: (i) The temperature, baryonic density field and peculiar velocity along the sightline is calculated using Eqs. 2.1 to 2.4 as explained in Section 2.2.2; (ii) The neutral hydrogen density ($n_{\rm HI}$) field along a sightline is obtained from the baryonic density field assuming photoionizing equilibrium with UVB; (iii) The $n_{\rm HI}$ field is then used for calculating the Lyα optical depth τ at each pixel accounting for peculiar velocity effects and the thermal and natural widths of the line profile; (iv) The transmitted flux is given simply by $F = \exp(-\tau)$. Note that the spectra thus generated depend on initial conditions of the model (at $z_1 = 2.1$ refer Table 4.2) and the photoionization rate Γ_{12} ($\Gamma_{\rm HI}$ in units of 10^{-12} s^{-1}) at the redshift of interest ($z < 0.5$). To arrive at different self-consistent combinations of the two parameters T_0 and γ at the redshift of interest, we vary these two parameters at the initial redshift $z_1 = 2.1$ and calculate the temperature for each gas particle using CITE[1] mentioned in Section 2.2.3 and Table 4.2.

In Fig. 2.3, we show a comparison of simulated absorption spectrum along a random line of sight through our simulation box (bottom panel) with observed Lyα forest spectrum towards a QSO 3C57 (Danforth et al. 2016). In order to enable fair comparison with the observational data, we resample the simulated spectra (by linear interpolation) to match the resolution of observed data and then convolve with the line spread function (LSF) of the instrument (in this case HST-COS, see Chapter 5 for details). Finally, we add random noise to each spectrum in accordance with the SNR of the observed data. Note that in

---

[1] There is an apparent inconsistency in our analysis because the temperature evolution is calculated for a fixed value of the photoionization rate (e.g., that given by KS15), while we vary the same quantity Γ_{12} at the redshift of interest treating it as a free parameter. This, however, does not affect our results as the obtained gas temperatures (at $z < 0.5$) are insensitive to the assumed value of the Γ_{12} (at $z < 2.1$).

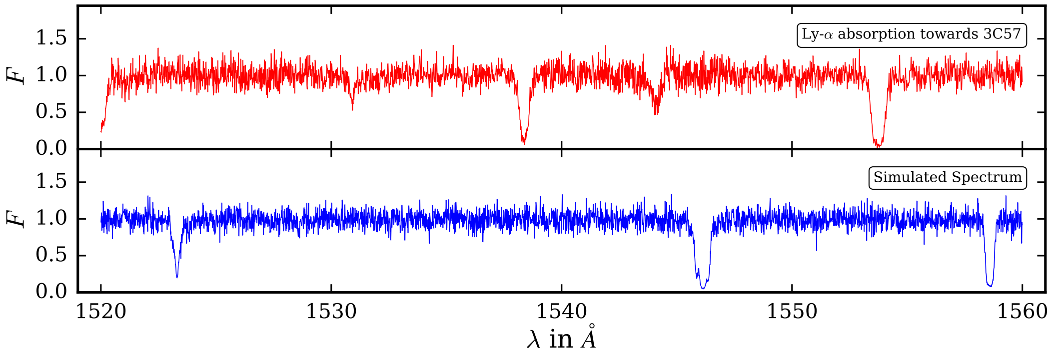

Figure 2.3: *Top panel* shows the observed Lyα forest (after the metal lines and higher order Ly series lines are removed and replaced by a continuum added with random noise) towards the QSO 3C57 (Danforth et al. 2016). *Bottom panel* shows the simulated spectrum towards a random line of sight in our simulation box. The simulated spectrum is convolved with the appropriate line spread function of HST-COS and added with noise having SNR similar to that of 3C57.

observed spectrum, the metal lines and higher order Ly series lines are removed and replaced by a continuum added with random noise of same median SNR. One can see that the simulated spectrum is qualitatively quite similar to the observed one shown in the top panel.

In Chapter 4, we show the consistency of our simulations with other low-z simulations available in the literature. Despite offering consistency and flexibility, the obvious shortcoming of CITE is that the diffuse gas is evolved dynamically at effectively zero pressure (because of its low temperature), rather than the pressure it would have if it were at $T \sim 10^4$ K typical of photoionized gas. Thus dynamical impact of diffuse IGM pressure is not modeled self-consistently in CITE. However, the consistency of CITE with other simulations discussed in Chapter 4 suggests that this is not severe shortcoming at low-z. This indicates that the evolution hydrodynamic simulation with low pressure is not distorting the properties of Lyα forest for the spatial resolution typically achieved in the low-z simulations that are used to reproduce HST-COS data.

2.3 High-z Lyα forest

While CITE works well in the low resolution simulation (~ 48 ckpc) as shown in 4, the dynamical evolution of SPH particles at finite pressure is an important effect when we consider high resolution simulations (e.g. ~ 10 ckpc, typically achieved with echelle spectrograph). In this section, we present a method to account for this effect by smoothing (in 3 dimensions) the density and velocity fields over a local Jeans scale.

2.3.1 Simulation

We again use the publicly available GADGET-2[1] (Springel 2005) to perform smoothed particle hydrodynamical simulations used in this study. The initial conditions were generated at $z = 99$ using the publicly available 2LPT[2] code (Scoccimarro et al. 2012). We use 1/30th of the mean inter-particle distance as the gravitational softening length. The GADGET-2 simulation does not include radiative heating and cooling of the SPH particles internally. As a result, the unshocked gas particles (in the low density regions) are evolved at very low temperature (~ 100 K) and pressure. However, the simulation allows one to set the minimum allowed gas temperature (referred as temperature floor) to higher values. In this work, we perform two simulations using GADGET-2: (i) G2-LTF with low temperature floor ~ 100 K and (ii) G2-HTF with high temperature floor ~ 10000 K (corresponding to typical IGM temperatures due to photoheating). An unique identification number is assigned to each particle in GADGET-2 and is used for tracing its density and temperature evolution.

We also perform a GADGET-3 simulation (a modified version of the publicly available GADGET-2 code, see Bolton et al. 2006) with the same initial conditions as the GADGET-2 simulations discussed above. Unlike GADGET-2, the GADGET-3 simulation includes radiative heating and cooling of SPH particles internally for any given UVB. To speed up the calculations, we ran the simulations with QUICK_LYALPHA flag that converts particles with $\Delta > 1000$ and $T < 10^5$ K into stars (Viel et al. 2004a) and removes them from subsequent calculations. We found that particle IDs are not unique when the QUICK_LYALPHA flag is

---

[1] http://wwwmpa.mpa-garching.mpg.de/gadget/

[2] http://cosmo.nyu.edu/roman/2LPT/

Table 2.1: Details of our simulations described in Section 2.3.1

Model	GADGET-3	G2-LTF	G2-HTF
N-body code	GADGET-3	GADGET-2	GADGET-2
Initial redshift[1]	99	99	99
Box size (h^{-1} c Mpc)	10	10	10
Number of particles	2×512^3	2×512^3	2×512^3
UVB[2]	HM12	HM12	HM12
Ionization evolution[2]	Equilibrium	Equilibrium	Equilibrium
T and Δ evolution	Internal	Post-process (CITE)	Post-process (CITE)
SFR Criteria[3]	QUICK_LYALPHA	–	–
Output redshifts	$6.0, 5.9, \cdots, 2.0$	$6.0, 5.9, \cdots, 2.0$	$6.0, 5.9, \cdots, 2.0$
Temperature floor[4]	–	~ 100 K	~ 10000 K
Smoothing kernel type[5]	SPH $W(r, h)$	Modified $W'(r, h, L_j)$	Modified $W'(r, h, 0.66 \times L_J)$

[1] All simulations (i.e. GADGET-3, G2-LTF and G2-HTF) are performed using same initial condition.

[2] The default run of GADGET-3 solves equilibrium ionization evolution equation using Haardt and Madau (2012, hereafter HM12) UVB.

[3] The QUICK_LYALPHA flag in GADGET-3 converts gas particles with $\Delta > 1000$ and $T < 10^5$K in to stars.

[4] The minimum allowed temperature of the gas particle in simulation is set by the temperature floor.

[5] To account for pressure smoothing in G2-LTF and G2-HTF model, the smoothing kernel is modified by convolving SPH kernel with Gaussian kernel of pressure smoothing in the post-processing step. The pressure smoothing is already accounted for in the default run of GADGET-3 model.

enabled. None of our simulations (i.e., GADGET-2 or GADGET-3) include AGN feedback, stellar feedback or outflows in the form of galactic wind. The details of our simulations are listed in Table 2.1.

2.3.2 Method

The Lyα optical depth is calculated by evaluating the overdensity (Δ), temperature (T) and velocity (v) on grid points along a given sightline in the simulation box. Unlike GADGET-3, the $T - \Delta$ relation obtained in GADGET-2 is not realistic as the radiative heating and cooling terms are not incorporated. At moderate to low resolution, the overdensity and velocity fields from GADGET-2 matches well with those from GADGET-3 as shown in panel (a) and (b) of Fig. 2.4. This resolution (gas particle mass $\delta m \sim 1.26 \times 10^7\, h^{-1}\, \mathrm{M}_{\odot}$, pixel size $\delta x \sim 48.8\, h^{-1}$ ckpc) is appropriate for low-z ($z < 0.5$) Lyα forest studies with instruments like the HST-COS (Gaikwad et al. 2017b,c). However GADGET-2 does not capture the effect of finite gas pressure in the hydrodynamical evolution of the photoionized gas. This effect becomes important at smaller scales probed well in high resolution (gas particle mass $\delta m \sim 1.01 \times 10^5\, h^{-1}\, \mathrm{M}_{\odot}$, pixel size $\sim 9.77\, h^{-1}$ ckpc) spectra of Lyα forest at high-z ($z > 1.6$). This is illustrated in the panel (c) and (d) of Fig. 2.4 where the density and velocity fields obtained in GADGET-3 can be seen to be smooth as compared to those in GADGET-2. Our method of evolving the gas temperature using GADGET-2 + CITE, as discussed in Gaikwad et al. (2017b), does not account for the effect of finite gas pressure on the evolution of density and velocity fields.

In this work, we present a method to account for the effect of gas pressure in GADGET-2 + CITE for high resolution Lyα forest simulations. Fig. 2.5 shows the outline of our procedure whose main steps are as follows: (1) First we estimate the temperature of the GADGET-2 particles accounting for the radiative heating and/or cooling (Gaikwad et al. 2017b). Depending on the requirements of the problem, the ionized fraction can be calculated either under equilibrium or non-equilibrium conditions. (2) We then calculate the Jeans length for each particle assuming the particles to be in local hydrostatic equilibrium (Schaye 2001). We smooth the density field by modifying the SPH kernel suitably to account for the pressure smoothing. (3) We then use the equation of state to calculate the temperature on the grids (Hui and Gnedin 1997) for particles that do not go through any shock heating. (4) Finally we calculate the Lyα optical depth using the density, velocity and temperature along the sightline (Choudhury et al. 2001). We discuss all these steps in more details below.

2.3.2.1 Temperature evolution in GADGET-2 using CITE:

We evolve the temperature of the particles in GADGET-2 using CITE as discussed in details in Section 2.2.3. However, in this case we start at initial redshift

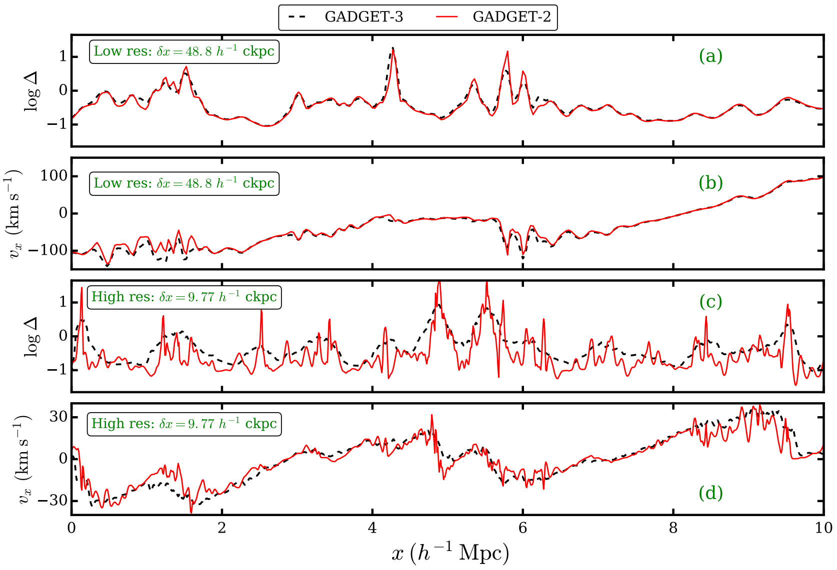

Figure 2.4: Panels (a) and (b) compare the line of sight density and velocity fields respectively from GADGET-3 (black dashed curve) and GADGET-2 (red solid curve) simulations for a low resolution simulation box at $z = 2.5$ (box size $L = 50h^{-1}$ cMpc, gas particle mass $\delta m \sim 1.26 \times 10^7\ h^{-1}\ \mathrm{M}_\odot$ and pixel size $\delta x = 48.8h^{-1}$ ckpc). Panels (c) and (d) are same as panels (a) and (b) respectively except that these are obtained from high resolution simulation boxes at $z = 2.5$ (size $L = 10h^{-1}$ cMpc, gas particle mass $\delta m \sim 1.01 \times 10^5\, h^{-1}\, \mathrm{M}_\odot$ and pixel size $\delta x = 9.77h^{-1}$ ckpc) used in this paper. GADGET-2 models for low and high resolution boxes were performed with the temperature floor of ~ 100 K.

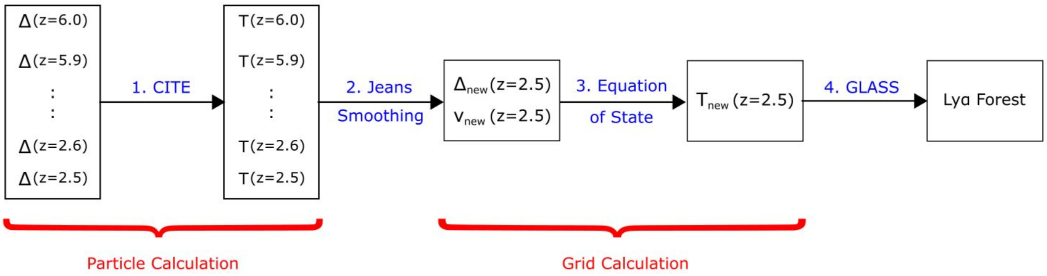

Figure 2.5: Schematic diagram showing main steps adopted in our post-processing method of obtaining Lyα forest spectra from GADGET-2 taking into account radiative cooling and heating effects externally. The basic steps involved in our method are as follows: (1) We calculate the temperature of each particle at each redshift using CITE and obtain the thermal history parameters T_0 and γ. (2) Given T and Δ of particles, we apply pressure smoothing to get new $\Delta_{\rm new}$ and $v_{\rm new}$ on grids for a simulation box at a redshift of interest. (3) For this new Δ on grid points, we apply power-law equation of state using thermal history parameters T_0 and γ obtained in the previous step. (4) We calculate Lyα optical depth from the simulation box using our routine GLASS.

$z_1 = 6.0$. We choose $T_0 = 7920$ and $\gamma = 1.52$ at $z_1 = 6.0$ in order to match with those from GADGET-3 at the same redshift for HM12 UVB. We follow the same steps given in Section 2.2.3 to obtain the temperature of the particle at lower redshifts ($2 \leq z \leq 4$). We solved both equilibrium (and non-equilibrium) ionization evolution equation and obtain the thermal history.

Fig. 2.6 shows comparison of $T - \Delta$ relation of SPH particles obtained from GADGET-3 (left panel), G2-LTF (middle panel) and G2-HTF (right panel) at $z = 2.5$ for equilibrium ionization evolution. Qualitatively, the $T-\Delta$ relation from G2-LTF and G2-HTF is remarkably similar to that from GADGET-3. The differences at $\Delta > 1000$ and $T < 10^5$ K can be attributed to the QUICK_LYALPHA flag employed in GADGET-3 (see Appendix 2.5.2 for more details). For each model, we calculate median temperature (black star points) in $\log \Delta$ bins with centers at $-0.375, -0.125, 0.125, 0.375$ and bin width 0.125 (indicated by magenta dashed vertical lines). We then fit power law relation $T = T_0\, \Delta^{\gamma-1}$ to obtain the best fit T_0 and γ (Hui and Gnedin 1997; McDonald et al. 2005). The fitted equation of state is shown by black dashed line in each panel.

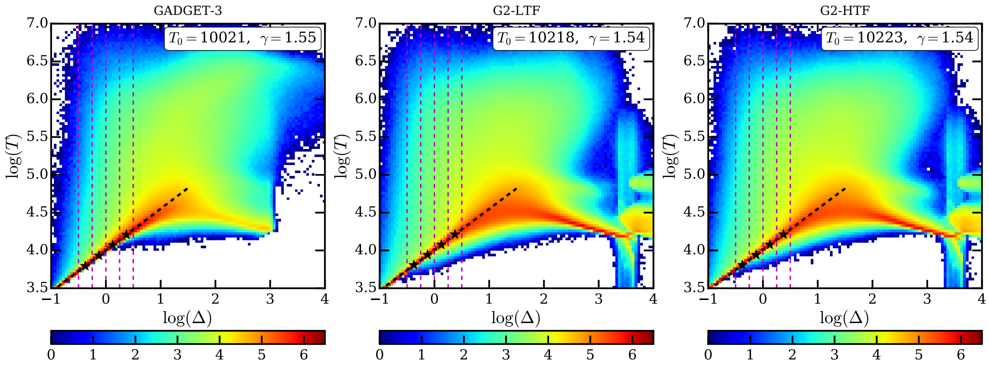

Figure 2.6: $T - \Delta$ relation of the SPH particles from GADGET-3 (left panel), G2-LTF (middle panel) and G2-HTF (right panel) at $z = 2.5$. The temperature in the G2-LTF and G2-HTF models are obtained in the post-processing step of GADGET-2 using CITE (see section 3.3). The magenta dashed vertical line shows bins in $\log \Delta$. We calculate median T (black stars) in these Δ bins and fit power-law relation $T = T_0\,\Delta^{\gamma-1}$ to obtain T_0 and γ. The resulting equation of state is shown by black dashed line. We use QUICK_LYALPHA flag in GADGET-3 which converts gas particles with $T < 10^5$ K and $\Delta > 1000$ into stars and removes from subsequent calculations. No such star formation criteria is applied in G2-LTF and G2-HTF models (see Appendix 2.5.1 for more details). The colour scheme represents density of points in logarithmic unit.

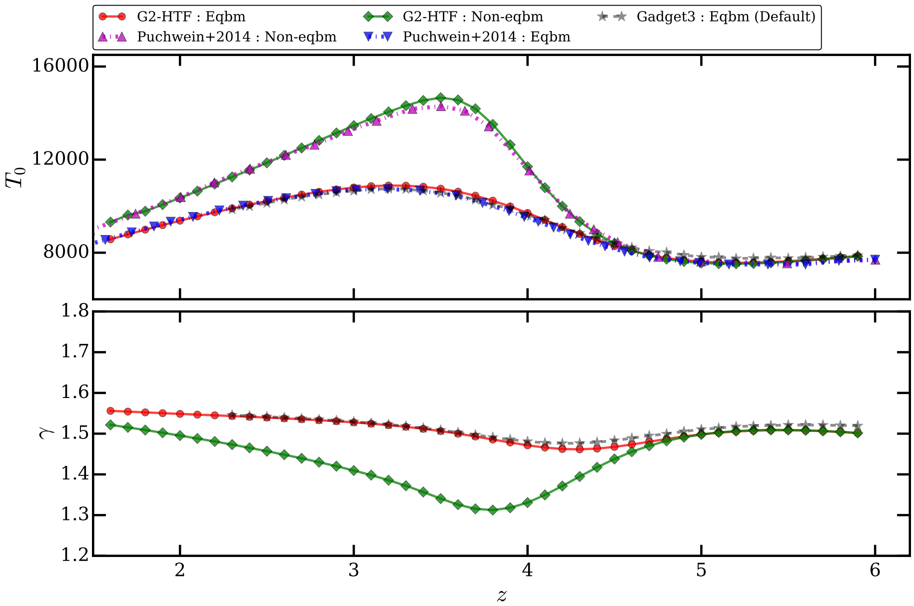

Figure 2.7: Comparison of the thermal history parameters (T_0 and γ) evolution from G2-HTF with GADGET-3 (gray stars) and Puchwein et al. (2015, magenta up-triangles for non-equilibrium and blue down-triangles for equilibrium ionization evolution). CITE is started at $z = 6.0$ with initial conditions $T_0 = 7920$ and $\gamma = 1.52$ same as those obtained in GADGET-3 at that redshift (see section 3.3 for details). We run CITE using equilibrium (red filled circles) and non-equilibrium (green diamonds) ionization evolution equation. Note that the default version of GADGET-3 solves equilibrium ionization evolution equation.

Fig. 2.7 shows the redshift evolution of best fit T_0 (top panel) and γ (bottom panel) for G2-HTF, GADGET-3 and Puchwein et al. (2015) models for equilibrium and non-equilibrium ionization evolution cases. The evolution of T_0 and γ obtained from CITE for the equilibrium ionization case is remarkably similar to those obtained from the GADGET-3 run and Puchwein et al. (2015)[1].

As mentioned earlier, we can also solve for non-equilibrium ionization evolution equation using CITE. The T_0 evolution for non-equilibrium case from Puchwein et al. (2015, magenta dashed curve) is also consistent with that from G2-HTF with the maximum difference being less than 2.5 per cent (seen at $z \sim 3.5$). Since the default version of GADGET-3 solves the ionization evolution equation under equilibrium conditions, hereafter we present the results for equilibrium ionization evolution equation while comparing our method with GADGET-3.

2.3.2.2 Jeans length of SPH particle in GADGET-2:

In this section, we explore the possibility of using local pressure smoothing in the GADGET-2 simulations to reduce the shortcomings highlighted in panels (c) and (d) of Fig. 2.4. We choose to smooth the density field in G2-LTF or G2-HTF on the scales of Jeans length of the particles to account for the pressure smoothing. Assuming the Lyα absorbers to be in local hydrostatic equilibrium, Schaye (2001) showed that the Jeans length can be obtained by equating dynamical time with sound crossing time and is given by

$$\frac{L}{1\,\mathrm{kpc}} \sim 0.52 \times \left[\frac{T}{10^4\,\mathrm{K}} \times \frac{1-Y}{0.76} \times \frac{f_g}{0.16} \times \frac{1\,\mathrm{cm}^{-3}}{n_H} \times \frac{0.59}{\mu}\right]^{1/2} \tag{2.8}$$

where T is temperature, n_H is number density of H, Y is He fraction by mass, $\mu = 4/(8-5Y)$ is the mean molecular weight (assuming fully ionized H and He) and f_g is fraction of total mass in gas phase. For the scales of interest here f_g is close to its universal value $\Omega_\mathrm{b}/\Omega_\mathrm{m} \sim 0.16$. It should be emphasized that the Jeans length depends on the density and temperature and hence is different for each particle. The above equation is not valid for Lyα absorbers with characteristic densities smaller than the cosmic mean ($\Delta \sim 1$, Schaye 2001). Hence we ignore

---

[1] The differences between the values of T_0 and γ calculated from G2-LTF and G2-HTF are less than 0.1 per cent.

the pressure smoothing for such particles and retain only the SPH smoothing. We now explain how the effect of pressure smoothing is incorporated in G2-LTF or G2-HTF by modifying the SPH kernel.

Smoothing kernel

The estimate of a quantity f at any grid point i in the SPH formulation (Monaghan 1992; Springel 2005) is given by ,

$$f_i = \sum_j f_j \frac{m_j}{\rho_j} S_{ij} \tag{2.9}$$

where the summation is performed over all particles. the quantities m_j, ρ_j, f_j are the mass, density and value of the quantity f of j^{th} particle, respectively. The quantity f could be overdensity (Δ), temperature (T) or any component of the velocity (v). The smoothing kernel, S_{ij}, has units of inverse of volume and in general depends on the distance (r_{ij}) between i^{th} grid point and j^{th} particle. It is necessary for S_{ij} to satisfy the following normalization condition in order to conserve the quantity f (in particular mass) in SPH formulation (Monaghan 1992),

$$\int\limits_{\mathcal{V}} S_{ij}\, d\boldsymbol{r} = 1 \tag{2.10}$$

where the integration is over volume $\mathcal{V}$.

We use the following smoothing kernels for various simulations,

$$S_{ij} \equiv \begin{cases} W(r_{ij}, h_j), & \text{For GADGET-3} \\ W'(r_{ij}, h_j, 1 \times L_j), & \text{For G2-LTF} \\ W'(r_{ij}, h_j, 0.66 \times L_j), & \text{For G2-HTF} \end{cases} \tag{2.11}$$

where h_j and L_j are smoothing length and Jeans length (given by Eq. 2.8) of the j^{th} particle respectively.

The smoothing kernel used for GADGET-3 is same as SPH kernel given in

Springel (2005) and has following form,

$$W(r,h) = W_0 \begin{cases} 1 - 6\left(\frac{r}{h}\right)^2 + 6\left(\frac{r}{h}\right)^3, & 0 \leq \frac{r}{h} \leq \frac{1}{2} \\ 2\left(1 - \frac{r}{h}\right)^3, & \frac{1}{2} \leq \frac{r}{h} \leq 1 \\ 0, & \frac{r}{h} > 1 \end{cases} \tag{2.12}$$

where $W_0 = 8/(\pi h^3)$ is normalization constant of SPH kernel.

The pressure smoothing can be well approximated by a Gaussian (Gnedin and Hui 1998; Kulkarni et al. 2015). Hence we modify the smoothing kernel by convolving SPH kernel with Gaussian kernel of pressure smoothing

$$W'(r,h,\sigma) = \int \mathrm{d}^3\boldsymbol{x_1}\, W(r_1,h)\, G(|\boldsymbol{r} - \boldsymbol{x_1}|,\sigma) \tag{2.13}$$

where the Gaussian kernel is assumed to be isotropic and is given by

$$\begin{aligned} G(|\boldsymbol{r} - \boldsymbol{x_1}|,\sigma) &= \frac{1}{(2\pi\sigma^2)^{3/2}}\ \exp\left[-\frac{|\boldsymbol{r} - \boldsymbol{x_1}|^2}{2\sigma^2}\right] \\ &= \frac{1}{(2\pi\sigma^2)^{3/2}}\ \exp\left[-\frac{(r^2 + r_1^2 - 2\,r\,r_1\,\mu)}{2\,\sigma^2}\right] \end{aligned} \tag{2.14}$$

with μ being the cosine of the angle between $\boldsymbol{r}$ and $\boldsymbol{x_1}$ and σ the width of the Gaussian which in turn depends on the Jeans length. At this point let us highlight some of the key properties of $W'(r,h,\sigma)$ which are relevant for our calculations:

- Both $W(r,h)$ in Eq. 2.12 and $W'(r,h,\sigma)$ in Eq. 2.13 satisfy the normalization condition given in Eq. 2.10.
- The kernel in eq. 2.13 does not have a closed form analytic solution, hence we need to calculate it numerically (See Appendix 2.5.2 for more details).
- Unlike $W(r,h)$, $W'(r,h,\sigma)$ does not have a compact support as the Gaussian is non-zero at large distances. Hence we put a cutoff such that if distance between particle and grid is more than $h+3\sigma$, the contribution of $W'(r,h,\sigma)$ is zero. Mathematically,

$$W'(r,h,\sigma) = \begin{cases} W'(r,h,\sigma) & 0 \leq r \leq (h+3\sigma) \\ 0, & r > (h+3\sigma) \end{cases} \tag{2.15}$$

We find that this cutoff does not have any significant effect on the density, velocity or temperature estimates as long as it is taken to be $\geq h + 3\sigma$.

- The amount of pressure smoothing in Eq. 2.13 is decided by the width σ of the Gaussian. The SPH particles in G2-HTF are evolved at relatively high temperature ($T \sim 10^4$ K) and pressure as compared to G2-LTF ($T \sim 100$ K). Assuming power-law $T - \Delta$ relation and using $L \propto (T/\Delta)^{1/2}$ (from Eq. 2.8), it can be shown that the additional pressure smoothing required in G2-HTF model is less than by a factor of ~ 1.5 as compared to G2-LTF model.

- This way of modifying smoothing kernel and estimating quantities along sightlines allows us to account for two important effects: (i) the variation in pressure smoothing for different particles at any epoch and (ii) the evolution of pressure smoothing scale for any particle at different epochs. Note that the pressure smoothing experienced by a particle in the GADGET-3 simulation depends on the whole thermal history and not only on the present temperature as we do in our case (Kulkarni et al. 2015; Lukić et al. 2015). However, as we will discuss later, this does not seem to have too large an effect on the Lyα forest statistics used in our analysis.

2.3.2.3 Estimation of the temperature field on a grid:

After calculating the overdensity (Δ) and velocity field (v) on a grid along a given sightline using Eqs 2.9-2.14, we can also estimate the temperature (T) along the same sightline using the same equations. However, the resultant $T - \Delta$ relation is not a power-law anymore. This is because the temperature of the particle from CITE in the first step is calculated using GADGET-2 density field that does not incorporate the pressure smoothing. Hence we need to calculate the temperature corresponding to the new density field with the pressure smoothing incorporated. In principle, we can again use CITE on the new smoothed density field and calculate the temperature. However, we found that this is computationally expensive because we need to calculate the smoothed density field on the grid along the sightline for all redshifts i.e. $z = 6.0, 5.9, \cdots, 2.6, 2.5$. Hence we adopt a simplified approach of applying power-law $T - \Delta$ relation (Choudhury et al. 2001; Hui and Gnedin 1997)

$$T = \begin{cases} T_0\, \Delta^{\gamma-1}, & \Delta \leq 10 \\ T_0\, 10^{\gamma-1}, & \Delta > 10 \\ T_{\rm shock}, & T_{\rm shock} > T \end{cases} \tag{2.16}$$

where T_0 and γ are obtained from fitting the $T - \Delta$ relation for particles as explained in Step (1) (also see Fig. 2.7). The last relation implies that if a particle is shock heated then its temperature is not updated. We note that this approach produces consistent results with running CITE on the new density field and does not make any significant difference to our results.

2.3.3 Transmitted Lyα flux

We used our module GLASS (see Section 2.2.4 for details) to calculate the transmitted Lyα flux that has signal-to-noise ratio and spectral resolution similar to the observational data. We treat the H I photoionization rate ($\Gamma_{\rm HI}$) as a free parameter. When comparing with observations, the Lyα flux field is linearly interpolated to match the wavelength sampling of observations. The Lyα flux field is then convolved with line spread function (LSF) of the spectrograph used in the observation. In this work we assume that the LSF is a Gaussian with a full width at half maximum, FWHM $\sim 6\ {\rm km\,s^{-1}}$, typical of UVES or HIRES spectra. Finally we add Gaussian random noise corresponding to a typical SNR=25 similar to what has been frequently achieved in echelle spectrographic observations with VLT and KECK that are used for Lyα forest studies.

A comparison of slices (width ~ 10 ckpc) of the overdensity ($\log \Delta$), velocity (along x axis, v_x) and temperature ($\log T$) fields on grids from a simulation box at $z = 2.5$ are shown in Fig. 2.8. The top, middle and bottom rows show slices for GADGET-3, G2-LTF and G2-HTF respectively. The $\log \Delta$, v_x and $\log T$ fields are sharper in the G2-LTF model (in particular in low density regions) as compared to the GADGET-3 model. On the other hand the $\log \Delta$, v_x and $\log T$ fields from G2-HTF model are similar to those from GADGET-3.

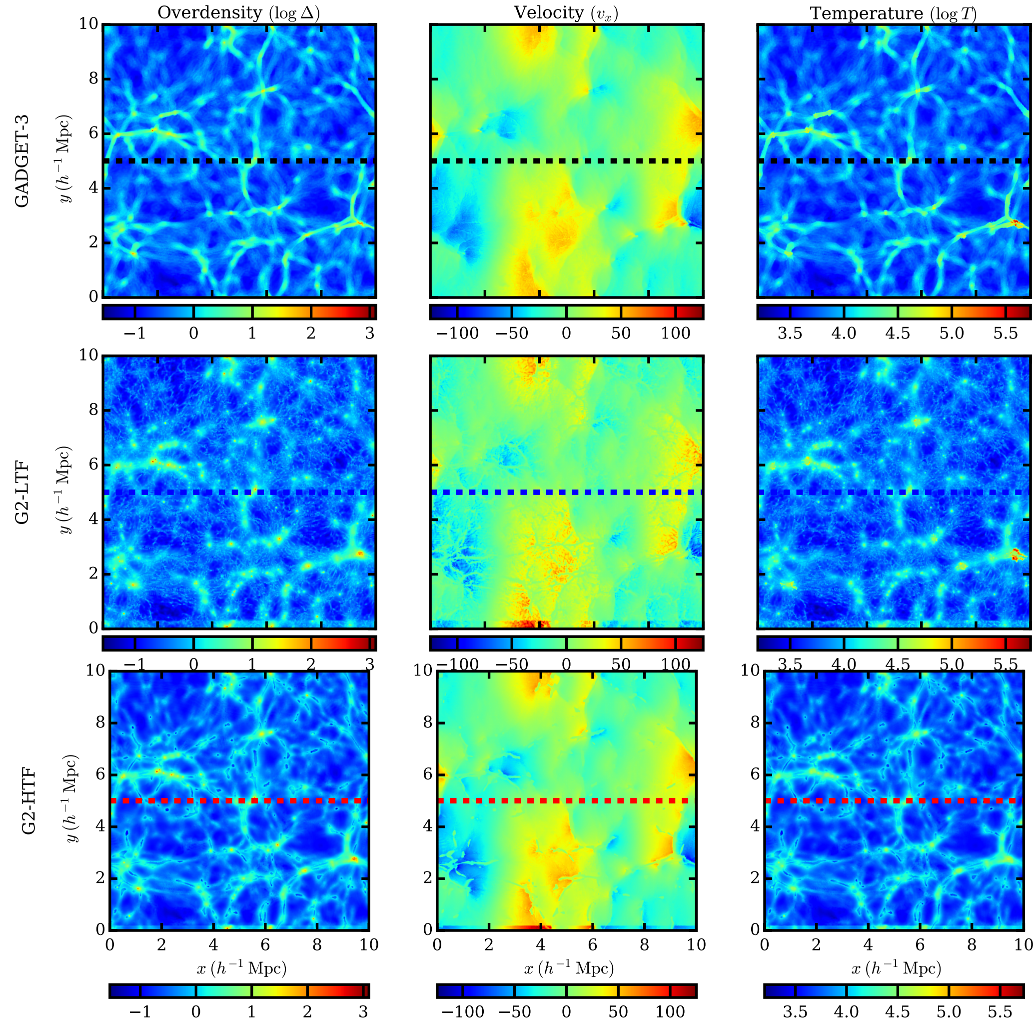

Figure 2.8: Slices of width $\sim$ 10 ckpc from a simulation box at $z = 2.5$ for GADGET-3 (top), G2-LTF (middle) and G2-HTF (bottom). Left, middle and right panels in each row show overdensity ($\log\Delta$), velocity component (v_x) along x axis and temperature ($\log T$) field respectively. The colour scheme represents density of points in logarithmic unit. We shoot a sightline parallel to x axis through simulation box in each model as shown by horizontal dashed line in each panel. We extracted $\log\Delta$, v_x and $\log T$ along this sightline for each model and plotted in Fig. 2.9 .

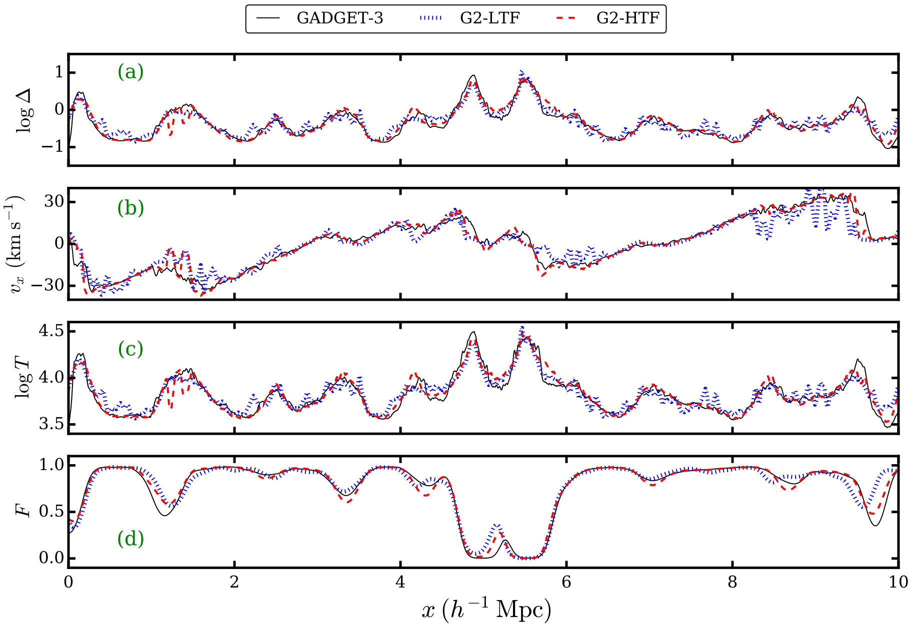

Figure 2.9: Line of sight comparison of overdensity (panel (a), $\log \Delta$), velocity (panel (b), v_x in km s^{-1}), temperature (panel (c), $\log T$) and Lyα flux (panel (d), F) for GADGET-3 (black solid line), G2-LTF (blue dotted line) and G2-HTF (red dashed line) from a simulation box at $z = 2.5$ as shown in Fig. 2.8. The Lyα flux is not convolved with any LSF and no noise is added to the flux.

We shoot a sightline from each of these slices as shown by horizontal dashed line and extract the $\log\Delta$, v_x and $\log T$ fields as shown in panel (a), (b) and (c) of Fig. 2.9 respectively. The line of sight $\log\Delta$, v_x and $\log T$ fields from G2-LTF and G2-HTF are very similar to those from GADGET-3. However, in general the variations in these fields for G2-LTF model are slightly more compared to GADGET-3 and G2-HTF models. The panel (d) of Fig. 2.9 shows the Lyα transmitted flux calculated along sightlines shown in Fig. 2.8 for GADGET-3, G2-LTF and G2-HTF models. Visually the Lyα transmitted fluxes from different models are very similar, despite the differences in $\log\Delta$, v_x, $\log T$ fields for these models. The Lyα flux in this example is not convolved with LSF and is free of noise.

We perform a quantitative comparison and consistency of the Lyα forest spectra extracted from different models using eight different statistics in Chapter 4. Treating the GADGET-3 model as the reference, we show the recovery of H I photoionization rate ($\Gamma_{\rm HI}$).

2.4 Summary

With the advent of high quality observations, an efficient method to simulate the Lyα forest would be useful for parameter estimation. Current state-of-art simulations like GADGET-3, though reproduce observational properties of Lyα forest very well, are computationally expensive for large parameter space exploration. We have developed a set of post processing modules to model the thermal, ionization history of the IGM and to generate Lyα optical depth along different sightlines. The main results of this chapter are as follows:

- We developed a new module “CITE” to evolve the temperature of the IGM from high redshift 2.1 to 0 in the post-processing step of the GADGET-2 simulation taking into account various photo-heating and radiative cooling processes. CITE provides flexibility to (i) simulate the non-equilibrium evolution of thermal and ionization state of the gas, (ii) incorporate the radiative cooling for a wide range of metallicities and (iii) include heating due to non-standard sources like cosmic rays and high energy γ-rays from Blazars.

- We further develop a module GLASS to calculate the Lyα transmitted flux that has SNR, spectral resolution and line spread function (LSF) effects similar to the observational data. The spectra generated using our method are remarkably similar to the spectra observed using HST-COS at $z < 0.5$.

- For the resolution used in the above study (at $z < 0.5$, gas particle mass $\delta m \sim 1.26 \times 10^7 \ h^{-1}$ M$_\odot$, pixel size $\delta x \sim 48.8h^{-1}$ ckpc), the pressure smoothing of baryons is not a major issue. However, for studying the high-z ($2 \leq z \leq 4$) Lyα forest one usually uses higher resolution echelle data. When we use appropriate high resolution (gas particle mass $\delta m \sim 1.01 \times 10^5 \, h^{-1}$ M$_\odot$, pixel size $\delta x \sim 9.77$ ckpc) simulation boxes, we notice that the density (Δ) and velocity (v) fields are smoother for GADGET-3 as compared to those from GADGET-2. This is because (i) the temperature of the SPH particles in GADGET-2 is not calculated self-consistently (photo-heating and radiative cooling terms are not accounted for), (ii) the gas (in particular the unshocked gas) is effectively evolved at zero pressure in GADGET-2 and (iii) the local Jeans length of the particles, responsible for pressure smoothing in GADGET-3, is comparable to the resolution of the simulation box.

- In this chapter, we show that by running a GADGET-2 simulation with elevated temperature floor and using local Jeans smoothing we are able to circumvent the above mentioned shortcomings of our method in the high resolution simulations. The basic idea is to apply additional smoothing in GADGET-2 by a local Jeans length at the epoch of our interest. However, it is well known that the smoothing in GADGET-3 is not only decided by the instantaneous density and temperature of the particles but also to some extent by the thermal history of the particles. To illustrate this, we perform three high resolution simulations (gas particle mass $\delta m \sim 1.01 \times 10^5 \, h^{-1}$ M$_\odot$, pixel size $\delta x \sim 9.77$ ckpc) with same initial conditions (i) G2-LTF: GADGET-2 with low temperature ($T \sim 100$ K) floor in which local Jeans length is decided by instantaneous density and temperature and (ii) G2-HTF: GADGET-2 with high temperature ($T \sim 10^4$ K) floor in which even the unshocked gas is evolved at with a pressure appropriate for a photoionized gas at $T = 10^4$ K and (iii) GADGET-3: a reference model for comparison with G2-LTF and G2-HTF model.

- For G2-LTF and G2-HTF models, we first estimate the temperature of SPH particles in GADGET-2 using our code CITE. We modify the smoothing kernel to account for pressure smoothing and estimated the density, velocity field on grids. We found that the line of sight density and velocity from our method matches well with that from GADGET-3. We generate the Lyα forest spectra by shooting random sightlines through simulation box in all the 3 models. The resulting Lyα forest spectra along sightline are remarkably similar in the G2-HTF and GADGET-3 methods.

- We obtain the evolution of thermal history parameters T_0 and γ by estimating the temperature of the SPH particles from CITE. The T_0 and γ evolution from G2-HTF and G2-LTF is in very good agreement with that from GADGET-3. CITE also provides us with enough flexibility to solve the non-equilibrium ionization evolution equation. The T_0 and γ evolution for non-equilibrium case is considerably different (T_0 is larger by ~ 60 percent and γ is smaller by 15 percent at $z = 3.7$) than that for equilibrium case. The T_0 evolution for non-equilibrium case from our method is consistent with that from Puchwein et al. (2015, difference less than 2.5 percent).

2.5 Appendices

2.5.1 Star formation criteria

To speed up the calculations in GADGET-3, we use QUICK_LYALPHA flag that converts gas with $\Delta > 1000$ and $T < 10^5$ K into stars. In order to study its effect on our method, we apply the same criteria to the G2-HTF model. The

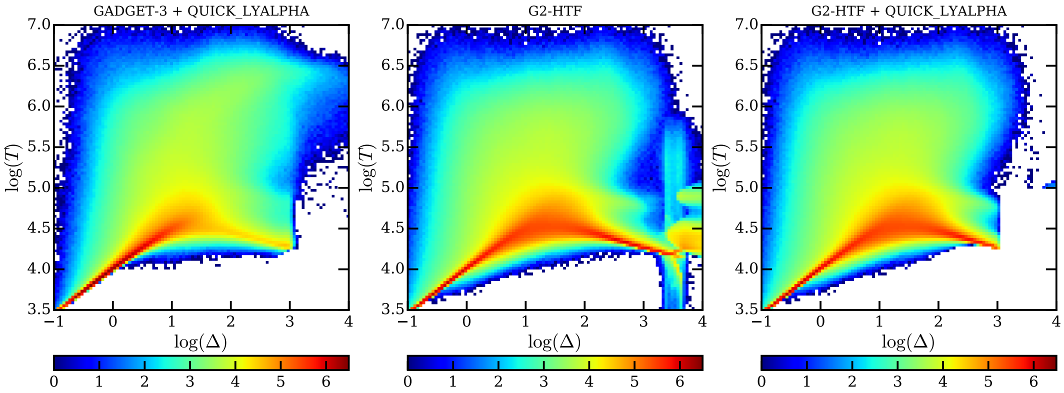

Figure 2.10: Left, middle and right panel shows the $T - \Delta$ relation for GADGET-3, G2-LTF and G2-LTF with SFR criteria used in QUICK_LYALPHA setting of GADGET-3. Particles with $\Delta > 1000$ and $T < 10^5$K are treated as stars and removed from further calculation. The $T - \Delta$ relation looks remarkably similar for GADGET-3 and G2-LTF with QUICK_LYALPHA setting.

left, middle and right panels in Fig. 2.10 show the $T - \Delta$ relation for GADGET-3, G2-HTF model without star formation criteria and G2-HTF model with star formation criteria similar to QUICK_LYALPHA respectively. The $T - \Delta$ relation for GADGET-3 and G2-HTF model with star formation are remarkably similar even at $\Delta > 1000$. We also generate Lyα forest from G2-HTF model with star formation and calculated various Lyα statistics. We found that the Lyα statistics are accurate to within 1.8 percent suggesting QUICK_LYALPHA is a good approximation. This is because the particles converted in to stars occupy small volume in the simulation box. The probability that a random sightline (along which Lyα optical depth is calculated) intersecting such region is small.

2.5.2 Convolution of SPH kernel with Gaussian Kernel

In this section, we show that the convolution integral in Eq. 2.13 can be recast in to an analytical form that is fast and easy to implement on computers. Let $W(r, h)$ be SPH kernel and $G(r, \sigma)$ be Gaussian kernel of pressure smoothing. Let $\widetilde{W}(k, h)$ and $\widetilde{G}(k, \sigma)$ be the Fourier transforms of $W(r, h)$ and $G(r, \sigma)$ respectively.

The convolution of $W(r,h)$ with $G(r,\sigma)$ is given by,

$$\begin{aligned}
\widetilde{W'}(k,h,\sigma) &= \widetilde{W}(k,h) \times \widetilde{G}(k,\sigma) \\
W'(r,h,\sigma) &= \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3}\ \widetilde{S}(\boldsymbol{k},h,\sigma)\ \mathrm{e}^{\,i\,\boldsymbol{k}\cdot\boldsymbol{r}} \\
W'(r,h,\sigma) &= \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3}\ \widetilde{W}(\boldsymbol{k},h) \times \widetilde{G}(\boldsymbol{k},\sigma)\ \mathrm{e}^{\,i\,\boldsymbol{k}\cdot\boldsymbol{r}}
\end{aligned} \tag{2.17}$$

Using the convolution theorem,

$$W'(r,h,\sigma) = \int \mathrm{d}^3\boldsymbol{x_1}\, W(r_1,h)\ G(|\boldsymbol{r}-\boldsymbol{x_1}|,\sigma) \tag{2.18}$$

The SPH kernel is given in Eq. 2.12. The Gaussian kernel in Eq. 2.14 can be written in following form,

$$\begin{aligned}
G(|\boldsymbol{r}-\boldsymbol{x_1}|,\sigma) &= \frac{1}{(2\pi\sigma^2)^{3/2}}\ \exp\left[-\frac{|\boldsymbol{r}-\boldsymbol{x_1}|^2}{2\sigma^2}\right] \\
&= \frac{1}{(\pi b^2)^{3/2}}\ \exp\left[-\frac{|\boldsymbol{r}-\boldsymbol{x_1}|^2}{b^2}\right] \\
&= G_0\ \exp\left[-\frac{|\boldsymbol{r}-\boldsymbol{x_1}|^2}{b^2}\right] \\
&= G_0\ \exp\left[-\frac{(r^2+r_1^2-2\,r\,r_1\,\mu)}{b^2}\right]
\end{aligned} \tag{2.19}$$

where, $b = \sqrt{2}\sigma$ and μ is cosine of angle between vector $\boldsymbol{r}$ and $\boldsymbol{x_1}$. The convolution integral can be recast in to the following form,

$$\begin{aligned}
W'(r,h,\sigma) = W_0'\Bigg[\ \sum_{n=0}^{4} \mathrm{A}_n^+\ \mathrm{I}_n\left(\frac{-r}{b},\frac{h/2-r}{b}\right) + \ \mathrm{A}_n^-\ \mathrm{I}_n\left(\frac{r}{b},\frac{h/2+r}{b}\right) \\
+\ \mathrm{C}_n^+\ \mathrm{I}_n\left(\frac{h/2-r}{b},\frac{h-r}{b}\right) + \ \mathrm{C}_n^-\ \mathrm{I}_n\left(\frac{h/2+r}{b},\frac{h+r}{b}\right)\Bigg]
\end{aligned} \tag{2.20}$$

where,

$$
\begin{aligned}
W_0' &= \frac{\pi b^2\, G_0\, W_0}{r} \\
\mathrm{A}_0^{\pm} &= \pm\frac{6\,b\,r^4}{h^3} \;-\; \frac{6\,b\,r^3}{h^2} \;+\; b\,r \\
\mathrm{A}_1^{\pm} &= \frac{24\,b^2\,r^3}{h^3} \;\mp\; \frac{18\,b^2\,r^2}{h^2} \;\pm\; b^2 \\
\mathrm{A}_2^{\pm} &= \pm\frac{36\,b^3\,r^2}{h^3} \;-\; \frac{18\,b^3\,r}{h^2} \\
\mathrm{A}_3^{\pm} &= \frac{24\,b^4\,r}{h^3} \;-\; \frac{6\,b^4}{h^2} \\
\mathrm{A}_4^{\pm} &= \pm\frac{6\,b^5}{h^3}
\end{aligned}
\tag{2.21}
$$

$$
\begin{aligned}
\mathrm{C}_0^{\pm} &= \mp\frac{2\,b\,r^4}{h^3} \;+\; \frac{6\,b\,r^3}{h^2} \;\mp\; \frac{6\,b\,r^2}{h} \;+\; 2\,b\,r \\
\mathrm{C}_1^{\pm} &= -\frac{8\,b^2\,r^3}{h^3} \;\pm\; \frac{18\,b^2\,r^2}{h^2} \;-\; \frac{12\,b^2\,r}{h} \;\pm\; 2\,b^2 \\
\mathrm{C}_2^{\pm} &= \mp\frac{12\,b^3\,r^2}{h^3} \;+\; \frac{18\,b^3\,r}{h^2} \;\mp\; \frac{6\,b^3}{h} \\
\mathrm{C}_3^{\pm} &= -\frac{8\,b^4\,r}{h^3} \;\pm\; \frac{6\,b^4}{h^2} \\
\mathrm{C}_4^{\pm} &= \mp\frac{2\,b^5}{h^3}
\end{aligned}
\tag{2.22}
$$

$$
\begin{aligned}
\mathrm{I}_n(l_1, l_2) &= \int_{l_1}^{l_2} t^n\ \mathrm{e}^{-t^2}\ dt \\
\mathrm{I}_0(l_1, l_2) &= \frac{\sqrt{\pi}}{2}\ \Big[\ \mathrm{erf}\ (t)\ \Big]_{l_1}^{l_2} \\
\mathrm{I}_1(l_1, l_2) &= \Big[\ -\frac{\mathrm{e}^{-t^2}}{2}\ \Big]_{l_1}^{l_2} \\
\mathrm{I}_2(l_1, l_2) &= \frac{1}{2}\ \mathrm{I}_0(l_1, l_2) - \frac{1}{2}\ \Big[t\ \mathrm{e}^{-t^2}\Big]_{l_1}^{l_2} \\
\mathrm{I}_3(l_1, l_2) &= \quad \mathrm{I}_1(l_1, l_2) - \frac{1}{2}\ \Big[t^2\ \mathrm{e}^{-t^2}\Big]_{l_1}^{l_2} \\
\mathrm{I}_4(l_1, l_2) &= \frac{3}{2}\ \mathrm{I}_2(l_1, l_2) - \frac{1}{2}\ \Big[t^3\ \mathrm{e}^{-t^2}\Big]_{l_1}^{l_2}
\end{aligned}
\tag{2.23}
$$

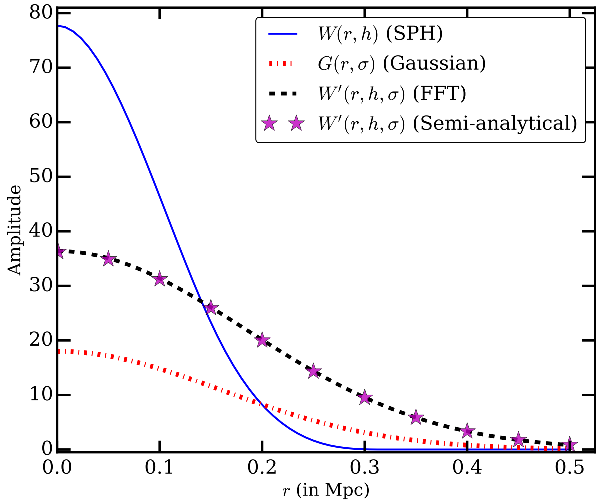

Figure 2.11: The SPH kernel (see Eq. 2.12) and Gaussian kernel (Eq. 2.14) for a particle are shown by blue solid curve and red dashed-dot curve respectively. These two kernels are convolve using FFT based method as shown by black dashed line. Red stars shows our the semi-analytical convolution approximation (given in Eq. 2.20). Our method of approximation is accurate within 2 percent of FFT based method.

These integrals involve error function and hence need to be evaluated numerically. To speed up the calculations we used error function approximation of the following form (Abramowitz and Stegun 1972)

$$\mathrm{erf}(x) = 1 - (a_1\, t + a_2\, t^2 + a_3\, t^3 + a_4\, t^4 + a_5\, t^5)\, \mathrm{e}^{-x^2} + \epsilon(x) \tag{2.24}$$

where,

$$\begin{gathered} t = \frac{1}{1 + p\, x} \quad \text{and} \\ |\epsilon(x)| \leq 1.5 \times 10^{-7} \end{gathered} \tag{2.25}$$

The small value of $|\epsilon(x)|$ indicates that the uncertainty in error function approximation is negligible. The values of the constants are

$$\begin{aligned} p &= 0.3275911 & a_1 &= 0.254829592 \\ a_2 &= -0.284496736 & a_3 &= 1.421413741 \\ a_4 &= -1.453152027 & a_5 &= 1.061405429 \end{aligned}$$

Eq. 2.13 can also be solved numerically using a 3D FFT based method. However, we found that this method is computationally expensive for large number of particles. Fig. 2.11 shows a comparison of semi-analytical approximation (red stars, given in Eq. 2.20) with FFT based method (black dashed curve) for a particle. The SPH kernel and Gaussian kernel for this particle are shown by blue solid curve and red dash dot curves respectively. For visual purpose the Gaussian kernel is rescaled to fit the graph. Our method of approximation for convolution given in Eq. 2.20 is accurate within 2 percent of FFT based method.

3 | VoIgt profile Parameter Estimation Routine

This chapter contains material that has been published in Gaikwad et al. (2017c).

3.1 Introduction

The observations of the Lyα forest have regularly been used to constrain cosmological and astrophysical parameters related to IGM physics. The strength and width of the Lyα absorption lines can be used to trace the ionization and thermal state of the neutral hydrogen (H I) in the IGM. Various statistics are used in the literature to constrain cosmological and astrophysical parameters from the Lyα forest observations. These statistics are broadly divided into two cases.

In the first case, Lyα transmitted flux is treated as a continuous field quantity. In particular, the mean flux, the flux probability distribution function (PDF) and the flux power spectrum (PS) have been used to constrain cosmological parameters such as $\Omega_{\rm m}$, $\Omega_{\rm b}h^2$, σ_8 and $n_{\rm s}$ (Choudhury et al. 2001; McDonald et al. 2000, 2005; Phillips et al. 2001; Seljak et al. 2006; Tegmark et al. 2004; Viel et al. 2009; Viel and Haehnelt 2006; Viel et al. 2006, 2004a,b), thermal history parameters[1] (Becker et al. 2011; Bolton et al. 2008; Calura et al. 2012; Garzilli et al. 2015; Lidz et al. 2010; Zaldarriaga et al. 2001) and H I photoionization rate ($\Gamma_{\rm HI}$; Becker and Bolton 2013; Bolton and Haehnelt 2007; Gaikwad et al. 2017b; McQuinn et al. 2011; Meiksin and White 2004; Pontzen et al. 2014; Rauch et al. 1997). Constraining such quantities by comparing flux statistics between observations and

---

[1] The thermal state of the IGM is described by the effective equation of state parameterized by the mean IGM temperature (T_0) and the slope (γ).

simulations are relatively easier and are frequently used in the high-z ($2 \leq z \leq 6$) studies.

In the second case, Lyα forest is decomposed into multiple Voigt profiles. The line width distribution function calculated from Voigt profile fitting is sensitive to the thermal history and the energy injected by various astrophysical processes in the form of heat and turbulent motions in the IGM (Davé et al. 2001; McDonald et al. 2001; Schaye et al. 1999, 2000). Similarly, the column density distribution function (CDDF) calculated from Voigt profile decomposition is sensitive to $\Gamma_{\rm HI}$ (Cooke et al. 1997; Gurvich et al. 2017; Kollmeier et al. 2014; Shull et al. 2015) and cosmological parameters (Penton et al. 2000; Shull et al. 2012; Storrie-Lombardi et al. 1996). While statistics based on parameters obtained using Voigt profile fitting are useful in deriving thermal history and equation of state of the IGM, the Voigt profile decomposition is usually a time consuming process. Therefore, a large parameter space exploration in simulations is usually difficult.

In order to obtain the CDDF, each Lyα absorption is usually decomposed into multiple Voigt profile components. Each Voigt profile is defined by 3 free parameters i.e., line center ($\lambda_{\rm c}$), H I column density ($\rm N_{HI}$) and line width parameter (b). The manual Voigt decomposition of the large number ($\sim$ 10000) of simulated Lyα forest is laborious and time consuming. Furthermore, the criteria used to fit the number of components ($N_{\rm Voigt}$) to a given identified Lyα absorption is subjective and need not be unique. Although there are several Voigt profile fitting codes available in the literature like VPFIT[1], ALIS (Cooke et al. 2014), GVPFIT (Bainbridge and Webb 2016) it will be invaluable to have a tailor made automatic module that will identify Lyα absorption regions and fit them with multiple component Voigt profiles where the best fit parameters of the individual components and the minimum number of required components are determined through objective criterion.

We have developed a parallel processing module "VoIgt profile Parameter Estimation Routine" (VIPER) to fit the Lyα forest with multiple Voigt profiles automatically. In VIPER, the blended and saturated features are fitted simultaneously with multi-component Voigt profiles. An objective criteria based on information theory is used to find the number of Voigt profiles needed to describe the Lyα

---

[1] `http://www.ast.cam.ac.uk/~rfc/vpfit.html`

forest. The parallel and automated nature of VIPER allows us to simultaneously fit large number of simulated spectra and to explore a wide parameter space efficiently. For consistency, we used the same code for analyzing the observed and simulated spectra in chapter 5. We calculated CDDF by consistently taking into account the redshift path length and the incompleteness of the observed sample. We show CDDF and line width distribution obtained for observed data using VIPER matches very well with those from Danforth et al. (2016).

The paper is organized as follows. In section 3.2, we explain the HST-COS data used in this work. Section 3.3 describes our module VIPER that automatically fits Voigt profiles to the Lyα forest. In section 3.4, we show the consistency of VIPER by comparing CDDF and linewidth distribution for the observed data from literature.. Finally we summarize our results in section 3.5.

3.2 HST-COS Data

We used publicly available HST-COS science data product[1] that consists of a sample of low redshift Lyα forest spectra towards 82 UV bright QSOs performed by Danforth et al. (2016, hereafter D16). These QSOs are distributed in the redshift range $z = 0.0628$ to 0.852. The sample covers the Lyα forest in the redshift range $0 \leq z \leq 0.48$ with a velocity resolution of $\sim 17\ \mathrm{km\,s^{-1}}$ (full width at half maximum). The median signal-to-noise ratio (SNR) per pixel varies from 6 to 17 for different sightlines.

D16 fitted the continuum to each spectrum and identified several thousand absorption line features that consists of Lyα lines, higher order Ly-series lines, metal lines from the IGM and the interstellar medium (ISM) of our Galaxy. As a part of the HST-COS science data product, D16 fitted each absorption feature with multiple component Voigt profiles and provided a table that contains line identifications (type of the specie and rest wavelength of the transition), redshift of the absorption system, column density, doppler-b parameter, equivalent width (along with associated fitting errors) and significance level of the absorption line detection. In this work we refer to their Lyα line catalog as “D16 line catalog”.

---

[1] `https://archive.stsci.edu/prepds/igm/`

As in Paper-I, we divided the sample into 4 different redshift bins (we denote them by roman numerals I, II, III and IV respectively) $\bar{z} \pm \Delta z \equiv (0.1125 \pm 0.0375, 0.2 \pm 0.05, 0.3 \pm 0.05, 0.4 \pm 0.05)$. The size and center of the lowest redshift bin is chosen in a way that avoids the contamination by geo-coronal line emission at $z \leq 0.075$. There are total $50, 31, 16$ and 12 lines of sight in the redshift bins I, II, III and IV respectively. Apart from the intervening Lyα lines all other lines in the Lyα forest are treated as contamination in our spectra. We replace all other lines except Lyα lines by the continuum added with a Gaussian random noise (see Fig. 2 in Paper-I). We use these clean spectra for further analysis.

3.3 Automatic Voigt profile fitting code

In this section we describe our automated Voigt profile fitting procedure "VoIgt profile Parameter Estimation Routine" (VIPER). The same code has been used to fit the observed and simulated Lyα forest spectra to constrain Γ_{12} in chapter 5 from CDDF. The algorithm is broadly divided into 3 steps; first we identify the absorption lines and region bracketing these lines, next in these regions we fit as many Voigt components as necessary based on an objective criteria. In the final step we accept the Voigt profile fit for a line based on a significance level of the fit. We now discuss each step in detail below.

1. **Line and region identification :** Following Schneider et al. (1993) and D16, first we estimate the "crude significance level" (hereafter CSL) to identify the lines,

$$\mathrm{CSL} = \frac{W(\lambda)}{\overline{\sigma}(\lambda)} \tag{3.1}$$

where $W(\lambda)$ and $\overline{\sigma}(\lambda)$ are "equivalent width vector" and "line-less error vector" respectively. The CSL defined in this way has the advantage that the unresolved features are unlikely to be identified as lines (see Schneider et al. 1993, for details). $W(\lambda)$ and $\overline{\sigma}(\lambda)$ are obtained by convolving the normalized flux and line-less error (i.e., error on flux if absorption lines were absent), respectively, with a representative line profile (see section 2.3 of D16). The representative line profile is a convolution of Gaussian (of a Doppler parameter of $b = 20\ \mathrm{km\,s^{-1}}$) with HST-COS line spread function

(LSF). We repeated the procedure with different values of the doppler $b = 50, 100\ \mathrm{km\,s^{-1}}$ to incorporate any missing narrower, broader and blended lines. The second panel from the top in Fig. 3.1 shows the CSL estimated for the spectrum shown in the top panel. Initially we identified all the lines with maxima satisfying CSL $\geq$ 1.5 shown by red stars in second panel of Fig. 3.1[1]. Next we find a threshold on either side of the maxima to enclose the line in a region. A minimum is accepted as a threshold if CSL $<$ 1.5. A minimum with CSL $>$ 1.5 indicates that the lines are blended hence we search for the next minimum until we meet the condition CSL $<$ 1.5 to accept it as a threshold. We then merge the overlapping regions (if any) into one bigger region for blended lines (e.g., see the yellow shaded region in second panel of Fig. 3.1). This procedure allows us to identify and fit the blended lines simultaneously.

2. **Voigt profile fitting :** In this step we fit each identified region by multiple Voigt profiles. Voigt profile, convolution of Gaussian with Lorentzian, is a real part of the "Faddeeva function" $w(z)$ (Armstrong 1967),

$$\begin{aligned} w(z) &= e^{-z^2}\,\mathrm{erfc}(-iz) \\ w(x+iy) &= V(x,y) + iL(x,y) \end{aligned} \tag{3.2}$$

where erfc(−iz) is the error function, $V(x,y)$ is Voigt profile and $L(x,y)$ is imaginary part of Faddeeva function. We used *wofz*[2] function in *python's scipy* package to compute Voigt profile. We convolve this Voigt profile with the appropriate HST-COS LSF before performing χ^2 minimization[3]. For χ^2 minimization, we used *leastsq*[4] function in *python's scipy* package. If F_{obs}, σ_{obs}, $F_{\mathrm{fit}}(\lambda_{\mathrm{c}}, b, \mathrm{N_{HI}})$ are observed flux, error in the observed flux and

---

[1] The cutoff used in this work is smaller than that used by D16 (CSL $\geq$ 3). This results in more number of identified lines in our initial line catalog as compared to D16. However, in the final step most of the extra identified lines at lower significance level are rejected.

[2] `http://ab-initio.mit.edu/wiki/index.php/Faddeeva_Package`

[3] We assumed that the observed fluxes in different pixels are uncorrelated. HST-COS LSF is not a Gaussian (`http://www.stsci.edu/hst/cos/performance/spectral_resolution/`). This function is slightly asymmetric around the center and has extended wings.

[4] `https://docs.scipy.org/doc/scipy-0.18.1/reference/generated/scipy.optimize.leastsq.html`

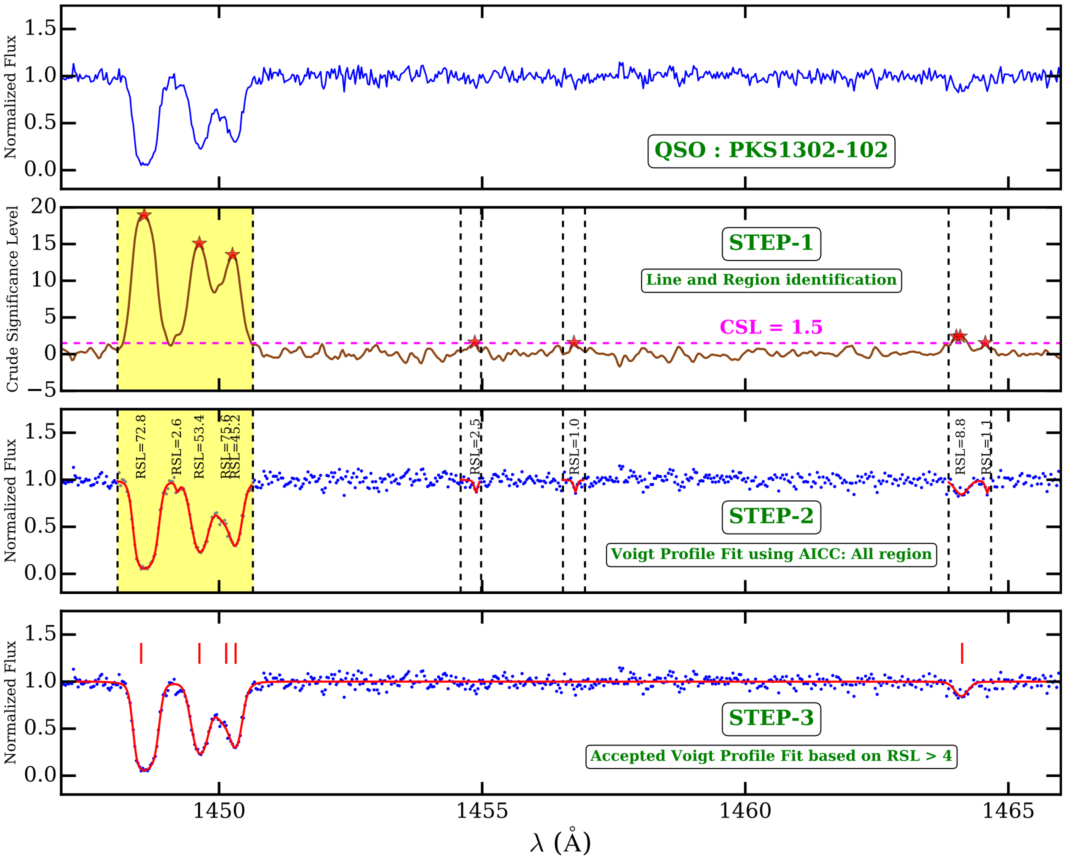

Figure 3.1: Illustration of different steps in the automatic Voigt profile fitting procedure used in VIPER. *Top* panel shows a portion of the observed HST-COS spectrum along the sightline towards QSO PKS1302-102. *Second* panel from top shows the estimation of crude significance level (CSL) using the Eq.3.1 (for $b = 20$ $\mathrm{km\,s^{-1}}$). All the identified peaks with CSL ≥ 1.5 (magenta dashed line) are shown by red stars. The identified regions enclosing the peaks are shown by black dashed vertical lines. Overlapping regions are merged accordingly to fit blended lines simultaneously (see yellow shaded region). All the identified regions are fitted with Voigt profile as shown in the third panel from top. The number of components used to fit the region is decided using AICC and demanding $\chi^2_{\mathrm{dof}} \sim 1$ (see section 3.3). Rigorous significance level (RSL) for each fitted line is calculated using Eq.3.4. *Bottom* panel shows the accepted fit with the RSL ≥ 4.

fitted Lyα flux respectively then we minimize the function $(F_{\rm obs} - F_{\rm fit})/\sigma^2_{\rm obs}$ in *leastsq* routine. The fit parameters are allowed to vary over the range $5 \leq b\,({\rm km\,s^{-1}}) \leq 150$, $10 \leq \log({\rm N_{HI}/cm^{-2}}) \leq 16.5$ and $\lambda_{\rm c}$ bounds are set by the wavelength of the region. We set initial guess values for lines in a given region by fitting individual line in that region with a Gaussian. Note that in the previous step each identified line is enclosed in two CSL minima hence we can fit each line separately. However initial guess values for any additional line required by the information criteria (explained below) are set randomly in the region.

A criteria based on information theory, Akaike Information Criteria with Correction (AICC) (Akaike 1974; King et al. 2011; Liddle 2007) is used to assess the optimum number of Voigt profile components required for an acceptable fit. If p is the number of parameters in a model used to fit the data with n pixels, then the AICC is given by,

$$\mathrm{AICC} = \chi^2 + \frac{2\,p\,n}{(n-p-1)} \quad . \tag{3.3}$$

The first term on right-hand side is a measure of loss of information while describing the data with a model. The second term in right-hand side quantifies the complexity of the model. Thus AICC incorporates the trade-off between loss of information and complexity of the model. We assign a model to be the best fit model over the previously assigned best fit model if AICC is lower by at least 5 (Jeffreys 1961). Since only the relative difference in the AICC values is important, according to Jeffreys (1961) $\Delta\mathrm{AICC} = 5$ is considered as the strong evidence against the weaker model. Fig. 3.2 illustrates our method of choosing a best fit model. Black points in left-hand panel of the Fig. 3.2 are data points which we want to fit by Voigt profile model. We fitted the data with different number of (say $N_{\rm Voigt} = 1, 2, 3, 4, 5$) Voigt profiles. The resulting AICC and χ^2 for each model as a function of $N_{\rm Voigt}$ is shown by star and circle respectively in the right-hand panel of the Fig. 3.2. For $N_{\rm Voigt} = 1$ the model is less complex but the χ^2 ($\chi^2_{\rm dof} \sim 1.5$) between model and data is large (see green curve in left-hand

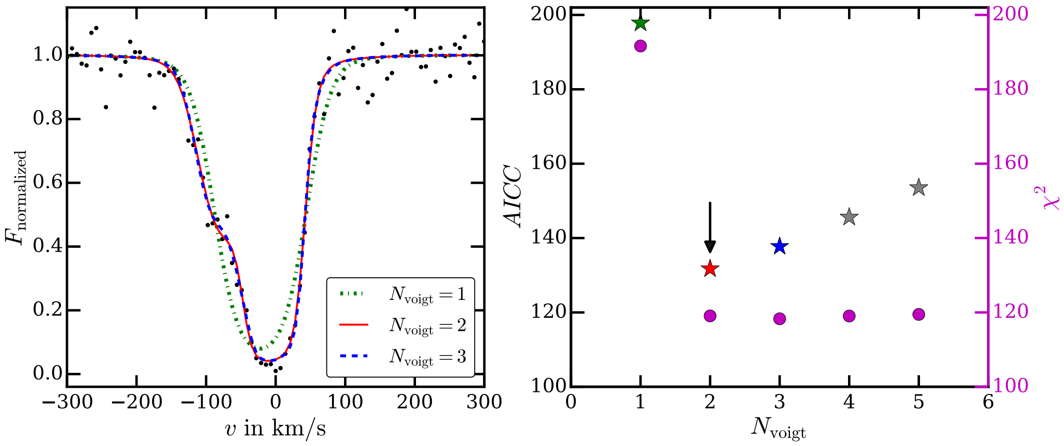

Figure 3.2: *Left-hand* panel shows the three different Voigt profile fits with $N_{\rm Voigt}=1,2$ and 3 (green dot-dashed, red continuous and blue dashed lines respectively) fitted to the observed data (black circle). The spectrum is shown in the velocity scale defined with respect to the redshift of the strongest line center. *Right-hand* panel shows the corresponding variation of AICC (stars) and χ^2 (magenta circles) for 5 different models. For legibility fits with $N_{\rm Voigt}=4,5$ (gray star points) are not shown in left-hand panel. For $N_{\rm Voigt}>2$, the χ^2 remains constant whereas AICC increases due to the second term on right-hand side of Eq.3.3. The best fit model corresponds to the minimum AICC (where $\chi^2_{\rm dof}\sim 1$ is also achieved) i.e., $N_{\rm Voigt}=2$ shown by black arrow in right-hand panel and red solid line in left-hand panel.

panel) resulting in a larger AICC value[1]. Whereas, for $N_{\rm Voigt}>2$, the χ^2 ($\chi^2_{\rm dof}\sim 1.0$) is small (see blue curve in left-hand panel) but with increasing $N_{\rm Voigt}$ the complexity of the model increases and hence AICC also increases. It is interesting to note that the χ^2 remains nearly constant for $N_{\rm Voigt}\geq 2$ whereas AICC systematically increases for $N_{\rm Voigt}>2$. A model simply based on χ^2 minimization, thus would be degenerate for $N_{\rm Voigt}\geq 2$. The minimum AICC occurs for $N_{\rm Voigt}=2$ (black arrow showing red star in right-

---

[1] The reduced χ^2 is given by,

$$\chi^2_{\rm dof}=\frac{\chi^2}{n-p}$$

where $n=134$ is number of pixels in the given region (left-hand panel of Fig. 3.2), $p=3\times N_{\rm Voigt}$ is number of free parameters where factor 3 accounts for the number of free parameters in each Voigt component ($\lambda_{\rm c}$, $\rm N_{HI}$ and b).

hand panel) which shows trade-off between goodness-of-fit ($\chi^2_{\rm dof} \sim 0.9$) and complexity of the model. The corresponding best fit model is shown by red curve in left-hand panel. Thus for minimum AICC, $\chi^2_{\rm dof}$ is also close to 1 and hence we chose it to be the best fit model. In the third panel from top of Fig. 3.1, we show the results of fitting each region with as many Voigt components as necessary for the minimum AICC and $\chi^2_{\rm dof} \sim 1$.

3. **Significance level of fitted lines :** Initially we fitted the lines that are identified using a simple approximation of "Crude Significance Level". However, we used a "Rigorous Significance Level" (hereafter RSL, see Appendix 3.6.1 for details) formula (Keeney et al. 2012) to include the lines in the final line catalog as given below,

$$\mathrm{RSL} = (\mathrm{SNR})_1 \frac{W_\lambda}{\Delta\lambda} \, \frac{\eta(x)}{x} \, f_c(x, \lambda, b) \quad . \tag{3.4}$$

where, x is width (in pixels) of discrete region over which equivalent width W_λ is calculated, $\Delta\lambda = \lambda/x$, λ is width (in Å) of the discrete region, $f_c(x, \lambda, b)$ is the fractional area of the HST-COS LSF contained within the region of integration, $\eta(x) = (\mathrm{SNR})_x/(\mathrm{SNR})_1$ takes care of the fact that noise property may not be purely Poissonian, $(\mathrm{SNR})_1$ is signal to noise ratio per pixel, $(\mathrm{SNR})_x$ signal to noise ratio average over discrete region containing x pixels. We used the parametric form of $\eta(x), f_c(x, \lambda, b)$ given by Keeney et al. (2012, their Eq. 4, Eq. 7 to Eq. 11 with parameters given in Table.1 for the coadded data).

To avoid the spurious detection, we retain only feature measured with RSL > 4 in the final line catalog. Other features are excluded from further analysis. Using this criteria, we find that the number of identified lines with $\log \mathrm{N_{HI}} \geq 12.4$ to be fitted by VIPER (1277 H I Lyα lines) are similar to those of D16 (1280 H I Lyα lines)[1]. In the third panel from top of Fig. 3.1, we show the RSL for each fitted component above the line. Bottom panel of Fig. 3.1 shows that the final accepted Voigt profile fit that contains only those components which have RSL > 4.

---

[1] The total number of identified lines above completeness limit (i.e., $\log \mathrm{N_{HI}} \geq 13.6$) in VIPER and D16 line catalog is 533 and 522 respectively.

3.4 Consistency of VIPER

The comparison of Voigt profile fit of VIPER with that of D16 method is illustrated in Fig. 3.3. The fit from VIPER and D16 method is shown by solid red and dashed blue line respectively. The line centers of components fitted by VIPER and D16 method are shown by solid red and dashed blue vertical ticks respectively. For VIPER the reduced χ^2 is small as compared to that for the components obtained by D16. The top row shows the example where VIPER fit (in terms of number of component and the values of fitted parameter along with the errorbar) matches well with D16 fit. In most situations the fitted parameters from VIPER matched well with those from D16 within errors. In few cases VIPER fitted data better than D16 method (bottom row of Fig. 3.3) in terms of reduced χ^2. We fitted all the observed spectra using VIPER and form a line catalog "VIPER catalog" (see Appendix for details). It should be noted that unlike D16, VIPER does not fit higher order Lyman series lines (e.g. Ly-β, Ly-γ) simultaneously for an accurate measurement of $\log \mathrm{N_{HI}}$ in the case of saturated Lyα lines. However, we show in the next section that the differences in CDDF and line width distributions from "VIPER catalog" and "D16 catalog" are very small.

3.4.1 Column density distribution function (CDDF):

Column density distribution function, $f(\mathrm{N_{HI}}, z)$, describes the number of absorption lines in the column density range $\log \mathrm{N_{HI}}$ and $\log \mathrm{N_{HI}} + \mathrm{d}\log \mathrm{N_{HI}}$ and in the redshift range z to $z+dz$. For a singular isothermal density profile of Lyα absorbers, the H I photoionization rate Γ_{HI} can be inferred from H I CDDF as (Schaye 2001; Shull et al. 2012),

$$f(\mathrm{N_{HI}}, z) = \frac{\partial^2 N}{\partial z \, \partial \log(\mathrm{N_{HI}})} \propto [\alpha(T)\, \Gamma_{\mathrm{HI}}]^{-1/2} \quad . \tag{3.5}$$

where $\alpha(T) \propto T^{-0.726}$, is Case A recombination rate coefficient which depends on temperature. We take into account the completeness of the sample

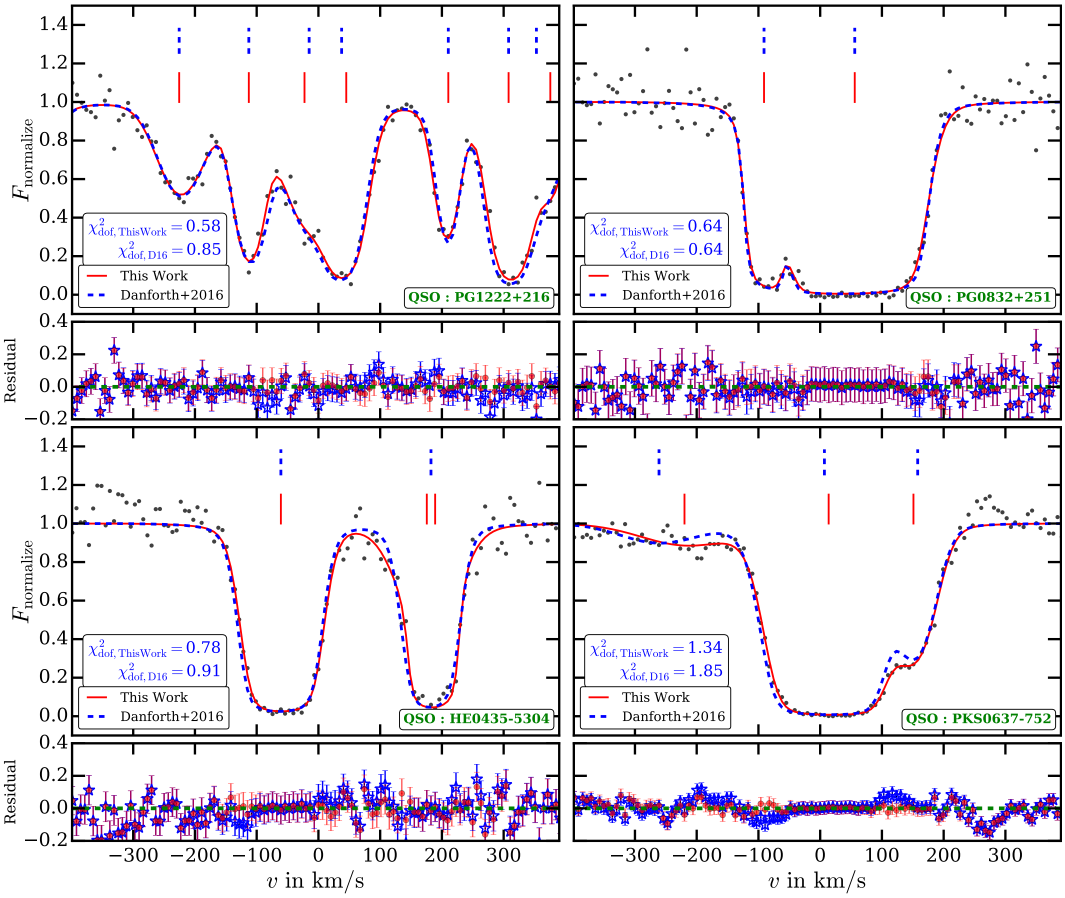

Figure 3.3: Comparison of Voigt profiles fitted using our procedure with those of D16 for four different regions in our sample. Black filled circles, the solid red line and the dashed blue line are observed data points, the best fit profile from VIPER and from D16 respectively. The spectra are shown in the velocity scale defined with respect to the redshift of the strongest component. Blue dashed and red continuous vertical ticks show the location of identified components by D16 and VIPER respectively. The residual between observed data and fitting from D16 (open blue stars) and VIPER (red filled circles) model are shown in the corresponding lower panel. In majority of cases ($\sim$ 89 percent, like upper row panels) our parameters within 1σ errors match with those from D16. However, for some cases our fit to the data using AICC (i.e., using criteria ΔAICC$\geq$ 5, see text for details) is found to be better (lower row panels). In all four cases shown above our χ^2_{dof} is better than the corresponding from D16.

while calculating the redshift path length as a function of $\log \mathrm{N_{HI}}$. Following D16, we calculate the CDDF in 13 $\log \mathrm{N_{HI}}$ bins with centers at $12.5, 12.7, \cdots, 14.7, 14.9$ and width $\mathrm{d}\log \mathrm{N_{HI}} = 0.2$.

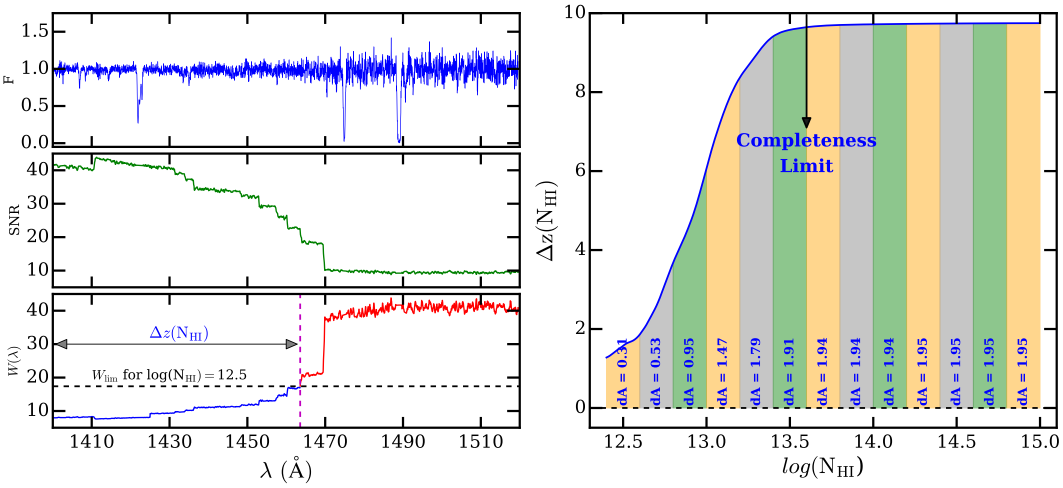

Figure 3.4: *Left-hand* panel illustrates our redshift path length calculation for sightline towards QSO H1821+643. Top, middle and lower left-hand panels show the flux, SNR per pixel and equivalent width vector respectively. Equivalent width vector $W(\lambda)$ is calculated for RSL = 4 in Eq.3.4. The limiting equivalent width ($W_{\rm lim}$), estimated from the curve of growth, corresponding to $\log(\mathrm{N_{HI}}) = 12.5$ is shown by horizontal black dashed line in the bottom panel. The redshift path length, $\Delta z(\mathrm{N_{HI}} = 10^{12.5}\ \mathrm{cm^{-2}})$, for this sightline is the redshift covered by region $W(\lambda) \leq W_{\rm lim}$. The total redshift path length is sum of the Δz measured along all QSO sightlines. *Right-hand* panel shows (blue curve) the total redshift path length as a function of $\log \mathrm{N_{HI}}$ (known as sensitivity curve). The completeness limit for the sample is $\log \mathrm{N_{HI}} = 13.6$ (shown by blue arrow). The fractional area in a given $\log \mathrm{N_{HI}}$ bin, $dA = dz\, d\log(\mathrm{N_{HI}})$, is area under the blue curve in the corresponding $\log \mathrm{N_{HI}}$ bin (shown by blue text) that is used in CDDF calculation.

Left-hand panel in Fig. 3.4 illustrates the procedure we adopt for calculating the redshift path length. The top, middle and lower sub-panels show the flux F, SNR per pixel and equivalent width vector $W(\lambda)$ respectively for a sightline towards QSO H1821+643. To calculate the equivalent width vector $W(\lambda)$, Eq.3.4 is rearranged and solved for $W(\lambda)$ by taking RSL=4 and $b = 17\ \mathrm{km\ s^{-1}}$ (corresponds velocity resolution) for each pixel. Next we calculate limiting equivalent width $W_{\rm lim}$ from curve of growth (b = 17 km s^{-1}) for different values of $\log {\rm N_{HI}}$. As an example we show $W_{\rm lim}$ for $\log {\rm N_{HI}} = 12.5$ by black dashed horizontal line in bottom panel of Fig. 3.4. The redshift path length $\Delta z({\rm N_{HI}} = 10^{12.5}$ cm^{-2}) for this sightline (shown by blue curve in bottom panel) is the sum of redshift range for which $W(\lambda) \leq W_{\rm lim}$. The total redshift path length $\Delta z({\rm N_{HI}})$ covered in the observed sample is a sum of all the redshift path length in individual sightlines. The total redshift path length $\Delta z({\rm N_{HI}})$ is then plotted as a function of $\log {\rm N_{HI}}$ ('Sensitivity curve') as shown in right-hand panel of Fig. 3.4. The completeness limit for the sample is $\log {\rm N_{HI}} = 13.6$ (shown by blue arrow) i.e., the lines with $\log {\rm N_{HI}} \geq 13.6$ are always detectable over the entire observed wavelength range for the full sample. The fractional area $dA = dz\times$ dlog ${\rm N_{HI}}$ in Eq.3.5 is calculated by integrating the sensitivity curve in the corresponding dlog ${\rm N_{HI}}$ bin. We shall refer to the CDDF obtained for our fitted parameters (over the redshift range $0.075 \leq z \leq 0.45$) in this way as "VIPER CDDF".

In the left-hand panel of Fig. 3.5, we compared the VIPER CDDF with CDDF given in Table.5 of D16 (also S15) and CDDF calculated from D16 line catalog. The CDDF given in Table.5 of D16 is calculated from 2256 H I absorbers in the redshift range $0 \leq z < 0.75$ whereas CDDF calculated from D16 line catalog contains 1280 H I absorbers in the redshift range $0.075 \leq z < 0.45$. We used our redshift path length estimation for CDDF calculation from D16 line catalog. The VIPER and the D16 line catalog CDDFs are consistent with each other within 1σ except at high and low $\log {\rm N_{HI}}$ bins. The median $\log {\rm N_{HI}}$ from VIPER and D16 is 13.39 ± 0.61 and 13.38 ± 0.63 respectively. The two sample KS test p-value between the $\log {\rm N_{HI}}$ distribution of VIPER and D16 line catalog is 0.83. The consistency between VIPER CDDF and D16 line catalog CDDF suggests that the number of components identified by VIPER in different dlog ${\rm N_{HI}}$ bins are similar to those from D16 line catalog. Whereas, the agreement between VIPER CDDF and CDDF from D16 paper indicates that our redshift path calculation is consistent with that from D16 paper.

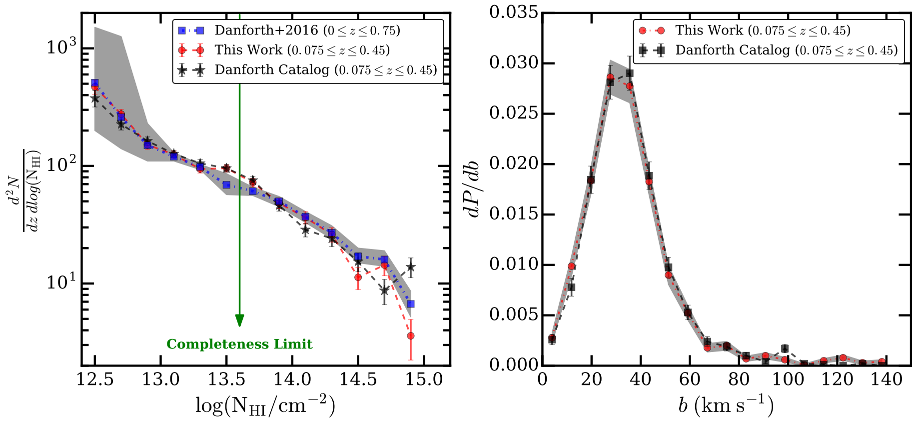

Figure 3.5: Left-hand panel shows comparison of CDDF from VIPER (red circle, $0.075 \leq z \leq 0.45$, 1277 H I Lyα lines), D16 line catalog (black stars, for $0.075 \leq z \leq 0.45$, 1280 H I Lyα lines) and Table.5 in D16 (blue square curve with gray shaded region, for $0 \leq z \leq 0.75$, 2256 H I Lyα lines). The two sample KS test p-value between VIPER and D16 line catalog for $\log N_{HI}$ distribution is 0.83. Thus within errors the CDDF from the two methods are consistent with each other. At high column densities the differences arises due to differences in the fitting procedure (multi-component fitting using AICC). Right-hand panel shows the b parameter distribution from VIPER (red curve for $0.075 \leq z \leq 0.45$) and D16 line catalog (black curve with shaded region for $0.075 \leq z \leq 0.45$). The two sample KS test p-value between VIPER and D16 line catalog for b parameter distribution is 0.41. Thus the b parameter distribution from VIPER is in good agreement with that of D16 line catalog validating our procedure. In both panels the error bars shown are computed assuming the Poission distribution.

We notice occasional component differences when the Lyα line is heavily saturated between D16 and our VIPER fits. Unlike D16, VIPER does not include simultaneous fitting of Lyα line and higher order Lyman series lines (such as Ly-β, Ly-γ). This could be the reason for minor mismatch of CDDF in the bins $\log N_{HI} = 13.5 - 13.7$ and $\log N_{HI} = 14.7 - 14.9$. Thus minor differences one notices in high and intermediate $\log N_{HI}$ bins in observed CDDF can be attributed to the differences in the multi-component fitting procedure in particular to the way the total number of components fitted to a given identified absorption region. However, there is an overall good agreement between CDDF derived by D16 and VIPER.

3.4.2 Linewidth distribution

In the right-hand panel of Fig. 3.5, we compared the doppler-b parameter distribution from VIPER catalog (red curve with shaded region) and D16 line catalog (black dashed line with error bars)[1]. In both distributions the errors are assumed to be poisson distributed. The median value of b parameter from VIPER and D16 is 32.9 ± 20.8 $\mathrm{km\,s^{-1}}$and 33.9 ± 18.3 $\mathrm{km\,s^{-1}}$respectively. The two sample KS test p-value between the b parameter distribution of VIPER and D16 line catalog is 0.41. Thus the two distributions i.e., CDDF and linewidth distribution are consistent with each other validating the consistency between VIPER and D16 line fitting methods.

3.5 Summary

In this chapter, we have developed a code "VoIgt profile Parameter Estimation Routine (VIPER)" to automatically fit the Lyα absorption lines with Voigt profile. This code is parallel and is written in *python*. In VIPER, the blended and saturated features are fitted simultaneously with multi-component Voigt profiles. An objective criteria based on information theory is used to find the number of Voigt profiles needed to describe the Lyα forest. The main results of this work are as follows

- Using VIPER, we fitted a sample of 82 QSO spectra obtained from Cosmic Origins Spectrograph on board Hubble Space Telescope and compiled a Lyα line catalog called "VIPER line catalog". The fitted parameters such as column density ($\mathrm{N_{HI}}$), line width parameter (b) and line width distribution from VIPER line catalog are found to be consistent with those from Danforth et al. (2016).

---

[1] In left-hand panel of Fig. 3.5, we compare our CDDF with that from D16 line catalog ($0.075 \leq z \leq 0.45$) and D16 paper ($0 \leq z \leq 0.75$). However, in right-hand panel of Fig. 3.5, we compare our b parameter distribution (i.e. in the redshift range $0.075 \leq z \leq 0.45$) with that from D16 catalog only as the b parameter distribution is not available in D16 paper.

- The median b parameter from VIPER (32.9 ± 20.8 km s^{-1}) is consistent with that from Danforth et al. (2016, 33.9 ± 18.3 km s^{-1}). Whereas, the median $\log \mathrm{N_{HI}}$ from VIPER (13.39 ± 0.61) is in good agreement with that from Danforth et al. (2016, 13.38 ± 0.63)

- We calculate the appropriate redshift path length $\Delta z(\mathrm{N_{HI}})$ and the sensitivity curve from HST-COS data. We calculate the CDDF after accounting for the incompleteness of the sample. Our calculated CDDF in the redshift range ($0.075 \leq z \leq 0.45$) is consistent (KS test p-value is 0.83) with that of Danforth et al. (2016) CDDF in the redshift range ($0 \leq z \leq 0.75$).

- In chapter 5, we used VIPER in simulations and observations to calculate CDDF and thereby to constrain Γ_{HI}.

3.6 Appendices

3.6.1 Rigorous significance level (RSL)

In this section we derive an expression for RSL used in the section 3.3. The equivalent width is defined as,

$$W = \int_{\lambda_{\mathrm{min}}}^{\lambda_{\mathrm{max}}} [1 - F(\lambda)]\, d\lambda \tag{3.6}$$

The above equation in discrete form can be rewritten as,

$$W = \sum_{i=1}^{x} [1 - F_i]\, \Delta\lambda \tag{3.7}$$

where x is number of pixels in the wavelength range $\lambda_{\mathrm{max}} - \lambda_{\mathrm{min}}$ and $\Delta\lambda$ is dispersion of the spectrum.

For very weak lines, $1 - F_i$ would be in a region dominated by noise. For these lines the equivalent width in a wavelength region λ_x is given by,

$$W = \lambda_x\, \sigma_x = \Delta\lambda\, x\, \sigma_x \tag{3.8}$$

where $\lambda_x = \lambda_{\max} - \lambda_{\min} = \Delta\lambda \times x$ and σ_x is average flux error in the wavelength range λ_x. In order to accept a feature as a absorption line we require it to have an equivalent width significantly larger than the above value of $\lambda_x \sigma_x$. Denoting this required significance level by N_σ, it is straightforward to see that the limiting equivalent width will be given by $W_{\rm lim} = N_\sigma \, \Delta\lambda \, x \, \sigma_x$. There is, however, an additional complication arising from the LSF which will affect the observed equivalent width.

In Fig.3.6, we show that the true equivalent width is obtained by dividing the observed equivalent width by the fractional area $f_c(x, \lambda, b)$ of the HST COS LSF contained within the region of integration.

$$W_{\rm lim} = \frac{N_\sigma}{f_c(x,\lambda,b)} \; \Delta\lambda \; x \; \sigma_x = \frac{N_\sigma}{f_c(x,\lambda,b)} \; \frac{\Delta\lambda \; x}{({\rm SNR})_x} \tag{3.9}$$

where $({\rm SNR})_x$ is signal to noise ratio at a resolution $\lambda_x = \Delta\lambda \times x$ i.e., it is averaged over wavelength range λ_x. The noise in HST COS is not purely Poissonian. Hence Keeney et al. (2012) define the measured relationship between the smoothed signal-to-noise $({\rm SNR})_x$ and the signal-to-noise per pixel $({\rm SNR})_1$ to be $\eta(x)$ as,

$$\eta(x) \equiv \frac{({\rm SNR})_x}{({\rm SNR})_1} \tag{3.10}$$

where $\eta(x) = x^{1/2}$ for purely Poissonian noise. The equivalent width is then given by,

$$W_{\rm lim} = \frac{N_\sigma}{f_c(x,\lambda,b)} \; \frac{\Delta\lambda \; x}{({\rm SNR})_1 \; \eta(x)} \tag{3.11}$$

Rearranging the terms in above expression to solve for the N_σ which we referred to as rigorous significance level RSL in paper,

$${\rm RSL} = N_\sigma = ({\rm SNR})_1 \, \frac{W_{\rm lim}}{\Delta\lambda} \, \frac{\eta(x)}{x} \, f_c(x,\lambda,b) \tag{3.12}$$

Note that RSL is significance level hence it is dimensionless.

3.6.2 VIPER line catalog

We formed a line catalog by fitting the observed spectra using VIPER. Table 3.1 shows few fitted parameters from the VIPER line catalog for spectra towards

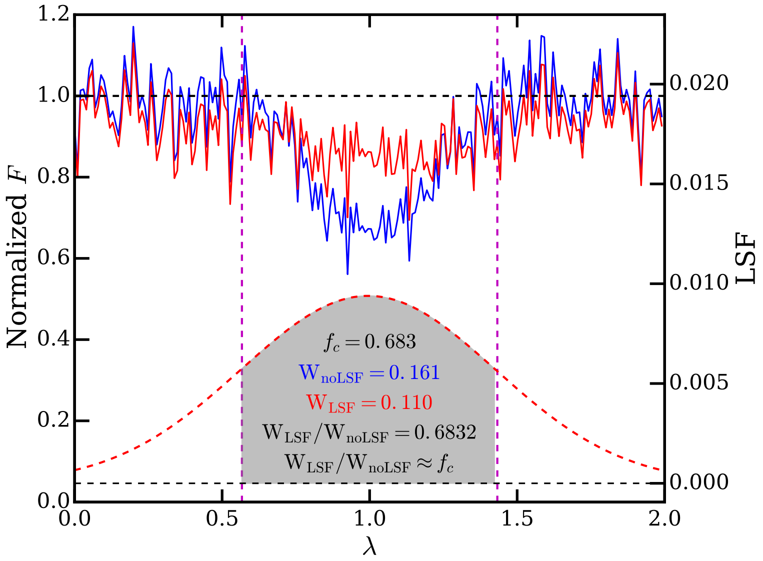

Figure 3.6: shows a toy model explaining the effect of fractional LSF area on equivalent width (W). Blue solid curve shows normalized flux when flux is not convolveed with LSF while the red solid curve shows normalized flux when flux is convolve with LSF. The red dashed curve shows LSF which is assumed to be Gaussian in this toy model. The magenta dashed vertical lines corresponds to $\lambda_{\min}$ and $\lambda_{\max}$ range over which equivalent width is calculated by integration. The fraction of the LSF area contributing to the line $f_c = 0.683$ is shown by gray shaded region. The true equivalent width (i.e., in the absence of LSF, blue solid curve) within $\lambda_{\min}$ and $\lambda_{\max}$ range is $W_{\mathrm{noLSF}} = 0.161$. The observed equivalent width due to LSF (red solid curve) in the same wavelength range is $W_{\mathrm{LSF}} = 0.110$. Thus in order to obtain the true equivalent width we need to divide the observed equivalent width (which is affected by LSF) by fractional area ($W_{\mathrm{noLSF}} = W_{\mathrm{LSF}}/f_c$). Convolving the instrumental LSF with the intrinsic line shape introduces a generic dependence on the Doppler b parameter. The HST COS LSF is wavelength dependent, introducing a wavelength dependence to f_c. Thus $f_c \equiv f_c(x, \lambda, b)$

Table 3.1: Few fitted parameters from the VIPER line catalog are given for spectra towards QSO 1ES1553+113. The full line catalog is available online in ASCII format.

λ (Å)	dλ (Å)	$\log \mathrm{N_{HI}}$ ($\mathrm{cm^{-2}}$)	dlog $\mathrm{N_{HI}}$ ($\mathrm{cm^{-2}}$)	b ($\mathrm{km\,s^{-1}}$)	db ($\mathrm{km\,s^{-1}}$)
1330.827	0.003	13.68	0.01	32.06	1.08
1339.936	0.002	14.23	0.01	38.09	0.80
1361.466	0.019	12.84	0.06	34.10	7.00
1361.922	0.013	12.94	0.05	28.91	4.99
1365.408	0.018	13.04	0.04	48.75	5.57

QSO 1ES1553+113. The first, second, third, fourth, fifth and sixth columns show fitted wavelength (λ in Å), error in wavelength ($d\lambda$ in Å), log of column density ($\log \mathrm{N_{HI}}$ in $\mathrm{cm^{-2}}$), error in log of column density (dlog $\mathrm{N_{HI}}$ in $\mathrm{cm^{-2}}$), b parameter (in $\mathrm{km\,s^{-1}}$) and error in b parameter (db in $\mathrm{km\,s^{-1}}$) respectively. The full VIPER line catalog is available online in ASCII format with this paper.

4 | Consistency of CITE and GLASS

The contents of this chapter are taken from Gaikwad et al. (2017a,b).

4.1 Introduction

The set of three modules CITE, GLASS and VIPER developed in Chapter 2 and 3, provides an efficient way to simulate the IGM properties. We validated VIPER in Chapter 3 by comparing observed CDDF and b parameter distribution from VIPER with that from literature. However in Chapter 2, we compared only the line of sight transmitted Lyα flux from our method with observations (in case of low-z Lyα forest) or other simulations (in case of high-z Lyα forest). In this chapter we validate our method, using GADGET-2 post-processed with CITE and GLASS, of modeling the Lyα forest. In particular for low-z ($z < 0.5$) Lyα forest, we compare our simulations with other simulations in the literature using 3 metric in Section 4.2. Whereas for high-z ($2 \leq z \leq 4$) Lyα forest, we show the consistency of our method with GADGET-3 by comparing 8 different statistics in Section 4.3. We summarize the chapter in Section 4.4

4.2 Low-z Lyα forest: Comparison with other simulations

We consider three predictions of our simulation that can be used for comparing different simulations. These are (i) fraction of baryons in different phases of the $T - \Delta$ diagram, (ii) predicted IGM equation of state at $z < 0.3$ and (iii) the relationship between H I column density ($N_{\rm HI}$) and baryon overdensity Δ. Some of these predictions depend on the adopted value of $\Gamma_{\rm HI}$. For the present purpose we used $\Gamma_{\rm HI}$ consistent with QSO dominated (i.e. $f_{\rm esc} = 0$) KS15 UVB radiation model.

4.2.1 Phase diagram of baryons

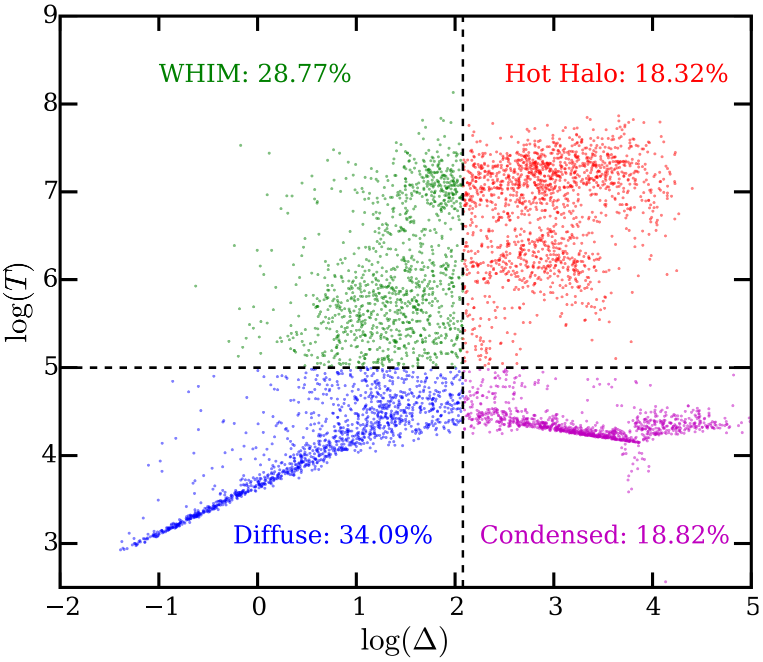

Figure 4.1: Phase diagram ($T - \Delta$ plane) of randomly selected 20000 GADGET-2 particle post-processed with our module CITE in our simulation at $z = 0$. The black dashed line cutoff at $T = 10^5$ and $\Delta = 120$ demarcates diffuse, WHIM, hot halo and condensed gas phase consistent with Davé et al. (2010) (different authors use different definitions, refer to Table 4.1 for details). Diffuse gas phase is mainly responsible for the H I absorption seen in the QSO spectrum in the form of Lyα forest. The percentage of baryons in different phases are given in legend.

The Lyα forest is produced by relatively low density and low temperature diffuse gas. According to FGPA the mean Lyα optical depth is,

$$\tau \propto \Gamma_{\rm HI}^{-1} \, (f_d \, \Omega_b \, h^2)^2 \, \Omega_m^{-0.5}. \tag{4.1}$$

Thus inferred $\Gamma_{\rm HI}$ from Lyα will be degenerate with fraction of baryons in diffuse phase (f_d). We found in the literature that different groups use different ranges in T and cutoff in Δ to demarcate the phase diagram (i.e., Δ vs T diagram) in 4 phases (see Fig. 4.1) namely diffuse, WHIM, hot halo and condensed phase. To make a fair comparison with other results we calculate the gas fraction in diffuse and WHIM phase as per the definitions used by the authors under consideration (see Table 4.1). Smith et al. (2011) label baryons at $z = 0$ as diffuse gas if $T < 10^5$ K and $\Delta < 1000$ and as WHIM if 10^7 K $> T > 10^5$ K and $\Delta < 1000$. We apply the same cutoff at $z = 0$ and find the diffuse and WHIM fraction to be ~ 39.11 per cent and 40.52 per cent respectively which is consistent with the ~ 40 per cent, $40 - 50$ per cent to that of Smith et al. (2011) and Shull et al. (2015). Note that the moderate feedback processes are included in the AMR (Adoptive Mesh Refinement) simulations of Smith et al. (2011) whereas ours is a SPH simulation without any feedback.

Similarly Davé et al. (2010) have incorporated momentum driven galactic outflows and various other wind models in their SPH simulations (with GADGET) which we lack. They treat baryon particles (at $z = 0$) as part of diffuse if $T < 10^5$ K and $\Delta < 120$ and as a part of WHIM if $T > 10^5$ K and $\Delta < 120$ and found the fraction to be $37 - 43$ per cent, $23 - 33$ per cent respectively. By applying similar cutoff on T and Δ at $z = 0$ our diffuse and WHIM fraction turns out to be 34.09 per cent and 28.77 per cent respectively which is in agreement with Davé et al. (2010).

Note that both the set of simulations discussed above have similar resolution like the one we consider here. Unlike our simulations the simulations from the literature discussed above incorporate feedback at different levels. The close matching of baryon fraction in the diffuse phase between different models reiterate the earlier findings that the contribution of feedback effects are minor in the derived $\Gamma_{\rm HI}$ (Davé et al. 2010; Kollmeier et al. 2014; Shull et al. 2015).

4.2.2 Equation of state

The uncertainties in the epoch of He II reionization are reflected in the values of T_0 and γ at the initial redshift $z_1 = 2.1$. To account for this, we vary T_0 and γ at $z_1 = 2.1$ by allowing them to take extreme values (for T_0 and γ measurement at high-z refer Becker et al. 2011; Boera et al. 2014; Lidz et al. 2010; Schaye

Table 4.1: Comparison of predictions of our low-z simulation with those from the literature

Parameters to[1] compare	This Work ($z = 0$)	(Smith et al. 2011) ($z = 0$)	(Davé et al. 2010) ($z = 0$)	Analytical[2] approximation
T_0[a]	3800 - 5100 K	$\sim$ 5000 K	$\sim$ 5000 K	$\sim$ 2555 K
γ[a]	1.46 - 1.62	$\sim$ 1.60	$\sim$ 1.60	$\sim$ 1.58
Diffuse (in per cent)[b]	34.09	-	37 - 43	-
WHIM (in per cent)[b]	28.77	-	23 - 33	-
Diffuse (in per cent)[c]	39.11	$\sim$40	-	-
WHIM (in per cent)[c]	40.52	40 - 50	-	-
Γ_{12}	0.12 ± 0.03	0.122	$\sim$ 0.2	0.12 ± 0.03
Δ_0[d]	34.8 ± 5.9	36.9	38.9	20.6 ± 4
α[d]	0.770 ± 0.022	0.650	0.741	0.744 ± 0.015

[a] The range in T_0 and γ corresponds to different initial T_0 (10000 to 25000 K) and γ (1.1 to 1.8) at $z_1 = 2.1$ see Table 4.2.

[b] WHIM is defined as $T > 10^5$ K and $\Delta < 120$ whereas the diffuse gas phase is defined as $T < 10^5$ K and $\Delta < 120$ in Davé et al. (2010).

[c] WHIM is defined as 10^7 K $> T > 10^5$ K and $\Delta < 1000$ whereas the diffuse gas phase is defined as $T < 10^5$ K and $\Delta < 1000$ in Smith et al. (2011) (also refer to Danforth and Shull 2008).

[d] The correlation between baryon overdensity Δ and H I column density is expressed as $\Delta = \Delta_0 N_{14}^{\alpha}$, where, Δ_0 is the normalization at a fiducial H I column density $N_{\rm HI} = N_{14} \times 10^{14} \rm cm^{-2}$. This relation is calculated for best fit Γ_{12} in the redshift range $0.2 < z < 0.3$ given in bracket (Fig. 5.16).

[1] We notice that the Paschos et al. (2009) have presented simulations for low-z IGM. While they check the consistency of the mean transmitted flux for their assumed $\Gamma_{\rm HI}$, no attempt was made to measure $\Gamma_{\rm HI}$. Moreover we could not have detailed comparison with their models as the metric we use for comparison are not available for their models.

[2] $T - \Delta$ relation can be obtained by equating Hubble time with net cooling time (Theuns et al. 1998b). Δ vs $N_{\rm HI}$ is calculated following Schaye (2001) assuming Lyα clouds are in hydrostatic equilibrium.

Table 4.2: Details of the thermal history considered in our simulation

	Initial free parameters		Final parameters obtained with CITE							
	$z_1 = 2.1$		$z = 0.1$		$z = 0.2$		$z = 0.3$		$z = 0.4$	
Model Name	T_0	γ	T_0	γ	T_0	γ	T_0	γ	T_0	γ
$T10 - \gamma1.1$	10000	1.10	4136	1.54	4326	1.53	4589	1.51	4844	1.50
$T10 - \gamma1.8$	10000	1.80	4133	1.61	4313	1.61	4568	1.60	4810	1.60
$T20 - \gamma1.1$	20000	1.10	4546	1.48	4971	1.46	5383	1.44	5811	1.42
$T20 - \gamma1.8$	20000	1.80	4493	1.62	4889	1.62	5279	1.61	5677	1.61
$T15 - \gamma1.3$	15000	1.30	4245	1.55	4583	1.54	4902	1.53	5220	1.51

et al. 2000) and obtain the temperatures at redshifts of our interest using CITE. The resulting values of T_0 and γ at lower redshifts as obtained from the $T-\Delta$ distribution are shown in Table 4.2 and Fig. 4.2. One can see that even for a widely different values of the two parameters at $z_1 = 2.1$, the equation of state at $z \sim 0.1 - 0.4$ are quite similar with $T_0 \sim 4000 - 6000$ K and $\gamma \sim 1.5 - 1.6$. This implies that the low-z IGM loses, to a large extent, any memory of the He II reionization. Our results are consistent with previous simulations by Davé et al. (2010) and Smith et al. (2011) who found that the equation of state parameters at $z = 0$ are $T_0 \sim 5000$ K and $\gamma \sim 1.6$.

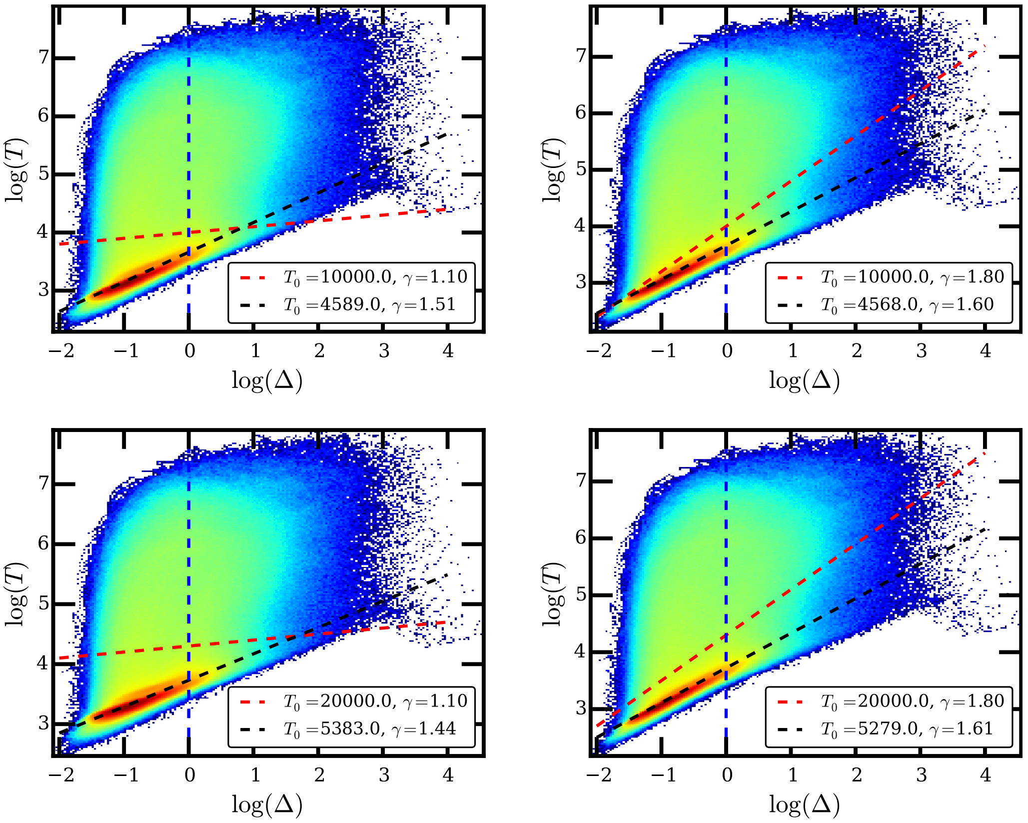

Figure 4.2: Same as Fig. 2.2 except the values of the free parameters T_0 and γ at initial redshift $z_1 = 2.1$ are different as indicated in each subpanel. Despite the large differences in initial T_0 and γ, the final equation of state parameters at $z = 0.3$ are quite similar (See Table 4.2).

The equation of state at low redshifts can be derived by equating net cooling time scale with Hubble timescale. Theuns et al. (1998b) derived such relation (at high z) in low density regime by assuming that the heating rate is dominated by photoheating and cooling rate is dominated by recombination cooling and inverse Compton cooling. The relationship between T and T_0 under this approximation turns out to be,

$$T \sim T_0\, \Delta^{\frac{1}{1+\beta}} \tag{4.2}$$

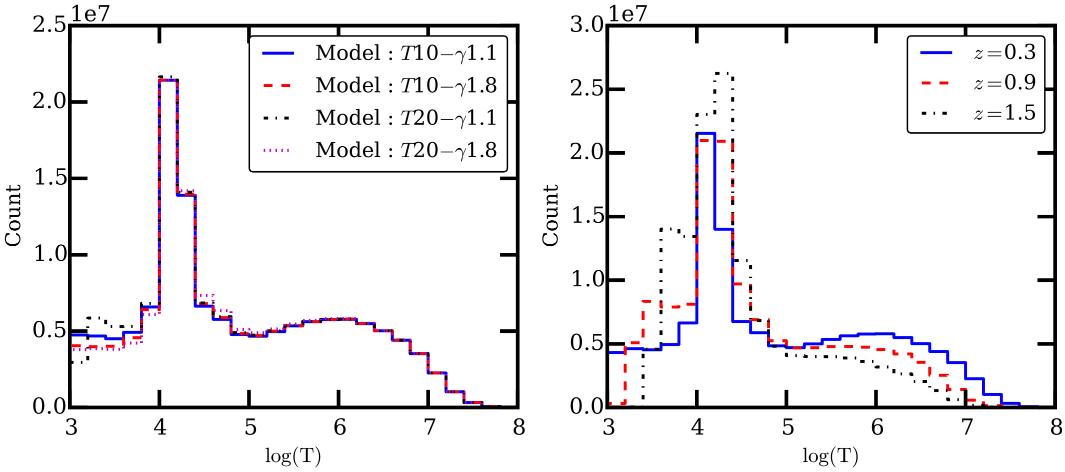

Figure 4.3: The temperature distribution of the gas particles after using CITE starting from varied initial condition at $z_1 = 2.1$ *Left-hand panel:* the final temperature distribution at $z = 0.3$ for different initial T_0 and γ at $z_1 = 2.1$ (see Table 4.2). *Right panel:* the temperature distribution at different redshifts $z = 1.5, 0.9, 0.3$. The initial equation of state ($z_1 = 2.1$) for right-hand panel corresponds to model $T15 - \gamma1.3$ in Table 4.2.

While deriving above equation, we have assumed that the H I recombination rate scales as $T^{-\beta}$. For $\beta = 0.7$ the slope of equation of state is $\gamma = 1.59$. This value is very much close to the one we obtained by evolving the IGM temperature using CITE thus validating our method (see Table 4.2). The mean IGM temperature in the above equation at $z = 0$ is $\sim$ 2555 K (see Eq. C21 in Theuns et al. 1998b). From Table 4.2, one can see our derived temperatures are higher by factor $\sim$ 2 because Eq. 4.2 neglects the heating due to other sources such as shock heating, adiabatic heating due to structure formation etc.

The distribution of the CITE temperatures for the gas particles is shown in Fig. 4.3. The left-hand panel shows the distribution at $z = 0.3$ for different initial values of T_0 and γ. We can see that the distributions at low-z are relatively insensitive to the initial equation of state. Some small differences can be seen at lower temperatures, consistent with the equation of state given in Table 4.2. The right-hand panel of Fig. 4.3 shows the CITE temperature distribution at different redshifts for model $T15-\gamma1.3$. As expected, the fraction of shock heated particles increases with decreasing redshift which is a direct consequence of structure formation shocks.

4.2.3 Δ vs $N_{\rm HI}$ relation

We further compared our simulations with other simulation using relation between baryon overdensity Δ and H I column density $N_{\rm HI}$. Conventionally this relation is fitted by power-law,

$$\Delta = \Delta_0 \, N_{14}^{\alpha} \tag{4.3}$$

where Δ_0 is the normalization at fiducial H I column density $N_{\rm HI} = N_{14} \times 10^{14}\ \rm cm^{-2}$. Assuming hydrostatic equilibrium, Schaye (2001) has derived the above relation analytically for optically thin gas. The normalization and slope of the Δ vs $N_{\rm HI}$ relation is given by,

$$\begin{aligned} \alpha &= \frac{1}{1.5 - 0.26(\gamma - 1)} \\ \Delta_0 &\sim \left[598\,\Gamma_{12}\,\mathrm{T}_{0,4}^{0.26} \left(\frac{1.25}{1+z}\right)^{4.5} \left(\frac{0.0221}{\Omega_{\rm b} h^2}\right)^{1.5} \left(\frac{0.16}{f_g}\right)^{0.5} \right]^{\alpha} \end{aligned} \tag{4.4}$$

where γ and $\mathrm{T}_{0,4}$ is slope of equation of state and mean IGM temperature in units of 10^4 K respectively and f_g is fraction of mass in the gas (excluding stars and molecules). For $\gamma \sim 1.6$, $T_{0,4} \sim 0.45$ and $\Gamma_{12} = 0.12 \pm 0.03$ (for $0.2 < z < 0.3$), the slope and normalization is given as $\alpha \sim 0.744$ and $\Delta_0 = 20.6 \pm 4$. This simple analytic approach is known to produce γ close to what has been seen in the simulations. However, its prediction of the normalization constant need not be accurate as one needs to take care of the baryon fraction in different phases.

To calculate such a relation in simulated spectra we fit the Voigt profile to the absorption lines using our automatic code VIPER (see Chapter 3). To associate the baryon overdensity with absorption line we calculate the optical depth (τ)

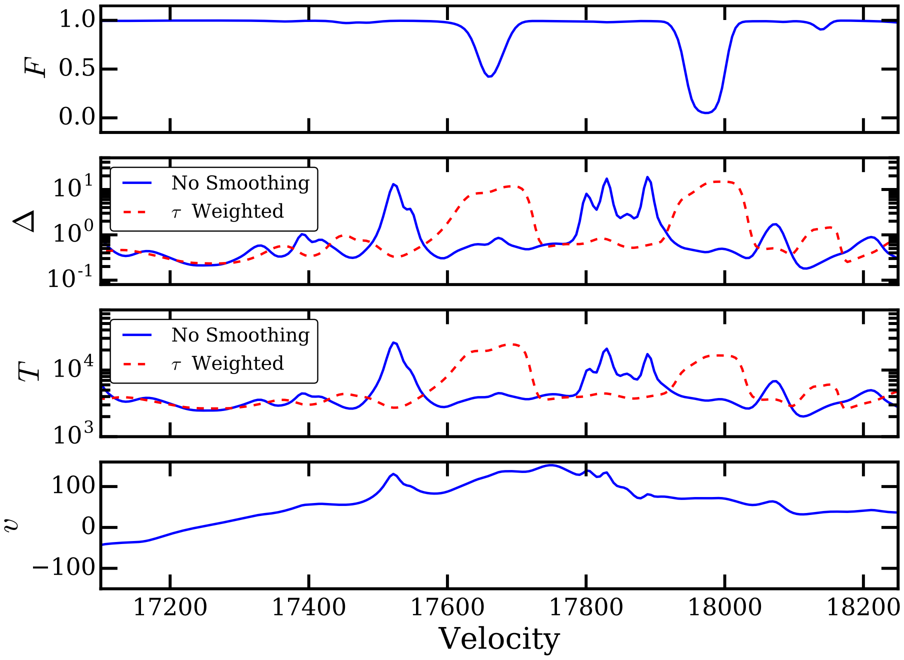

Figure 4.4: Illustration of assigning optical depth (τ) weighted overdensity and temperature to the absorption lines in simulated spectrum (see section 4.2.3). The flux shown in the *top* panel is computed from overdensity (Δ) (blue solid line), temperature (T) (blue solid line) and peculiar velocity (v) (blue solid line) given in 2^{nd},3^{rd} and 4^{th} panel from top respectively. The τ weighted overdensity (see Eq. 4.5) and τ weighted temperature are shown by red dashed lines in 2^{nd} and 3^{rd} panel from top respectively. To calculate τ weighted temperature, we replaced Δ_i in Eq. 4.5 by T_i.

weighted overdensity $\widetilde{\Delta}$ (Schaye et al. 1999) as follows. Let τ_{ij} be the optical depth contribution of overdensity Δ_i at the wavelength corresponding to the pixel i to the optical depth at pixel j. Then the τ weighted overdensity $\widetilde{\Delta}_j$ at pixel j is given by,

$$\widetilde{\Delta}_j = \frac{\sum\limits_{i=1}^{N} \tau_{ij}\Delta_i}{\sum\limits_{i=1}^{N} \tau_{ij}}, \tag{4.5}$$

where N is the total number of pixels in the spectrum. The total optical depth at a pixel j is given by,

$$\tau_j = \sum_{i=1}^{N} \tau_{ij}. \tag{4.6}$$

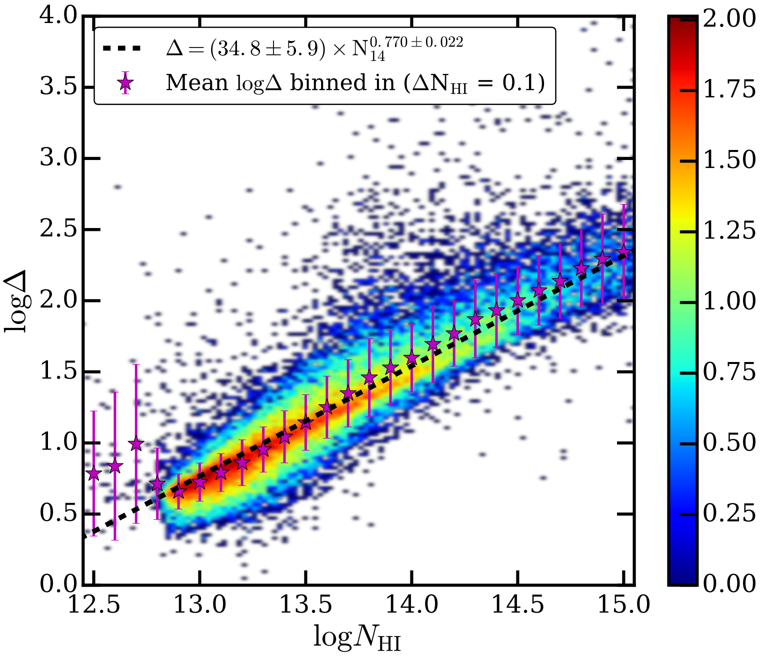

Figure 4.5: Correlation between τ weighted overdensity Δ (see section 4.2.3) and column density $N_{\rm HI}$ for 4000 simulated Lyα forest spectra (SNR=50) in the range $0.2 < z < 0.3$ for $\Gamma_{12} = 0.12$ (consistent with our final measurements see Fig. 5.16). The color scheme represents the density of points in logarithmic units. Magenta star points with errorbar are mean τ weighted overdensity binned in $N_{\rm HI}$ with width $\Delta \rm N_{HI} = 0.1$. Black dashed line shows our best fit to the mean τ weighted overdensity. The errorbar on best fit values corresponds 1σ variation in Γ_{12} which is 0.03 (see Fig. 5.16).

Fig. 4.4 demonstrates our procedure for assigning τ weighted overdensity to an absorption line. $1^{\rm st}$,$2^{\rm nd}$,$3^{\rm rd}$ and $4^{\rm th}$ panel from top shows flux, overdensity (Δ), temperature (T) and peculiar velocity (v) respectively. The flux in the top panel is calculated from Δ, T and v (all solid blue lines). As expected due to power-law equation of state, Δ and T are correlated (solid blue lines). We calculate the τ weighted temperature by replacing Δ_i in Eq. 4.4 by T_i. The τ weighted overdensity and the τ weighted temperature (shown by red dashed line in $2^{\rm nd}$ and $3^{\rm rd}$ panel from top) are also correlated. We then associate this τ weighted overdensity at the absorption line center to the column density of that line (obtained by Voigt profile fitting). Fig. 4.5 shows the density plot of τ weighted overdensity Δ and column density $N_{\rm HI}$ for 4000 simulated Lyα

spectra (SNR = 50) in the range $0.2 < z < 0.3$ for $\Gamma_{12} = 0.12$. The magenta errorbar shows the mean τ weighted overdensity (with 1σ error) in each bin of size $\Delta \log N_{\rm HI} = 0.1$. The black dashed line shows the power-law fit with $\Delta_0 = 34.8 \pm 5.9$ and $\alpha = 0.77 \pm 0.022$. The error in Δ_0 corresponds to 1σ range in Γ_{12} (see Chapter 5). We find that α is less sensitive to Γ_{12}, the error in α accounts for different thermal history (i.e., different values of γ).

The power-law index from our simulation ($\alpha = 0.77 \pm 0.022$) matches well (within 2σ) with Davé et al. (2010) and Tepper-García et al. (2012)($\alpha \sim 0.786$) and theoretically expected values but is slightly higher than Shull et al. (2015) (see Table 4.1). The normalization parameter Δ_0 in our simulation 34.8 ± 5.9 is in agreement (within 1σ) with Shull et al. (2015); Smith et al. (2011) and Davé et al. (2010) simulation but less than Tepper-García et al. (2012) ($\Delta_0 \sim 48.9$). Note that Tepper-García et al. (2012) used a simulation box with factor 2 lower resolution. Also Δ_0 is sensitive to the value of Γ_{12} used to generate simulated spectra (see Eq. 4.4).

We also found that the mean flux decrement (DA) calculated from our simulation is within 30 percent to that found from Paschos et al. (2009) and within 20 percent to that from Kollmeier et al. (2014); Viel et al. (2017) when we use their Γ_{12} in our simulations. The higher values of Paschos et al. (2009) can be attributed to the additional heating incorporated in their simulations. Whereas the differences in DA with Kollmeier et al. (2014); Viel et al. (2017) can be attributed to the differences in fraction of baryons in diffuse phase, mean IGM temperature of the gas in simulation and cosmological parameters used.

We summarize the comparison between our simulation results and others in Table 4.1. As can be seen, our estimates of T_0 and γ, the fraction of baryons in diffuse gas and in WHIM (accounting for differences in the precise definition), as well as Δ vs $N_{\rm HI}$ relation match quite well with those from other simulations, thus validating our method of evolving the gas temperature using CITE. The advantages of using CITE for the low-z IGM studies are as follows:

- Because CITE is based on post-processing the GADGET-2 output, the method is computationally less expensive.

- CITE allows us to explore large thermal history parameter space without performing the full SPH simulation from high-z.
- We are able to include the ionization and thermal evolution in GADGET-2 as post processing steps.
- It is easy to incorporate the radiative cooling for a wide range of metallicities in CITE.

All these allow us to explore a wide range of parameter space thereby enable us to have a reliable estimation of errors associated with the derived $\Gamma_{\rm HI}$ as discussed in Chapter 5.

4.2.4 Feedback Processes

Our simulations does not include any feedback processes such as galactic winds or AGN feedback. Previously Shull et al. (2015) used the simulations of Smith et al. (2011) which included two types of feedback namely *local* and *distributed* feedback. They found that the phase diagram converges as long as simulation box size $L \geq 50h^{-1}$ cMpc. While the fraction of baryons in WHIM and condensed phase changes considerably with feedback prescription, the fraction of baryons in diffuse phase that is responsible for Lyα forest remains similar. Because of this they found that both feedback processes affect column density distribution mildly and hence Γ_{12} constraint remains similar (see Fig. 6 in Shull et al. 2015). The feedback method however, can affect the clumpiness of the IGM on small scales whereas large scale correlation between parameters such as Δ and $N_{\rm HI}$ are nearly unaffected. Similarly Davé et al. (2010) and Kollmeier et al. (2014) found that the properties of Lyα forest are largely insensitive to wind models and feedback prescriptions. Keeping this in mind, now use our simulation to derive $\Gamma_{\rm HI}$ from the HST-COS data in Chapter 5.

4.3 High-z Lyα forest: Comparison with GADGET-3

While CITE works well for low-z Lyα forest (and low resolution simulation ~ 48 ckpc) as shown in previous section, the dynamical evolution of SPH particles at finite pressure is an important effect when we consider high resolution simulations (e.g. ~ 10 ckpc) at high-z ($2 \leq z \leq 4$). In this section, we show the consistency of our method with that from GADGET-3 (Bolton et al. 2006, in which the thermal effects on the hydrodynamical evolution of baryonic particles are taken care of in a self-consistent manner) by comparing different Lyα flux statistics frequently used in the literature.

In particular, we compare the Lyα forest generated from G2-LTF, G2-HTF with that from GADGET-3 using eight statistics, namely, (i) the line of sight baryonic density field ($\delta = \Delta - 1$) power spectrum (PS), (ii) the flux probability distribution function (PDF), (iii) the flux power spectrum (PS), (iv) the wavelet statistics, (v) the curvature statistics, (vi) the column density distribution function (CDDF), (vii) the line width (b) distribution function and (viii) the b vs $\log \mathrm{N_{HI}}$ correlation. The statistics (i)-(v) are obtained assuming Lyα transmitted flux to be a continuous field whereas, the statistics (vi)-(viii) are based on Voigt profile decomposition of Lyα forest. For this purpose we use our automatic Voigt profile fitting code VIPER described in Chapter 3 (also see Gaikwad et al. 2017c). Note that the statistics (ii)-(viii) are relatively straightforward to obtain from observations as well as from simulations.

4.3.1 Generation of mock Lyα forest spectra

To perform a quantitative comparison of the Lyα forest spectra extracted from different models, we shoot random sightlines through the simulation. We then splice together the lines of sight in such a way that it covers a redshift path $z \pm 0.05$, where $z = 2.5, 3.0, 3.5, 4.0$ are redshifts of the simulation box. Following Gaikwad et al. (2017b,c); Rollinde et al. (2013), we generate a mock sample of $N_{\mathrm{spec}} = 20$ Lyα forest spectra for the GADGET-3, G2-LTF and G2-HTF models. We repeat the procedure by choosing different random sightlines and generate

$N = 100$ such mock samples. The collection of N mock samples constitute a "mock suite" that consists of $N \times N_{\rm spec} = 2000$ simulated spectra. We estimate the covariance matrix for different statistics using the simulated spectra.

4.3.2 Line of sight δ power spectrum

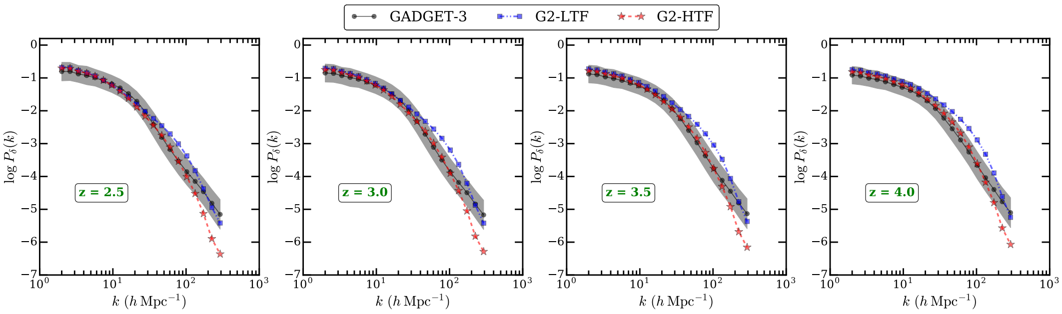

Figure 4.6: Each panel shows the comparison of the line of sight density field power spectrum from GADGET-3 (black circle), G2-LTF (blue squares) and G2-HTF (red stars) models. The gray shaded region in each panel represents 1σ uncertainty (diagonal elements of covariance matrix given in Eq. 4.9) on the PS from GADGET-3. We present the results for 4 redshifts that are identified in each panel.

The statistical properties of the Lyα forest are sensitive to the line of sight 1D density field. We calculate the power spectrum of the 1D density fluctuations using sightlines of comoving length equal to the simulation box size $10\,h^{-1}$ cMpc. This is done by first computing the Fourier transform $\delta(k)$ of the density field $\delta(x)$, the corresponding power is simply given by $P_\delta(k) \propto |\delta(k)|^2$. We normalize the density PS (Zhan et al. 2005) as,

$$\sigma^2_{F_\delta} = \int\limits_{-\infty}^{\infty} \frac{dk}{2\pi}\, P_\delta(k) \tag{4.7}$$

where $\sigma^2_{F_\delta}$ is variance of the 1D density field. We bin the density PS in 20 equispaced logarithmic bins with centers at $k = 2, 2.6, 3.38, \cdots, 224.92, 292.38$ and bin width $\Delta k = 1.3$ (Kim et al. 2004).

Following Gaikwad et al. (2017b); Rollinde et al. (2013), we take the average of all density power spectra along different sightlines in a mock sample (consisting

Table 4.3: Reduced χ^2 between G2-LTF, G2-HTF model and reference model GADGET-3 for different statistics.

	$z = 2.5$		$z = 3.0$		$z = 3.5$		$z = 4.0$	
Statistics[1]	G2-LTF	G2-HTF	G2-LTF	G2-HTF	G2-LTF	G2-HTF	G2-LTF	G2-HTF
Density Power spectrum (δ)	0.41	0.98	0.67	0.79	0.82	0.60	1.21	0.58
Flux PDF	1.26	0.76	1.41	0.80	1.79	0.76	1.87	0.81
Flux PS	0.50	0.36	0.68	0.25	1.82	0.22	3.23	0.45
Wavelet PDF	0.20	0.42	0.60	0.17	4.46	0.27	12.84	0.74
Curvature PDF	0.30	0.19	1.34	0.49	6.81	0.82	17.07	0.78
CDDF[2]	0.28	0.27	0.25	0.23	0.40	0.21	1.61	0.14
b parameter distribution	0.36	0.37	0.34	0.23	0.37	0.24	0.26	0.16
b vs $\log \mathrm{N_{HI}}$ correlation	0.17	0.25	0.22	0.22	0.23	0.17	0.50	0.15

[1] For a given redshift, all the astrophysical parameters (T_0, γ, $\Gamma_{\rm HI}$) are same for G2-LTF, G2-HTF and GADGET-3 models. Reduced χ^2 is calculated using full covariance matrix for flux PDF and flux PS. However, for other statistics we used diagonal elements of the covariance matrix as off diagonal elements are noisy.

[2] The reduced χ^2 for CDDF, b parameter distribution and b vs $\log \mathrm{N_{HI}}$ correlation is calculated from lines in a sample above completeness limit at each redshift.

of 20 lines of sight). We then calculate the mean density PS and the associated errors from the mock suite (which consists of $N = 100$ mock sample). Let $P_{\delta,n}(k_i)$ denotes the value of density PS in i^{th} bin of n^{th} mock sample, then the average density PS in i^{th} bin is given by,

$$\overline{P}_{\delta}(k_i) = \frac{1}{N} \sum_{n=1}^{N} P_{\delta,n}(k_i) \quad . \tag{4.8}$$

The covariance matrix element $C(i,j)$ between the i^{th} and j^{th} bins is given by,

$$C(i,j) = \frac{1}{N-1} \sum_{n=1}^{N} [\overline{P}_{\delta}(k_i) - P_{\delta,n}(k_i)][\overline{P}_{\delta}(k_j) - P_{\delta,n}(k_j)] \tag{4.9}$$

where, i and j can take values from 1 to the number of bins.

Fig. 4.6 shows the density PS for the GADGET-3 (black circles), G2-LTF (blue squares) and G2-HTF (red stars) models at four different redshifts as mentioned in each panel. The grey shaded region corresponds to the 1σ error on the GADGET-3 density PS. We find that at all redshifts the density PS for G2-HTF is within 1σ ($\chi^2_{\rm dof} \leq 1$, see Table 4.3) of that from GADGET-3. However on scales below 36 h^{-1} ckpc ($k \sim 175\, h$ Mpc^{-1}), the power predicted from G2-HTF is smaller than that from GADGET-3. This is due to the fact that the minimum temperature (irrespective of the density) in G2-HTF is $\sim$ 10000 K, while the particles with

$\Delta < 1$ are at temperature smaller than 10000 K for GADGET-3 model because of the $T - \Delta$ relation. This higher temperature in G2-HTF model leads to an additional pressure smoothing, thus the power on scales below 36 h^{-1} ckpc is smaller than that from GADGET-3. On the other hand, model G2-LTF has higher power on the scales in the range 35-135 ckpc ($k \sim 180 - 47\,h$ Mpc^{-1}). This power is more prominent for highest redshift bin $z = 4$. This highlights the need for an appropriate smoothing of the density field on scales larger than pressure smoothing scale for G2-LTF for higher redshifts. Thus unlike the G2-HTF model, a single Jeans length expression for particles in G2-LTF model does not reproduce the GADGET-3 density field PS at all redshifts.

4.3.3 Flux probability distribution function

The flux PDF is one of the flux statistics that is sensitive to $\Gamma_{\rm HI}$ and is relatively straightforward to calculate from observations as well as simulations (Desjacques et al. 2007; Gaikwad et al. 2017b; Jenkins and Ostriker 1991; Kim et al. 2007; McDonald et al. 2000; Rollinde et al. 2013). We calculate the flux PDF in 21 bins with bin centers at $F = 0.05, 0.1, 0.15, \cdots, 0.95, 1.0$ and bin width $\Delta F = 0.05$ (consistent with Kim et al. 2007). The pixels with $F < 0$ ($F > 1$) are included in the first (last) bin. Let $P_n(F_i)$ denote the value of flux PDF in i^{th} bin of n^{th} mock sample then average flux PDF in i^{th} bin (denoted as $\overline{P}(F_i)$) is given by Eq. 4.8 where we replace $P_{\delta,n}(k_i)$ with $P_n(F_i)$. Similarly the covariance matrix element $C(i, j)$ between the i^{th} and j^{th} bins is obtained from Eq. 4.9 by replacing $P_{\delta,n}(k_i)$, $\overline{P}_\delta(k_i)$ with $P_n(F_i)$, $\overline{P}(F_i)$ respectively. We use the full covariance matrix for χ^2 calculations involving the flux PDF.

The left hand panels in Fig. 4.7 show the flux PDF from the GADGET-3 (black circles), G2-LTF (blue squares) and G2-HTF (red stars) models. The grey shaded region corresponds to 1σ error (diagonal element of covariance matrix) on the GADGET-3 flux PDF. Note that $\Gamma_{12} = 1$ is same for all the three models at all redshifts. At all redshifts the flux PDF for G2-HTF and G2-LTF is within 1σ ($\chi^2_{\rm dof} \sim 1$, refer Table 4.3) with that from GADGET-3. Though flux PDF in G2-LTF model is in good agreement with that from GADGET-3 at $z = 2.5$ and 3.0, it differs by ~ 20 percent (in the bin $F = 1.00$) and ~ 50 percent (in the bin

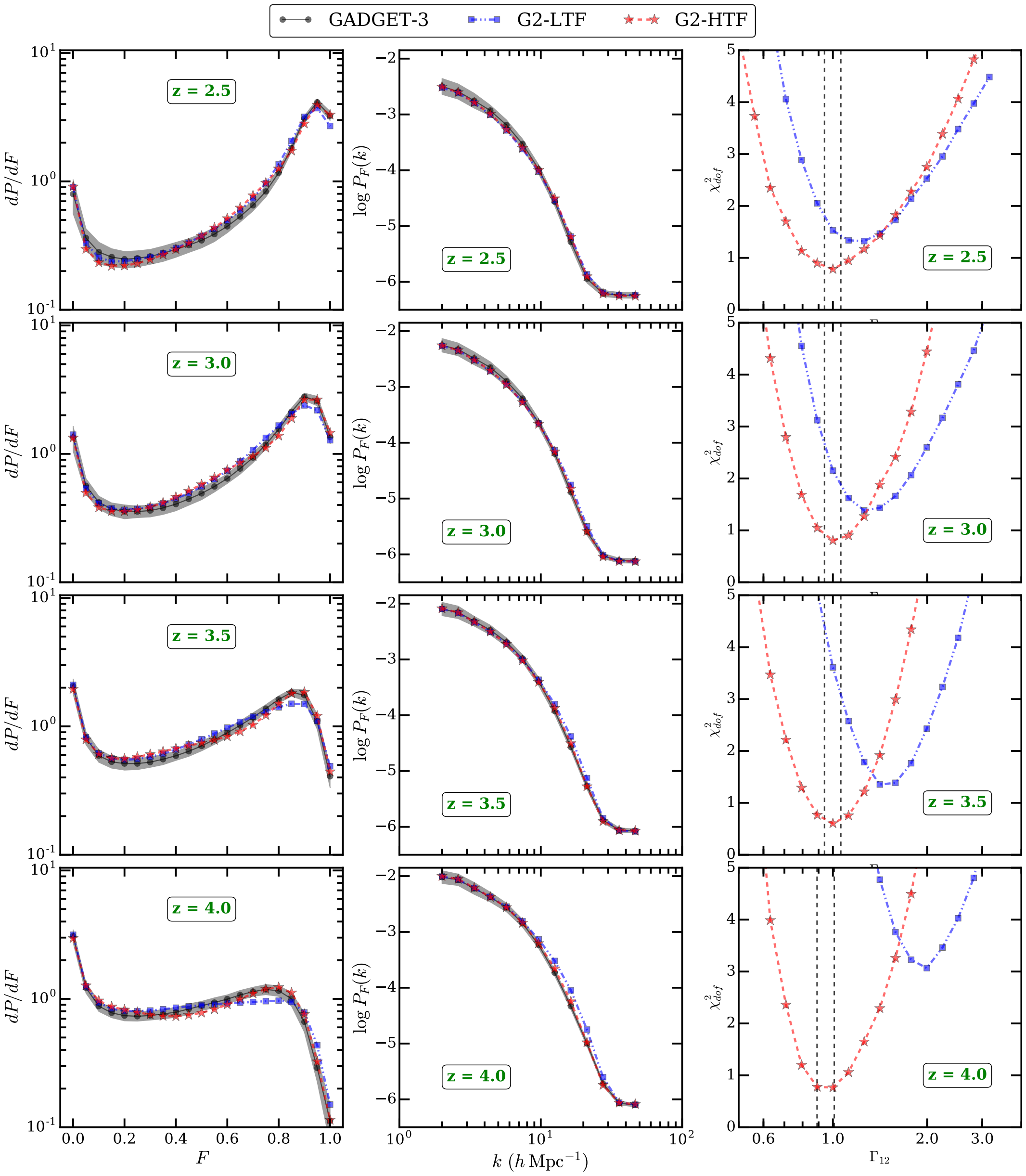

Figure 4.7: Left and middle panels show comparison of the Lyα forest flux PDF and flux PS respectively from GADGET-3 (black circle), G2-LTF (blue squares) and G2-HTF (red stars) where we use $\Gamma_{12} = 1$. The gray shaded region in each panel represents 1σ uncertainty (diagonal elements of the covariance matrix) on the respective statistics from GADGET-3. Right panels show the combined (for flux PDF and PS) reduced χ^2 as a function of Γ_{12} for G2-LTF (blue squares) and G2-HTF (red stars) model. GADGET-3 is used as the reference model with $\Gamma_{12} = 1$. The χ^2 is calculated between statistics from GADGET-3 and G2-LTF or G2-HTF models (see Table 4.3). The 1σ statistical uncertainty on the recovered Γ_{12} for G2-HTF model is indicated by black dashed vertical lines. First, second, third and fourth row from top corresponds to $z = 2.5, 3.0, 3.5$ and 4.0 respectively.

$F = 0.95$) at $z = 3.5$ and 4.0 respectively. This is mainly because the density PS at these redshifts are different for the G2-LTF model as compared to those in the GADGET-3 model (see Fig. 4.6).

4.3.4 Flux PS

Like the density PS, the flux PS is a two point correlation function between pixels of the Lyα transmitted flux (Arinyo-i-Prats et al. 2015; Croft et al. 1998; Kim et al. 2004; McDonald 2003; McDonald et al. 2000; Zhan et al. 2005). The flux PS is known to be sensitive to the astrophysical parameters such as Γ_{HI}, T_0 and γ (Viel et al. 2004a; Zaldarriaga 2002; Zaldarriaga et al. 2001). The procedure for calculating the flux PS is identical to that of the density PS except that the field is the Lyα transmitted flux instead of the density. If we denote the value of flux PS in i^{th} bin of n^{th} mock sample as $P_{F,n}(k_i)$ then the average flux PS in i^{th} bin is obtained from Eq. 4.8 by replacing $P_{\delta,n}(k_i)$ with $P_{F,n}(k_i)$. In similar vein, the covariance matrix elements $C(i,j)$ are obtained from Eq. 4.9. The χ^2 is calculated using the full covariance matrix.

The middle panels in Fig. 4.7 show the flux PS from GADGET-3 (black circles), G2-LTF (blue squares) and G2-HTF (red stars) models for different redshift bins. The grey shaded region corresponds to 1σ error (i.e., diagonal elements of covariance matrix C) on GADGET-3 flux PS. Note that at the redshift of interest the astrophysical parameters Γ_{HI}, T_0 and γ are same for different models. The flux PS for different models behave in a way similar to the density PS. The flux PS from G2-HTF model is consistent within 1σ ($\chi^2_{\mathrm{dof}} \sim 0.5$ from Table 4.3) to that from the GADGET-3 model at all redshifts. However, G2-LTF model at $z = 3.5$ and 4.0 ($\chi^2_{\mathrm{dof}} \sim 3$ from Table 4.3) has slightly more power at scales in the range $220 - 650$ ckpc ($k \sim 30 - 10\,h$ Mpc^{-1}). We also see the suppression of fluctuation in flux as compared to baryon density i.e. errorbars in flux PS are smaller as compared to density PS as noted by Zhan et al. (2005). This is because the transformation (logarithmic suppression) between baryon density and flux is non-linear.

4.3.5 Γ_{12} recovery

The difference we see between the models discussed above in the flux PDF and the flux PS will have direct consequence in the Γ_{HI} values derived. To study this, we treat GADGET-3 as the reference model and see how the value of Γ_{HI} is recovered when we use the G2-LTF and G2-HTF models. For this, we vary Γ_{12} in G2-HTF (or G2-LTF) model and calculate the flux PDF and flux PS. The χ^2 between the flux PDF / PS calculated from GADGET-3 and that from G2-HTF (or G2-LTF) model can be written in the matrix form as (for similar method see Gaikwad et al. 2017b),

$$\chi^2_{(\Gamma_{12})} = [P_{(\Gamma_{12})} - P_{\text{fid}}]\, C^{-1}\, [P_{(\Gamma_{12})} - P_{\text{fid}}]^{\text{T}} \tag{4.10}$$

where P_{fid} and $P_{(\Gamma_{12})}$ is flux statistics (either flux PDF or PS) from GADGET-3 and G2-HTF (or G2-LTF) model respectively. C is covariance matrix as given in Eq. 4.9. Note that we use full covariance matrix for χ^2 estimation.

The right panels in Fig. 4.7 show χ^2 as a function of Γ_{12} from G2-HTF (red stars) and G2-LTF (blue squares) model at four different redshifts. The black dashed vertical lines show the statistical uncertainty in Γ_{12} for G2-HTF model[1]. The Γ_{12} is recovered within 1σ ($d\Gamma_{12} \sim 0.1$) in G2-HTF model at all redshifts whereas G2-LTF model fails to recover the Γ_{12} within 1σ[2]. The Γ_{12} recovered from G2-LTF model at $z = 3.0$ and $z = 4.0$ is higher by a factor of 1.7 and 2 respectively. The minimum χ^2_{dof} for G2-HTF model is also close to 1 indicating the goodness-of-fit.

4.3.6 Wavelet statistics

The wavelet statistic has been used in the past to constrain the thermal history parameters of the IGM (Garzilli et al. 2012; Lidz et al. 2010; Theuns and Zaroubi 2000; Theuns et al. 2002b; Zaldarriaga 2002). Wavelets have finite support in both real and Fourier space and thus can be used to extract the power at scales of interest. This is necessary because large scales (small k) are not sensitive to T_0,

---

[1] The statistical uncertainty corresponds $\chi^2 = \chi^2_{\text{min}} + \Delta\chi^2$ $\Delta\chi^2 = 1$ (Press et al. 1992).

[2] Γ_{12} recovery and its associated error depends on the SNR used in the analysis. The error in the recovered Γ_{12} are quoted for SNR = 25.

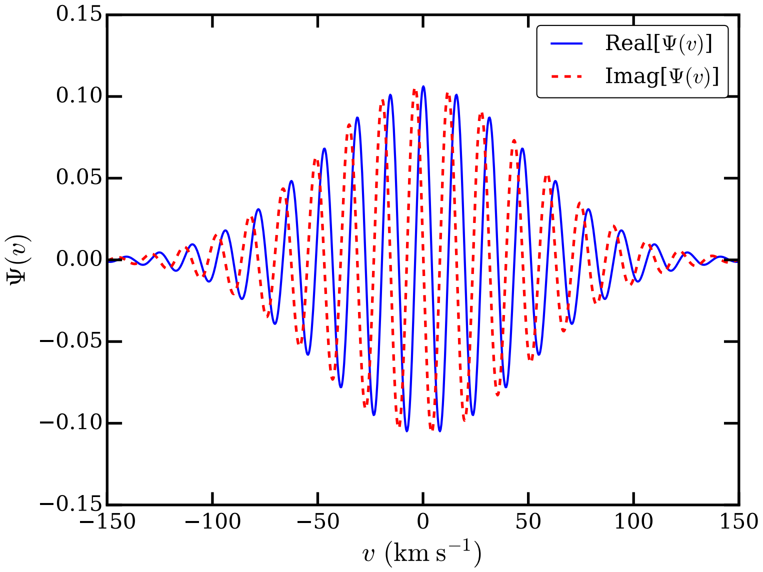

Figure 4.8: An example of a Morlet wavelet ($s_n = 50$ km s^{-1}). Morlet is a sine (or a cosine) function damped by Gaussian. The wavelet of suitable scale is convolved with transmitted Lyα flux. The resulting wavelet amplitudes are sensitive to the mean temperature of the IGM.

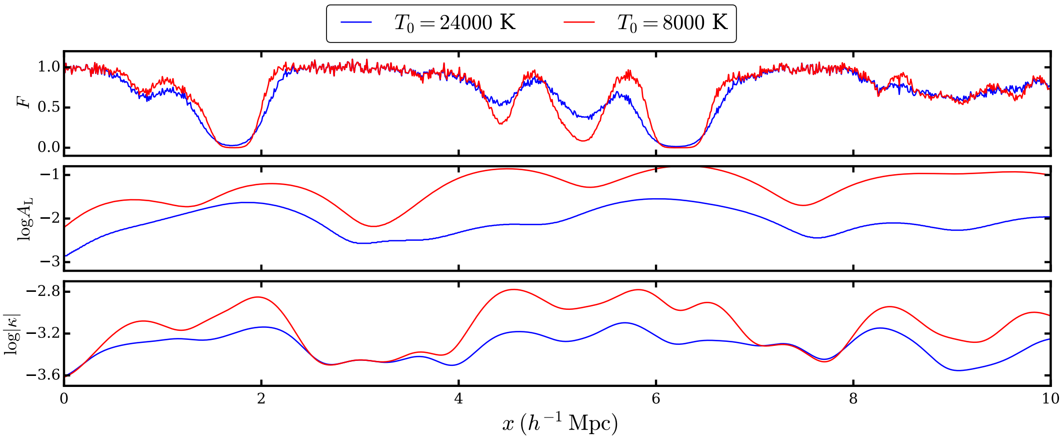

Figure 4.9: Top panel shows the transmitted Lyα flux for mean IGM temperature $T_0 = 24000$ K (blue curve) and $T_0 = 8000$ K (red curve). The middle panel shows the smoothed wavelet power (see Eq. 4.14) for the two models. The bottom panel shows the curvature ($\log|\kappa|$ see Eq. 4.15) for the same two models. It is clear from the bottom two panels that the higher the mean IGM temperature, the smaller the wavelet power and curvature values.

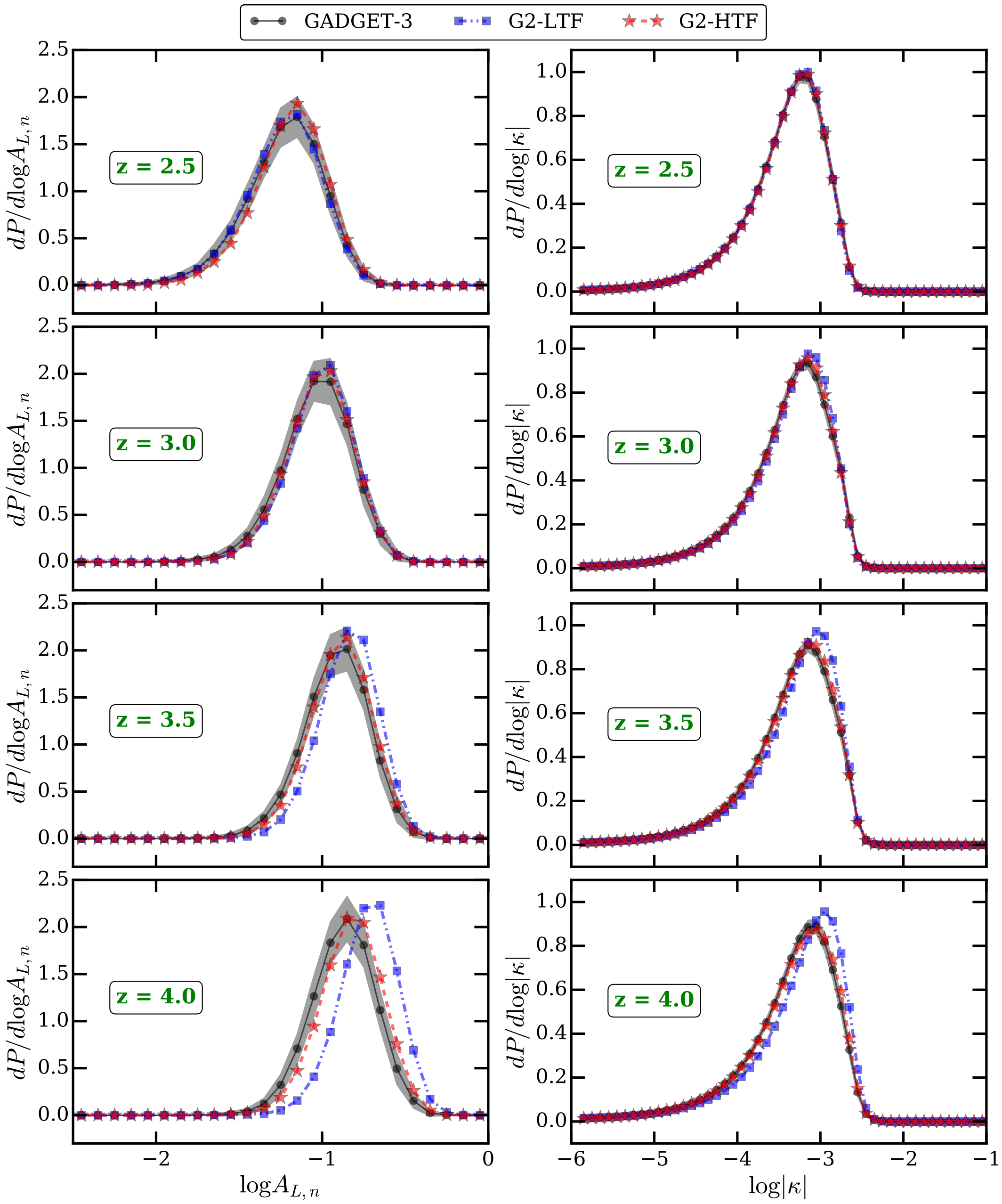

Figure 4.10: Left and right panels show the comparison of the smoothed wavelet power PDF and PDF of curvature parameter of Lyα transmitted flux respectively from GADGET-3 (black circle), G2-LTF (blue squares) and G2-HTF (red stars). The gray shaded region in each panel represents 1σ uncertainty on the respective statistics from GADGET-3. First, second, third and fourth row from top corresponds to $z = 2.5, 3.0, 3.5$ and 4.0 respectively.

γ variation whereas small scales (large k) are contaminated by noise and metal lines in observations (Lidz et al. 2010). We use the "Morlet" wavelet (see Fig. 4.8), usually a sine (or a cosine) function damped by Gaussian, which has the form

$$\Psi(x) = A\ \exp(-i\, k_0\, x)\ \exp\left[-\frac{x^2}{2\, s_n^2}\right] \tag{4.11}$$

where $s_n = 50\ \mathrm{km\,s^{-1}}$, and $k_0 = s_n/2\pi$ is the scale over which power is extracted. As shown by Lidz et al. (2010), this scale is sensitive to T_0 and γ variations. A is a normalization constant fixed by,

$$\int_{-\infty}^{\infty} |\Psi(x)|^2\, dx = 1 \quad . \tag{4.12}$$

The wavelet coefficients are obtained by convolving the Lyα flux (F) with Morlet as,

$$a_n(x) = \int_{-\infty}^{\infty} F(x')\ \Psi(x - x')\, dx' \tag{4.13}$$

The wavelet power is then given by $A_n(x) = |a_n(x)|^2$. Following Lidz et al. (2010), we smooth the wavelet power on scales of $L = 1000\ \mathrm{km\,s^{-1}}$to avoid noisy excursions in wavelet power

$$A_{L,n}(x) = \frac{1}{L} \int_{-\infty}^{\infty} \Theta(|x - x'|; L/2)\ A_n(x')\, dx' \tag{4.14}$$

where $\Theta(|x{-}x'|; L/2)$ is the top-hat filter. Fig. 4.9 shows that the wavelet power is anti-correlated with the temperature of the IGM. The left panels in Fig. 4.10 show the PDF of the smoothed wavelet power ($A_{L,n}$) from GADGET-3 (black circles), G2-LTF (blue squares) and G2-HTF (red stars) models for different redshift bins. As the IGM equation of state parameters T_0 and γ evolve with redshift, the peak of the PDF also shifts accordingly. The wavelet PDF for G2-HTF model is in very good agreement with GADGET-3 model ($\chi^2_{\rm dof} \sim 0.5$ see Table 4.3) at all redshifts. In contrast, the wavelet PDF is systematically different at higher redshifts in the G2-LTF than in the GADGET-3 model ($\chi^2_{\rm dof} > 4$ at $z = 3.5$ and 4). This is because density field power spectrum has larger power at small scales in G2-LTF model (see Fig. 4.7) thus affecting the $A_{L,n}$ measurement. We rerun G2-LTF with best fitted Γ_{12} (shown in Fig. 4.7) and found the same results.

4.3.7 Curvature statistics

Similar to the wavelet analysis, Becker et al. (2011) introduced a curvature statistics to measure the amount of small-scale structure in the Lyα forest. The curvature κ is defined as,

$$\kappa \equiv \frac{F''}{[1 + (F')^2]^{3/2}} \tag{4.15}$$

where F', F'' is first and second derivative of Lyα transmitted flux respectively. Fig. 4.9 shows that, for higher IGM temperature, the value of the curvature κ is smaller. This statistics is suitable for obtaining the IGM temperature at characteristic overdensity which is found to be an almost one-to-one function of the mean curvature regardless of γ (Becker et al. 2011; Boera et al. 2014; Padmanabhan et al. 2014, 2015; Upton Sanderbeck et al. 2016). Following the earlier works, Padmanabhan et al. (2015) have shown that the mean and the percentiles of the curvature distribution function can be used to obtain constraints on the equation of state. The right panels in Fig. 4.10 show the PDF of the curvature from GADGET-3 (black circles), G2-LTF (blue squares) and G2-HTF (red stars) models for different redshift bins. Again the G2-HTF model is in good agreement with GADGET-3 model at all redshifts ($\chi^2_{\rm dof} \sim 0.9$ from Table 4.3). On the other hand, curvature is systematically more at higher redshifts in G2-LTF than GADGET-3 model. The reduced χ^2 in this case is as high as ~ 32. This is because small scale fluctuations in density field are large for G2-LTF model (see Fig. 4.7) which affects the curvature κ measurement.

4.3.8 Column Density Distribution function (CDDF)

The next three statistics treat the Lyα forest as a composition of discrete Lyα absorbers. Each Lyα line can be fitted with Voigt profile having 3 free parameters column density ($N_{\rm HI}$), linewidth (b) parameter and line center (λ_c). We used VIPER to decompose the Lyα forest into multi-component Voigt profile. We refer the reader to Chapter 3 for more details on VIPER. A sample of Voigt profile decomposition using VIPER is shown in Fig. 4.11. Visually, the match between input and fitted spectra is good. We fit 24000 (at each z) spectra for the three

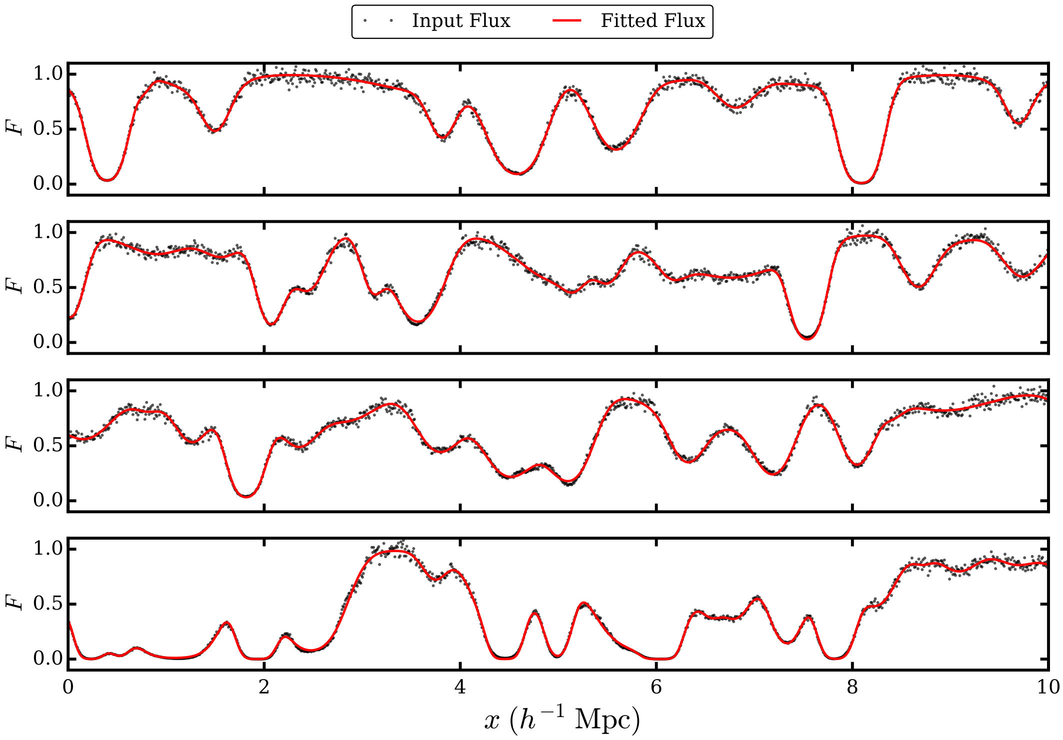

Figure 4.11: Voigt profile decomposition using VIPER. The input flux is shown by black circles and fitted flux is shown by red solid curve. First, second, third and fourth panel from top shows the spectra from simulation box at $z = 2.5, 3.0, 3.5, 4.0$ respectively. All the spectra are shown for G2-HTF model.

models at four different redshifts which can be done quite efficiently using our automated routine.

The CDDF, $f(\mathrm{N_{HI}}, z)$, is a bivariate distribution that describes the number of absorption lines in the column density range $\log \mathrm{N_{HI}}$ and $\mathrm{d}\log \mathrm{N_{HI}}$ and in the redshift range z to $z + dz$. The CDDF is sensitive to Γ_{HI} (Gaikwad et al. 2017c; Gurvich et al. 2017; Kollmeier et al. 2014; Schaye 2001; Shull et al. 2015, 2012; Viel et al. 2017). Left panels in Fig. 4.12 show the CDDF from GADGET-3 (black circles), G2-LTF (blue squares) and G2-HTF (red stars) models for different redshift bins. We have not accounted for incompleteness of the sample in the calculation of redshift path length. However, the reduced χ^2 is calculated from lines above completeness limit. Again the CDDF from G2-HTF model is in good agreement ($\chi^2_{\mathrm{dof}} \sim 0.3$ from Table 4.3) with that from GADGET-3 at all

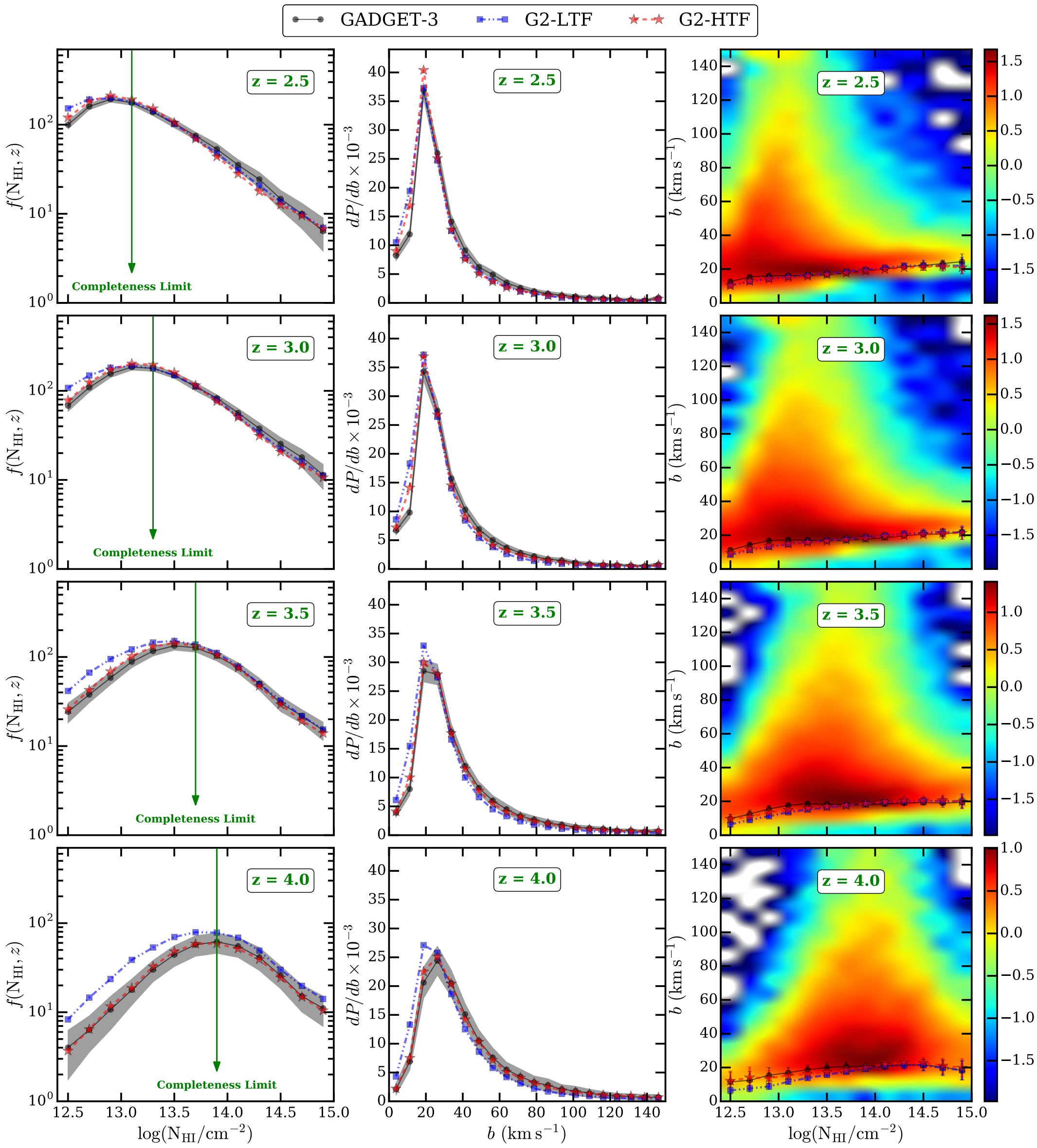

Figure 4.12: Left, middle and right panels show the comparison of the CDDF, b parameter distribution function and b vs $\log N_{HI}$ correlation respectively from GADGET-3 (black circle), G2-LTF (blue squares) and G2-HTF (red stars). The gray shaded region in first two columns (except b vs $\log N_{HI}$ panels) represents 1σ uncertainty on the respective statistics from GADGET-3. First, second, third and fourth row from top corresponds to $z = 2.5, 3.0, 3.5$ and 4.0 respectively. The incompleteness of the sample is not accounted for in the calculation of CDDF. The completeness limit is shown by solid vertical green line. In the right panel, the colour scheme indicates density of points in logarithmic units for GADGET-3. The lower envelope in right panel is obtained by calculating 10$^{\text{th}}$ percentile of b values in $\log N_{HI}$ bin. We calculated the χ^2 for the three statistics from sample of lines only above the completeness limit. However, in the plot we show the results for full sample.

redshifts. However, G2-LTF model predicts more number of lines at lower column densities $\log \mathrm{N_{HI}} < 13.5$. This is because density field in G2-LTF is less smooth as compared to that from GADGET-3 model. The features arising from variation in density field of G2-LTF model is identified and fitted by VIPER as lines with smaller column densities (for example see the region between $8-9$ cMpc in Fig. 2.9). Thus the CDDF from G2-LTF and G2-HTF models are consistent within 1σ with that from GADGET-3 model above the completeness limit.

4.3.9 Linewidth (b parameter) distribution function

The middle panels in Fig. 4.12 show the linewidth distribution, which is sensitive to thermal history and unknown turbulent motions in the IGM (Davé and Tripp 2001; Gaikwad et al. 2017c; McDonald et al. 2001; Schaye et al. 1999, 2000; Viel et al. 2017), from GADGET-3 (black circles), G2-LTF (blue squares) and G2-HTF (red stars) models for different redshift bins. Again, unlike the G2-LTF model the linewidth distribution from G2-HTF model is consistent within 1σ with that from GADGET-3 model. The reduced χ^2 between G2-HTF (G2-LTF) and GADGET-3 model is ~ 0.3 (~ 0.35). Note that for the χ^2 analysis, we use the b parameter distribution calculated from lines only above the completeness limit. However, the b parameter distribution plotted in Fig. 4.12 is calculated from all the lines in sample.

4.3.10 b versus $\log \mathrm{N_{HI}}$ scatter

The right panels in Fig. 4.12 show the b versus $\log \mathrm{N_{HI}}$ scatter for GADGET-3 model. The color scheme represents density of points in logarithmic units. We summarize the χ^2 calculated from this 2D distribution between G2-LTF, G2-HTF and GADGET-3 model in Table 4.3. The $\chi^2_{\mathrm{dof}} \sim 0.25$ for the G2-LTF and G2-HTF models indicates good agreement with the GADGET-3 model. Another way to assess the goodness of fit is to match the lower-envelope in b versus $\log \mathrm{N_{HI}}$ plot. The lower-envelope in the b versus $\log \mathrm{N_{HI}}$ plot has been used in the past to constrain the thermal history parameters T_0 and γ (McDonald et al. 2001; Schaye et al. 1999, 2000). Following Garzilli et al. (2015), we obtain the lower-envelope by calculating the 10^{th} percentile of b values in $\log \mathrm{N_{HI}}$ bin. The lower-envelope in b versus $\log \mathrm{N_{HI}}$ plot from G2-HTF (red stars) is again in agreement with GADGET-3 (black circles) within 1σ. On the other hand, at $z = 3.5$ and 4.0, the lower-envelope from G2-LTF (blue stars) is consistently smaller at $\log \mathrm{N_{HI}} < 13.5$ than that from GADGET-3 model. This can again be attributed to extra absorption line features with smaller $\log \mathrm{N_{HI}}$ identified by VIPER. It should be noted that, above the completeness limit the lower envelope from both the G2-LTF and G2-HTF models are in very good agreement with that from GADGET-3 model at all redshifts.

4.3.11 Effect of change in thermal history

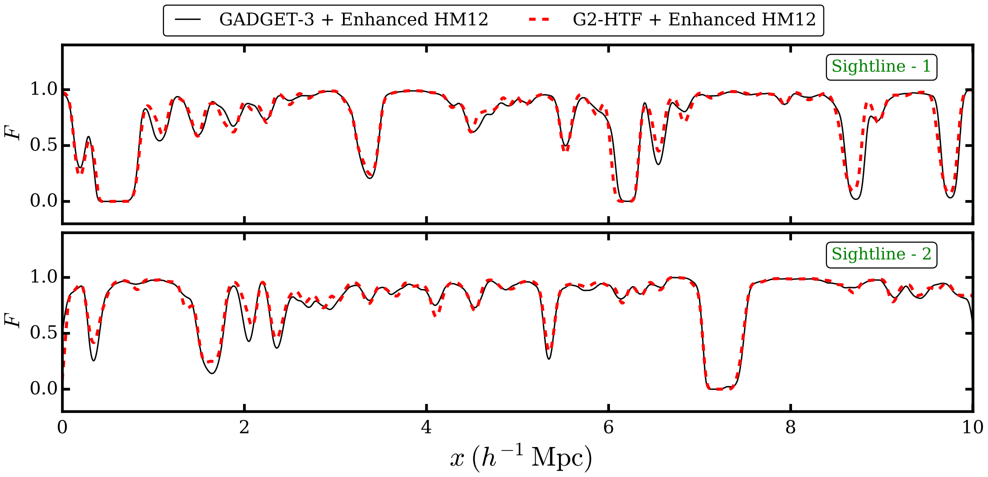

Figure 4.13: Line of sight comparison of Lyα flux (F) for GADGET-3 (black solid line) and G2-HTF (red dashed line) simulation boxes at $z = 2.5$ along two different sightlines as shown in top and bottom panels. GADGET-3 simulation is performed with an enhanced photo-heating rates (see section 4.3.11 for details). For G2-HTF model, we used enhanced HM12 photo-heating rates in CITE. The Lyα flux F along the sightline match very well for the two models. The Lyα flux is not convolved with any LSF and no noise is added to the flux.

The comparison between the different models discussed till now has been performed for the HM12 UVB model. It is, however, important to validate our method for different UVB models where thermal the history is significantly different from that in the case of the HM12 UVB. In order to explore this, we follow

Becker et al. (2011) and modify the photo-heating rates of species i =[H I, He I, He II] as $\epsilon_i = a \times \epsilon_i^{\rm HM12}$ where, $\epsilon_i^{\rm HM12}$ is HM12 photo-heating rates of specie i. We choose $a = 2.933$ such that the T_0 is increased by factor of ~ 2 while γ remains same at all redshifts. With this updated photo-heating rates we performed a GADGET-3 (with QUICK_LYALPHA flag) and G2-HTF simulation with the initial conditions same as described in section 2.3.1. It is important to emphasize here that we do *not* perform a GADGET-2 simulation again, rather we only modify the HM12 photo-heating rates while running CITE in the post-processing stage on the same simulation run earlier. Note that at the initial redshift $z = 6$, we use $T_0 = 14543$ K and $\gamma = 1.51$ in CITE consistent with GADGET-3 for enhanced HM12 photo-heating rates at that redshift. Fig. 4.13 shows comparison of Lyα flux from GADGET-3 and G2-HTF model for enhanced HM12 photo-heating rates. The flux from the two models match very well with each other. We also calculate the line of sight density PS, flux PDF, flux PS, wavelet PDF, curvature PDF, CDDF, b parameter distribution and b vs $\log \rm N_{HI}$ distribution for these models and calculated the χ^2. The reduced χ^2 for these statistics are $\sim 0.33, 0.67, 0.34, 0.41, 0.59, 0.23, 0.37$ and 0.28 respectively. Using this model we were also able to recover Γ_{12} within 1σ from flux PDF and flux PS statistics. This shows that G2-HTF model is consistent with GADGET-3 model for a significantly different thermal history. Thus for a range of physically motivated photo-heating rates from UVB calculations (such as Khaire and Srianand 2015a,b), we can easily probe the $T_0 - \gamma$ parameter space and calculate Lyα flux in G2-HTF model without performing full GADGET-3 simulation.

We now highlight the advantages of using our method for simulating Lyα forest:

- **Efficiency :** Table 4.4 summarizes the CPU time consumption in various parts of the code. Significant fraction of time is spent in evolution of Δ, v and T in both codes. However, unlike GADGET-3 we need to evolve Δ, v and T in G2-HTF (or G2-LTF) only once. To vary astrophysical parameters in G2-HTF, we just need to vary UVB in CITE. This allows one to probe T_0 and γ parameter space efficiently. For example, the time (per core) required to simulate Lyα forest for 10 different UVB in GADGET-3 is ~ 67 days whereas for GADGET-2 is ~ 8 days.

- **Accuracy :** The Lyα flux statistics from our method is accurate within 5 percent with that from GADGET-3. We have shown that using our method $\Gamma_{\rm HI}$ can be recovered within 1σ statistical uncertainty.
- **Flexibility :** In addition to HM12, it is straightforward to incorporate other UVB such as Faucher-Giguère et al. (2009); Khaire and Srianand (2015a,b) in CITE and evolve the temperature without performing full hydrodynamic simulation. CITE can be run in either equilibrium or non-equilibrium ionization evolution mode. It is easy to incorporate cooling due to metals in CITE by changing cooling rate tables (Gaikwad et al. 2017b; Wiersma et al. 2009, for similar analysis)[1].

4.4 Summary

In this chapter we validated our method (GADGET-2 post-process with CITE and GLASS) of modeling the Lyα forest. For low-z (and low resolution) Lyα forest we compare our results with other simulations in the literature. Whereas for high-z (and high resolution) Lyα forest where pressure smoothing effects are important, we show the consistency by comparing our simulations with full hydrodynamic simulation GADGET-3. The main results of our analysis are as follows.

- We compare our results with other low-z simulations in the literature using three predictions. These are, (i) thermal history parameters: our simulation predicts $T_0 \sim 5000$ K and $\gamma \sim 1.6$ in the redshift range $z = 0.1$ to 0.45. These values are shown to be insensitive to our choice of T_0 and γ at an initial redshift, $z_1 = 2.1$; (ii) distribution of baryons in phase diagram at $z = 0$: We find $\sim$ 34 per cent of baryons are in diffuse phase, $\sim$ 29 per cent in WHIM, $\sim$ 18 per cent in hot halo and $\sim$ 19 per cent in condensed phase and (iii) the correlation between baryon overdensity Δ vs H I column density, $N_{\rm HI}$, in the redshift range $0.2 < z < 0.3$: we find $\Delta = 34.8 \pm 5.9 \, (N_{\rm HI}/10^{14})^{0.770 \pm 0.022}$. We show that all these predictions compare well with those of low-z simulations in the literature that include

---

[1] http://www.strw.leidenuniv.nl/WSS08/

Table 4.4: Consumption of CPU time (in hours) per core for various tasks of the code for a cosmological run from $z = 99$ to $z = 2.0$

Step	Description[a]	GADGET-3	G2-HTF
1	Δ, v and T Evolution	156	108
2	CITE (T Evolution)[b]	–	3.5
3	Grid calculation[c]	3.5	4
4	GLASS[d]	1	1
-	Total time to run 10 different UVB model[e]	1605 (67 days)	193 (8 days)

[a] The analysis was done using 256 core on IUCAA PERSEUS cluster.

[b] CITE evolves the temperature of the SPH particles from $z = 6.0$ to $z = 2.0$. Temperature is evolved internally in GADGET-3.

[c] We used modified smoothing kernel for G2-HTF or G2-LTF as given in Eq. 2.11. The time is given for 10240 random sightlines through simulation box.

[d] We apply equation of state as given in Eq. 2.16 for G2-HTF and G2-LTF models. The numbers are given for total 10×2048 simulated Lyα forest spectra. We splice 5 sightline to cover redshift path length for a single spectra.

[e] The total time required to run 10 UVB model for GADGET-3 is sum of time consumed by steps 1, 3 and 4 (i.e. 160.5×10 hours). Unlike GADGET-3, step 1 is performed only once for G2-HTF or G2-LTF models. For different UVB models, we follow step 2-4 in the post-processing stage. Hence the total time required to run 10 UVB model for G2-LTF or G2-HTF model is (108 hours + 8.5 hours $\times$ 10 = 193 hours).

different feedback processes at varied levels. Feedback processes such and galactic winds or AGN feedback are not incorporated in our simulations. However, as shown by Shull et al. (2015) these processes are not expected to severely influence the Lyα forest statistics.

- We generate the Lyα forest spectra by shooting random sightlines through simulation box in all the 3 models. We compare the G2-LTF and G2-HTF with the GADGET-3 model using 8 different statistics, namely: (i) 1D density field PS, (ii) flux PDF, (iii) flux PS, (iv) wavelet PDF, (v) curvature PDF, (vi) column density distribution function, (vii) linewidth distribution and (viii) b vs $\log \mathrm{N_{HI}}$ correlation, at four different redshift $z = 2.5, 3.0, 3.5$ and 4.0. Treating the GADGET-3 model as the reference, we demonstrate that the H I photoionization rate (Γ_{HI}) can be recovered, using flux PDF and flux PS statistics, within 1σ statistical uncertainty using the G2-HTF model. We find that the G2-HTF model is in general very good agreement (within 1σ) with GADGET-3 model at all redshifts.
- Using enhanced HM12 photo-heating rates, we obtain a thermal history such that T_0 is increased by a factor of ~ 2. We show that our method for such significantly different thermal history is consistent (in 1σ) with GADGET-3 simulation.

Our method to simulate the Lyα forest is computationally less expensive, flexible to incorporate changes in UVB, metallicity, non-equilibrium ionization evolution etc. and accurate to within 5 percent. This method can be used in future to explore T_0, γ and Γ_{HI} parameter space and to simultaneously constrain these quantities from observations.

5 | Parameter Estimation: $\Gamma_{\rm HI}$ at $z < 0.5$

This chapter contains material that has been published in Gaikwad et al. (2017b,c). In this chapter we use our tools (i.e., CITE, GLASS, VIPER and Lyα statistics module) developed for efficient modeling and analyzing the Lyα in Chapter 2 and 3 to constrain the astrophysical parameter from observations.

5.1 Introduction

After the H I reionization ($z \geq 5.5$, Becker et al. 2001; Bolton et al. 2011; Fan et al. 2001, 2006; Khaire et al. 2016; Madau et al. 1999; Planck Collaboration et al. 2016; Robertson et al. 2010), the IGM at subsequent epoch is maintained at a highly ionized state by UVB radiation ($\lambda \leq 912$ Å). This UVB is contributed by radiation from blackhole accretion in QSO and stellar light escaping from galaxies (see for example, Fardal et al. 1998; Haardt and Madau 1996; Miralda-Escude and Ostriker 1990; Shull et al. 1999). The stellar contribution to UVB depends crucially on the fraction of ionizing photons escaping the galaxies known as the escape fraction ($f_{\rm esc}$). The $f_{\rm esc}$, in principle, depends on various physical factors such as the galaxy mass, morphology, composition of the interstellar medium (ISM), spatial distribution of gas and supernova rates (Benson et al. 2013; Cen and Kimm 2015; Fernandez and Shull 2011; Gnedin et al. 2008; Kim et al. 2013; Ricotti and Shull 2000; Roy et al. 2015). As a result, there is no consensus among different models of $f_{\rm esc}$. Measuring $f_{\rm esc}$ directly from observations too is quite challenging. The reported values of $f_{\rm esc}$ at $2 \leq z \leq 4$ vary between 0.01 to 0.2 (Boutsia et al. 2011; Cooke et al. 2014; Grazian et al. 2016; Iwata et al.

2009; Micheva et al. 2015; Mostardi et al. 2015; Nestor et al. 2013; Siana et al. 2015; Smith et al. 2016; Vasei et al. 2016). At $z < 2$, apart from the detection of high $f_{\rm esc}$ in few individual galaxies (Bergvall et al. 2006; Borthakur et al. 2014; Deharveng et al. 2001; Izotov et al. 2016b; Leitet et al. 2013; Leitherer et al. 2016), the 3σ upper limits on average $f_{\rm esc}$ obtained by stacking samples of galaxies is ≤ 0.02 (Bridge et al. 2010; Cowie et al. 2009; Rutkowski et al. 2016; Siana et al. 2010).

An alternate way of constraining $f_{\rm esc}$ (and hence the stellar contribution to the UVB) is by measuring H I photoionization rate ($\Gamma_{\rm HI}$) (see Inoue et al. 2006; Khaire et al. 2016). The $f_{\rm esc}$ also determines the shape of the UVB (Khaire and Srianand 2013) which is important for modeling the distribution of ions in the IGM detected in the QSO spectra (Finn et al. 2016; Oppenheimer et al. 2016; Rahmati et al. 2016; Shull et al. 2014). The UVB estimated at any given redshift z_0 depends on emissivities of radiating sources and the IGM opacity (contributed mainly by the partial Lyman Limit Systems (LLS) with $N_{\rm HI} \geq 3 \times 10^{16}$ cm^{-2} and LLS with $N_{\rm HI} \geq 10^{17}$ cm^{-2}) over the large redshift range $z \geq z_0$. Therefore, measurements of $\Gamma_{\rm HI}(z_0)$, can be useful to constrain the IGM opacity evolution at $z \geq z_0$, especially at low-z where it is ill-constrained.

$\Gamma_{\rm HI}$ is usually constrained in the literature using three methods: (i) The first method uses the H I absorption in the proximity of QSOs (Bajtlik et al. 1988; Calverley et al. 2011; Dall'Aglio et al. 2009; Kulkarni and Fall 1993; Scott et al. 2000; Srianand and Khare 1996). The main uncertainties in this method arise because of the anisotropies in the QSO emission (Kirkman and Tytler 2008; Schirber et al. 2004) and possible density enhancements around the QSO host galaxies (Faucher-Giguère et al. 2008c; Guimarães et al. 2007; Rollinde et al. 2005). (ii) The second method, mainly useful at low-z, is based on the measured Hα surface brightness from the outskirts of nearby galaxies ($z \sim 0$) and high velocity cloud at the edges of our galaxy (Adams et al. 2011; Kutyrev and Reynolds 1989; Madsen et al. 2001; Songaila et al. 1989; Vogel et al. 1995; Weymann et al. 2001). However this measurement too is uncertain because of the assumptions made about the geometries of the Hα emitting gas (see Shull et al. 2014). (iii) The third method of constraining $\Gamma_{\rm HI}$ is by simulating the observed properties of the Lyα forest (far away from the proximity of QSOs) such as the H I column density distribution function (CDDF) (Kollmeier et al. 2014; Shull et al. 2015), the flux probability distribution function (PDF) and the flux power spectrum (PS), which forms the basis of the analysis presented in this paper.

The basic idea behind using the Lyα forest to constrain $\Gamma_{\rm HI}$ is that, under the fluctuating Gunn-Peterson approximation (hereafter FGPA; see Croft et al. 1998; Gunn and Peterson 1965; Weinberg et al. 1998), the Lyα optical depth scales as $\Gamma_{\rm HI}^{-1}$ and hence can be used to constrain it. Observed statistical properties of the Lyα forest are compared with those from an appropriate model with $\Gamma_{\rm HI}$ as one of the free parameters. The other free parameters in these models are those describing the thermal history of the IGM and the cosmological parameters which can be degenerate with $\Gamma_{\rm HI}$. To estimate the uncertainty in $\Gamma_{\rm HI}$ at $2 \leq z \leq 4$ due to its degeneracy with other parameters, Bolton and Haehnelt (2007); Bolton et al. (2005); Faucher-Giguère et al. (2008b) used scaling relations derived from their hydrodynamical simulations (Table 4 in Bolton and Haehnelt 2007). At these redshifts, it is well known from numerical simulations that the Lyα forest arises from the low-density diffuse medium which accounts for 90 per cent of the baryons (Bolton and Becker 2009; Cen et al. 1994; Faucher-Giguère et al. 2008a; Hernquist et al. 1996; Miralda-Escudé et al. 1996; Paschos and Norman 2005; Rauch et al. 1997; Zhang et al. 1995). Hence a relatively simple model (Bi and Davidsen 1997; Choudhury et al. 2001) is sufficient to reliably constrain $\Gamma_{\rm HI}$. On the other hand at low-z ($z \leq 1.6$), only a small fraction ($\sim 30-40$ per cent) of the baryons are in diffuse medium responsible for Lyα forest (Davé et al. 2010; Davé and Tripp 2001; Shull et al. 2015; Smith et al. 2011; Tepper-García et al. 2012; Theuns et al. 1998a). It turns out that a significant fraction ($\sim 30-50$ per cent) of the baryons are in a phase known as the warm hot intergalactic medium (WHIM) (Cen and Fang 2006; Cen and Ostriker 1999, 2006; Davé et al. 2001, 2010; Lehner et al. 2007; Shull et al. 2012; Smith et al. 2011) and they are difficult to detect in either emission or absorption in the UV/optical bands. Hence to measure $\Gamma_{\rm HI}$ at low-z, one needs simulation incorporating all these effects. Furthermore there is also a strong possibility that SNe and AGN feedback processes can inject thermal energy into the IGM which may change the density-temperature distribution (i.e the phase diagram) of the baryons (Davé et al. 2010; Shull et al. 2015; Smith et al. 2011; Tepper-García et al. 2012), thus probably affecting the Lyα forest observable.

To observe the low-z ($z \leq 1.6$) Lyα forest, one needs the UV spectrograph onboard space based telescope. Thanks to a large survey using the Cosmic Origins Spectrograph (COS) onboard the Hubble Space Telescope (HST) (Danforth et al. 2016), there are now constraints on $\Gamma_{\rm HI}$, e.g., by using column density distribution of low-z Lyα forest (Kollmeier et al. 2014; Shull et al. 2015), and by modeling the observed metal abundances of ions (Shull et al. 2014) using CLOUDY (Ferland et al. 1998). It turns out that there is a tension between the inferred $\Gamma_{\rm HI}$ by Kollmeier et al. (2014) and Shull et al. (2015), where both use Lyα forest data by Danforth et al. (2016) but different simulations. The inferred $\Gamma_{\rm HI}$ values disagree by a factor of ~ 2.5. Given such a wide disagreement, it is worth taking an independent closer look at the $\Gamma_{\rm HI}$ measurements at low-z using the Lyα forest, in particular a careful analysis of the systematics in the data as well as modeling uncertainties.

The main aim of this chapter is to measure $\Gamma_{\rm HI}$ from the Lyα forest data by Danforth et al. (2016), using three different statistics, namely the flux PDF, the flux PS and CDDF. Our method to model Lyα forest at low-z and its consistency with other simulations in the literature is described in Chapter 2 and 4 respectively. The advantage of this method is that it is computationally less expensive and sufficiently flexible to account for variations in the thermal history. Our analysis allows us to study the degeneracy between $\Gamma_{\rm HI}$ and parameters related to the thermal history. The other significant step in our analysis is that we calculate the errors on $\Gamma_{\rm HI}$, unlike Kollmeier et al. (2014); Shull et al. (2015), by estimating the error covariance matrix from the simulations using a method similar to Rollinde et al. (2013), thus avoiding any non-convergence that may arise from the limited sample of the observed data.

The plan of this chapter is as follows: The observational data used in our analysis are discussed in Section 5.2. Details of the simulated Lyα forest spectra are discussed in Section 5.3. The three statistics used in this paper (i.e., the flux PDF, the flux PS and CDDF) and the associated errors are discussed in Section 5.4. The main results of our work are discussed in Section 5.5, where we match the simulations with the observed data to constrain $\Gamma_{\rm HI}$. We also discuss the various statistical and systematic uncertainties in the measured $\Gamma_{\rm HI}$. Finally we use the ionizing background computed by Khaire and Srianand (2015b) using the updated emissivities and IGM opacities to constrain $f_{\rm esc}$ from the evolution of $\Gamma_{\rm HI}$. We summarize our findings in Section 5.7. We use Γ_{12} to express $\Gamma_{\rm HI}$ in units of 10^{-12} s^{-1}.

5.2 HST-COS QSO absorption spectra

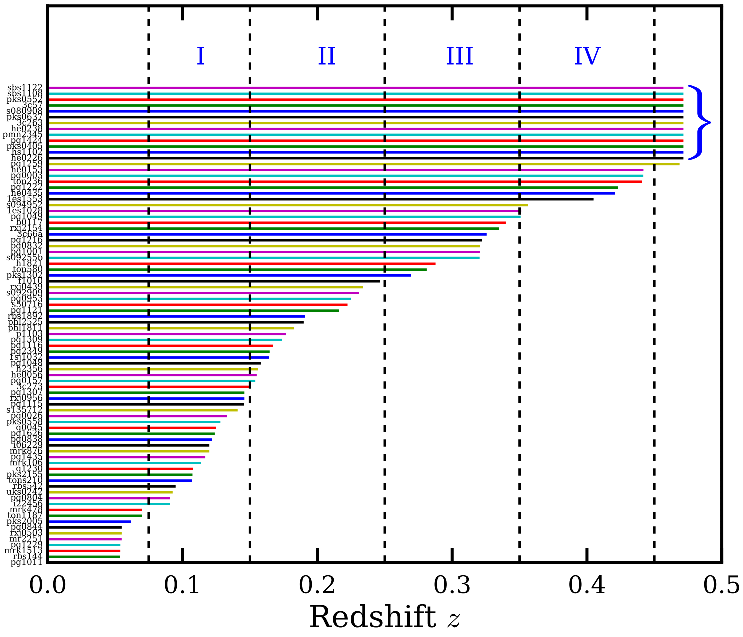

Figure 5.1: The redshift range covered by the Lyα forest for the 82 HST-COS spectra used in this work (see section 3.2). The vertical dashed lines show the redshift bins with centers at $z = 0.1125, 0.2, 0.3, 0.4$ and width $\Delta z = 0.075, 0.1, 0.1, 0.1$ respectively. The redshift bins used in this work are shown by roman numerals. The sharp cutoff shown by blue curly bracket is arising from the red wavelength cutoff of the COS-160M grism used (at $z = 0.48$). In these cases the COS spectra do not cover the Lyα emission from the QSOs.

We used the publicly available data from a survey[1] of low redshift Lyα performed by Danforth et al. (2016) using HST-COS. The sample consists of 82

---

[1] https://archive.stsci.edu/prepds/igm/

UV-bright QSO sightlines, with the QSOs being distributed across the redshift range 0.0628 to 0.852. The observations were carried out between July 2009 and August 2013. Using accurate data reduction process and careful subtraction of the background, Danforth et al. (2016) have produced high signal to noise ratio (SNR) Lyα forest spectra in the observed wavelength range 1100Å to 1800Å. In addition, Danforth et al. (2016) have fitted the continuum and identified several thousand absorption features using a semi-automated procedure. These absorption features arise not only from H I Lyman series lines from the IGM but also from other intervening absorbers and from the Galactic interstellar medium. The redshift range for Lyα lines covered by each sightline is shown in Fig.5.1. The sharp cutoff shown by blue curly bracket in the figure is because of the limited wavelength range covered by the spectrograph. We assume that a region of comoving size of up to $25h^{-1}$ cMpc around a QSO can be affected by the proximity effect of the QSO itself (Lidz et al. 2007), hence we exclude the corresponding section blueward of the Lyα emission line.

We divide the data into 4 different redshift bins. Three bins are centered on $z = 0.2, 0.3, 0.4$ with a width of $\Delta z = 0.1$. We chose the lowest redshift bin $z = 0.1125$ with $\Delta z = 0.0375$ to avoid the contamination from the foreground geo-coronal line emission at $z < 0.075$. We are then left with Lyα absorption from $50, 31, 16, 12$ sightlines each in the redshift bins with $z = 0.1125, 0.2, 0.3, 0.4$, respectively. The redshift bins chosen for analysis are indicated by roman numerals I, II, III and IV in Fig. 5.1. The values of the SNR for the Lyα forest spectra vary between 5 and 17. Each observed spectrum has a resolution of ~ 17 km s^{-1}. Table 5.1 summarizes the properties of the observed Lyα forest data in the four identified redshift bins. The table contains the redshift range of observation ($z_{\rm obs}$), the redshift of simulation box used ($\overline{z}_{\rm sim}$) for comparison, number of sightlines used and SNR range for each bin. A sample observed spectrum (towards the QSO 3C57) is shown in the top panel of Fig. 5.2. In addition to the Lyα absorption, the spectrum contains higher Lyman-series and various metal absorption lines as shown in the figure. We fit these lines with gaussian and replace them with appropriate continuum added with gaussian random noise (with the same SNR) as shown in middle panel of Fig. 5.2. We use these clean spectra (i.e., metal line and higher order Ly-series line removed) to match observations with

Table 5.1: Details of the HST-COS data used in different redshift bins. ($\overline{z}_{\rm sim}$ is redshift of simulation box used for comparison)

Redshift bin	$z_{\rm obs}$	$\overline{z}_{\rm sim}$	Number of QSOs	SNR Range
I	0.075 - 0.15	0.1	50	14.5 - 16.9
II	0.15 - 0.25	0.2	31	13.0 - 14.4
III	0.25 - 0.35	0.3	16	6.3 - 13.3
IV	0.35 - 0.45	0.4	12	5.8 - 6.9

simulations using flux PDF and flux PS. We fitted all the observed spectra with multi-component Voigt profile using our automated code VIPER details of which are given in in Chapter 3.

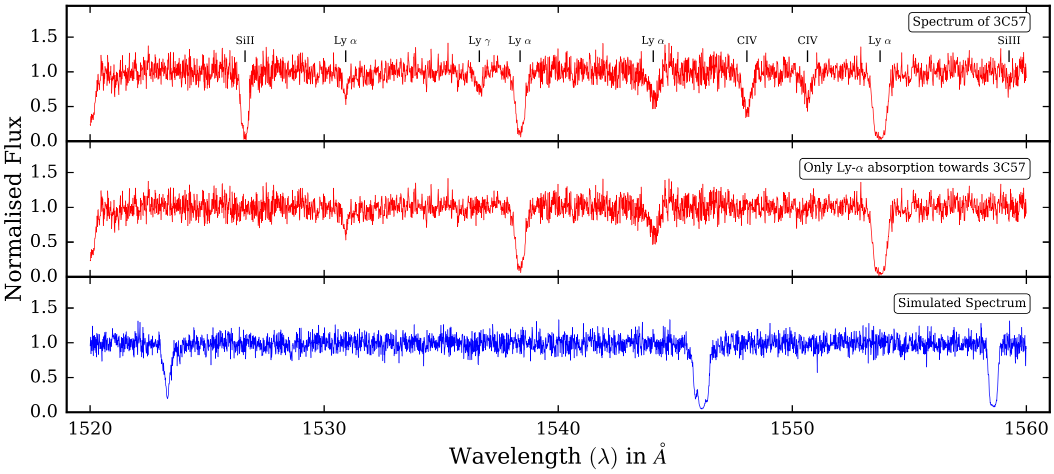

Figure 5.2: *Top panel* shows the observed Lyα forest towards the QSO 3C57. H I Ly series and metal lines as identified by Danforth et al. (2016) are also marked. *Middle panel* shows the same spectrum after these lines are removed and replaced by a continuum added with random noise with the same SNR as in the original spectrum (see section 5.2). *Bottom panel* shows the simulated spectrum towards a random line of sight in our simulation box. The simulated spectrum is convolved with the appropriate line spread function of HST-COS and added with noise having SNR similar to that of 3C57 (see section 5.3).

5.3 Simulation

We use GADGET-2 simulation post-processed with CITE and GLASS to generate the models of Lyα forest as discussed in Chapter 2 and 4. In order to enable fair comparison with the observational data, we prepare a sample of mock spectra which has properties resembling as close as possible to the observed ones. Each redshift bin contains different number of observed spectra (see Table 5.1). Let us assume that there are $N_{\rm spec}$ observed spectra at the redshift of interest z. For given thermal history and a free parameter $\Gamma_{\rm HI}$, we first generate $N_{\rm spec}$ simulated spectra using the method described above, which we call a "mock sample". We repeat the procedure by choosing different random sightlines and generate N such mock samples. We take $N = 500$ in this work. The collection of N mock samples constitute a "mock suite". Thus at z, the mock suite consists of $N \times N_{\rm spec}$ simulated spectra. The velocity separation of pixels in the simulated spectra is ~ 5 km s^{-1} which is set by the resolution of the box, whereas the velocity resolution of observations is ~ 17 km s^{-1}. We therefore resample the simulated spectra (by linear interpolation) to match the observed data and then convolve with the line spread function (LSF) given for HST-COS spectra. HST-COS LSF[1] is given at various wavelength line centers (λ_c) e.g. from 1150 Å to 1750Å in steps of 50Å. For our purpose we assume that the broadening function is not changing over the range $\lambda_c \pm 25$Å. Finally, we add random noise to each spectrum in accordance with the SNR of the observed data, e.g., a mock sample of $N_{\rm spec}$ spectra corresponds to $N_{\rm spec}$ different values of SNR as in the observed spectra. We found that SNR varies across the spectrum in the observed data. For each spectrum we calculate the SNR in 5 different regions and choose the median SNR. We use this observed median SNR in simulated spectra to mimic observations. For comparison, we show a simulated absorption spectrum along a random line of sight through our simulation box in the bottom panel of Fig. 5.2. One can see that the simulated spectrum is qualitatively quite similar to the observed one (the one with removal of all other lines except Lyα) shown in the middle panel.

---

[1] http://www.stsci.edu/hst/cos/performance/spectral _resolution/

5.4 Lyα forest: Flux statistics and Voigt statistics

The basic idea behind constraining $\Gamma_{\rm HI}$ is to calculate the χ^2 between the observed statistics of the Lyα forest and the same statistics calculated by modeling the Lyα forest in a cosmological simulation. The $\Gamma_{\rm HI}$ corresponding to minimum value of χ^2 (i.e. $\chi^2_{\rm min}$) gives the best fit $\Gamma_{\rm HI}$ whereas the associated statistical error is obtained by calculating the parameter values corresponding to $\chi^2_{\rm min} \pm 1$ (Press et al. 1992). In this section, we discuss the Lyα forest statistics sensitive to the $\Gamma_{\rm HI}$ and that are relatively easy to obtain from observations and simulations. These statistics are broadly divided into two cases. In the first case, Lyα transmitted flux is treated as a continuous field quantity and are called as Flux statistics. In the second case, Lyα forest is decomposed into multiple Voigt profiles and statistics are derived from the fitted parameter (Voigt statistics).

5.4.1 Flux statistics

In order to carry out comparison between simulations and observed data at each redshift bin given in Table 5.1, we consider two statistics of the Lyα transmitted flux, namely the flux PDF and the flux PS. In the following subsections we describe the method of calculating the flux PDF and PS and appropriate covariance matrix from the simulation.

5.4.1.1 Flux PDF

We compute the flux PDF of the observed and simulated spectra for all four redshift bins given in Table 5.1. We evaluate the distribution using 11 flux bins of width $\Delta F = 0.1$ with the first bin center at $F = 0$ and last one at $F = 1$. The pixels with $F > 1$ (respectively, $F < 0$) are included in the last (respectively, first) bin. Note that the flux bin widths used in this work are larger than those used previously at high-z. This is mainly because the SNR in the present sample is much lower than what is typically achieved in high-z echelle spectra (Bolton et al. 2008; Desjacques et al. 2007; Jenkins and Ostriker 1991; Kim et al. 2007;

McDonald et al. 2000; Rollinde et al. 2013). This will influence flux PDF in the bins near continuum and possibly introduce a strong correlation between different bins if bin width is small.

Any meaningful statistical comparison requires errors on the observed flux PDF at each flux bin, along with the noise covariance between different bins. One standard way of estimating the error covariance matrix from the observed data is to use the jack-knife method. However, we found that the errors obtained using this method are considerably underestimated. This is possibly because the cosmic variance is not properly accounted for in the jack-knife method (Rollinde et al. 2013). Hence, we use the simulated mock samples to compute the covariance matrix as explained below:

As discussed earlier, for each redshift bin (see Table 5.1) we generate $N = 500$ simulated mock samples for the free parameter Γ_{12} and for each model given in Table 4.2. We remind the reader that each mock sample consists of number of sightlines equal to observed number of sightlines in the corresponding redshift bin. Let $P_n(F_i)$ denote the value of the flux PDF in the i^{th} bin of n^{th} mock sample, where n takes values from 1 to N. Let the flux PDF in the i^{th} bin averaged over all mock samples be denoted as $\overline{P}_i$. The covariance matrix element $C(i,j)$ between the i^{th} and j^{th} bins is given by,

$$C(i,j) = \frac{1}{N-1} \sum_{n=1}^{N} \left[P_n(F_i) - \overline{P}_i\right] \left[P_n(F_j) - \overline{P}_j\right] \tag{5.1}$$

where, i and j can take values from 1 to the number of bins (which in this case is 11). The covariance matrix C is calculated for free parameter Γ_{12} and for each initial $T_0 - \gamma$ (at $z_1 = 2.1$) model given in Table 4.2.

To visualize the covariance matrix we calculate the correlation matrix defined as

$$Corr(i,j) = \frac{C(i,j)}{\sqrt{C(i,i)\,C(j,j)}}. \tag{5.2}$$

The left-hand panel of Fig. 5.3 shows the correlation matrix for the flux PDF for simulated Lyα forest at $z = 0.3$ for a model $T15 - \gamma1.3$ (refer Table 4.2) and $\Gamma_{12} = 0.1$. It is clear from the figure that the off-diagonal terms of the matrix are not negligible, thus showing that the errors in different bins are correlated.

This implies that the full covariance matrix should be used to compute the χ^2 while comparing the simulation with the data. We find that the correlation is strongest for the immediately neighboring bins. Also, the correlation between the neighboring bins increases for higher flux bins since large number of pixels are in the continuum.

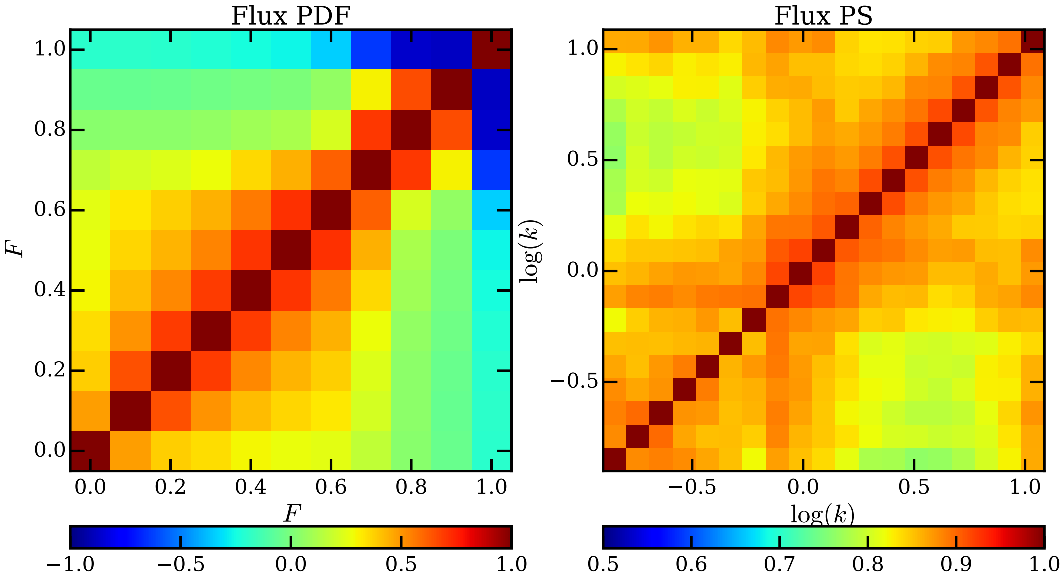

Figure 5.3: Correlation matrix for the flux PDF (*left-hand panel*) and the flux PS (*right-hand panel*). Both the correlation matrices are calculated using the covariance obtained from the simulated mock samples (see section 5.4.1.1 and section 5.4.1.2 for more details). The correlation matrices are shown for simulated Lyα forest at $z = 0.3$, with $\Gamma_{12} = 0.1$ for $T15 - \gamma1.3$ model given in Table 4.2.

In the observed spectra a large number of pixels are found to be in the continuum. Consequently, the errorbar on the flux PDF in the flux bins close to continuum $F \geq 0.9$ are very small. Any χ^2 minimization procedure thus tries to give more weight to the flux PDF bins around $F \geq 0.9$. In addition, these bins near continuum are affected by noise (where, SNR varies from 5 to 15) and continuum fitting uncertainty. This can introduce an additional correlation between bins near the continuum. To avoid such difficulties, during χ^2 minimization we used the flux PDF in the range $0 \leq F \leq 0.8$. However to normalize flux PDF we used all the bins. Note that Rollinde et al. (2013) have used a similar cutoff to calculate the flux PDF at $z \sim 2-3$. In our case it is not only the continuum uncertainty but also the relatively poorer SNR of the observed spectra, which affects the flux PDF calculation for bins with $F \geq 0.9$, are important. We have checked and found that ignoring the points near the continuum does not affect our constraints on Γ_{12}, except for a marginal increase in the errorbars on Γ_{12}.

The χ^2, which will be used for quantifying the match between the observed flux PDF and with the simulated one, can be written in the matrix form as,

$$\chi^2_{(T_0,\gamma,\Gamma_{12})} = [P_{(T_0,\gamma,\Gamma_{12})} - P_{obs}]\, C^{-1}\, [P_{(T_0,\gamma,\Gamma_{12})} - P_{obs}]^T, \tag{5.3}$$

where P_{obs} denotes observed flux PDF obtained from all the spectra in the relevant redshift bin and $P_{(T_0,\gamma,\Gamma_{12})}$ is the flux PDF obtained from our simulations for a particular model given in Table 4.2 and a free parameter Γ_{12}. Note that both P_{obs} and $P_{(T_0,\gamma,\Gamma_{12})}$ are row vectors with their i^{th} element being the flux PDF in the i^{th} bin. We re-emphasize that the covariance matrix C is calculated in each redshift bin for each model in Table 4.2 and Γ_{12}. Note that, because the flux PDF is normalized, the covariance matrix is singular. Hence we use the singular value decomposition method (Press et al. 1992) to compute the χ^2. The $\chi^2_{(T_0,\gamma,\Gamma_{12})}$ can be calculated for each combination of the free parameters $T_0 - \gamma$ at an initial redshift and Γ_{12} at the redshift of our interest. The best-fit parameters are obtained by finding the location of the minimum of the $\chi^2_{(T_0,\gamma,\Gamma_{12})}$. The 1σ confidence level corresponds to the region between $\chi^2_{1\sigma} = \chi^2_{\rm min} \pm \Delta\chi^2_{1\sigma}$, where $\chi^2_{\rm min}$ is the minimum value of the χ^2 and $\Delta\chi^2_{1\sigma} = 1$ (Press et al. 1992).

5.4.1.2 Flux PS

We compute the flux PS from observational data and simulations in redshift bins same as those used for estimating the flux PDF. However, there is a crucial difference in how we treat the sightlines while calculating the flux PS from that for the flux PDF. In the case of the flux PDF, we spliced different sightlines from the simulation box to construct a redshift path length as large as the observed redshift range. However, such splicing may introduce spurious effects while computing the two point correlation properties of the flux. Hence for calculating the flux PS we use sightlines of comoving length equal to the simulation box size $50\, h^{-1}$ cMpc.

In order to ensure that the simulations and the observations are treated on equal footing, we divide the observed spectra within each z bin into segments which have comoving length equivalent to $50h^{-1}$ cMpc. To calculate the flux PS, we first compute the Fourier transform $F(k)$ of the transmitted flux. The corresponding power is given by $P(k) \propto |F(k)|^2$, where the normalisation is calculated from the condition

$$\sigma_F^2 = \int\limits_{-\infty}^{\infty} \frac{dk}{2\pi}\, P(k) \tag{5.4}$$

with σ_F^2 being the variance of the transmitted flux. Once we compute $P(k)$ for each segment (of comoving length $50h^{-1}$ cMpc) of the spectra, we take an average over all the segments to get an estimate for the flux PS.

A reliable estimate of the flux PS can only be obtained in a limited range of scales because of the finite length of the spectra and other systematic effects. For the observed spectra, the small scale power is affected by the presence of the narrow metal lines (Arinyo-i-Prats et al. 2015). Following McDonald et al. (2000), we choose to work with scales corresponding to $k < k_c \sim 8\ h$ cMpc^{-1}. Similarly, the large scale power in the observed spectra is affected by the uncertainties in the continuum fitting (Kim et al. 2004). In addition, while computing the flux PS from our simulations, the finite periodic box size of $50\,h^{-1}$ cMpc implies that scales with $k < k_t \sim 0.2h$ cMpc^{-1} may not be sampled properly. Hence, to make any meaningful comparison between the observations and simulations we restrict the flux PS measurements to scales $0.209 \leq k/(h\ \mathrm{cMpc}^{-1}) \leq 8$.

Care must be exercised while binning the $P(k)$ over different k-ranges. We find that for $k \gtrsim 1h$ cMpc^{-1}, the k modes are sampled densely enough to divide the $P(k)$ into bins. In that case we use logarithmic bins of size $\log_{10}(1.24)$ similar to Kim et al. (2004). On the other hand, for $k \lesssim 1h$ cMpc^{-1}, the k modes are sparsely sampled and hence we do not bin the data as done in Arinyo-i-Prats et al. (2015). We ensure that the same procedure is followed while dealing with the observed and simulated spectra.

The procedure for computing the errorbars on the observed flux PS, i.e., estimating the error covariance matrix C, is same as the one discussed for the flux PDF. The matrix C is estimated from the suite of mock spectra. We also attempted to estimate it from the observed spectra using the jack-knife method.

However, since the number of observed spectra is relatively small, the off-diagonal terms in the covarinace matrix are found to be noisy. The right-hand panel of Fig. 5.3 shows the flux PS correlation matrix. The flux PS correlation matrix is dominated by diagonal terms consistent with McDonald et al. (2000) and Zhan et al. (2005). Since the neighboring k-modes are likely to be correlated, the neighboring bins show strong correlation. As one moves away from the diagonal terms the correlations between the different mode decrease. The smaller scales show slightly stronger correlations because the underlying density field is more non-linear and non-gaussian at these scales (Zhan et al. 2005).

We follow the same procedure as discussed in the previous section for calculating χ^2 between model and observed flux PS.

5.4.1.3 Tests with the mock spectra

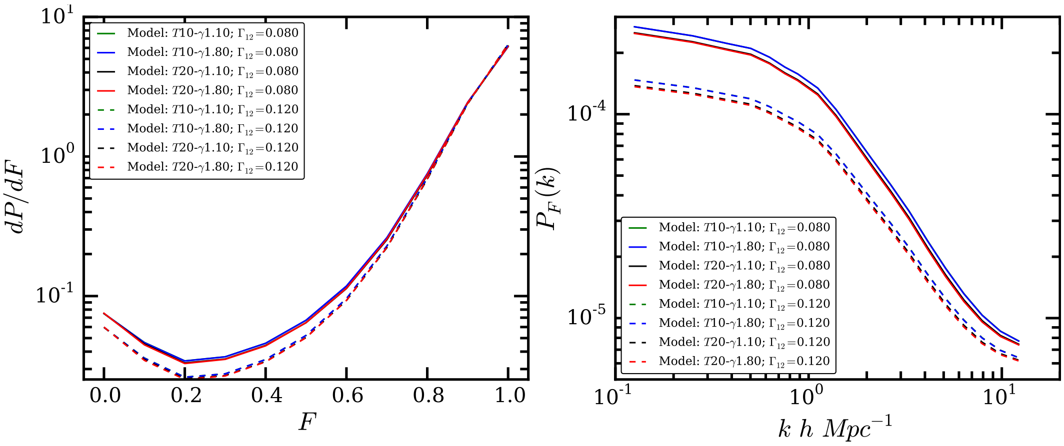

Figure 5.4: *Left* and *right-hand panels* show respectively the variation of the flux PDF and PS (at $z = 0.3$) for different models given in Table 4.2 and $\Gamma_{12} = 0.08$ and 0.12 at $z = 0.3$. The values of Γ_{12} and model corresponding to different lines are indicated in the legend. It is clear from the figure that flux PDF and PS are more sensitive to Γ_{12} (at $z = 0.3$) than initial values of T_0 and γ (at $z_1 = 2.1$) because the final equation of states at $z = 0.3$ are very similar (see Table 4.2).

Before using the observed and simulated spectra, to constrain the photoionization rate Γ_{12} at $z < 0.45$, we carry out a few tests on the simulated quantities. First, we check the sensitivity of the flux PDF and flux PS on the parameters

$T_0 - \gamma$ (at initial redshift $z_1 = 2.1$) and Γ_{12}. Fig. 5.4 shows the dependence of the flux PDF (left-hand panel) and flux PS (right-hand panel) at $z = 0.3$ on these three parameters. Note that the values of T_0 and γ at the redshift of interest are obtained by varying these two parameters at the initial redshift $z_1 = 2.1$. The solid and dashed lines, for a given combination of T_0 and γ, in the Fig. 5.4 correspond to $\Gamma_{12} = 0.08$ and 0.12 respectively at $z \sim 0.3$.

Even though we varied T_0 at $z_1 = 2.1$ by factor of 2, the value of T_0 obtained at $z \sim 0.3$ using CITE differ by only ~ 9 per cent. Similarly, for given T_0 at $z_1 = 2.1$ even when we change initial γ between 1.1 and 1.8 values of T_0 and γ obtained at $z \sim 0.3$ are nearly identical. Thus the flux PDF and PS are fairly insensitive to our choice of $T_0 - \gamma$ at $z_1 = 2.1$. In other words the flux PDF is insensitive to the He II reionization history. It is also clear from Fig. 5.4 that both the above statistics are sensitive to assumed value Γ_{12}. Therefore they can be used to constrain Γ_{12}.

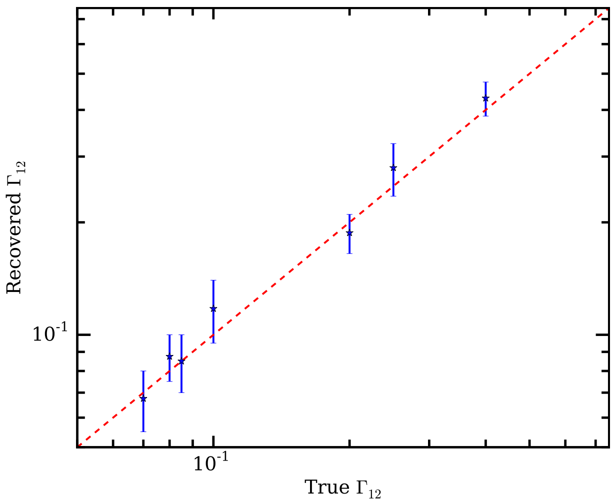

Figure 5.5: Recovery of the Γ_{12} at $z = 0.3$ using the flux PDF and PS and χ^2 statistics. The x-axis represents the true Γ_{12}, i.e., the one used in the input model. The points with errorbars show the recovered Γ_{12} with the 1σ confidence interval for each input model. The red dashed line indicates the case where there is perfect match between the input and the recovered Γ_{12}. The input and mock data are drawn from two different simulations with same cosmological parameters but different initial condition. The typical uncertainty in recovered Γ_{12} is ± 0.015.

We next study how well the two statistics can be used for constraining the Γ_{12}. For a particular model given in Table 4.2 and a fixed value of Γ_{12} (at $z = 0.3$), we construct a mock sample from the simulations, i.e., a set of sightlines which have properties similar to the observed ones. This mock sample can be treated as the "input" data for which the two statistics (flux PDF and PS) can be calculated. We then draw sightlines from other simulation box (parameters T_0 and γ are different from that of the input data) to construct a large number ($N = 500$) of mock samples and compute the two statistics along with the error covariance matrix. The input data and mock samples are drawn from two different simulation boxes with identical cosmological parameters but different initial conditions. The idea is to vary the value of Γ_{12} for these samples and compare with the input data. The minimization of the χ^2 should enable us to obtain the best-fit value of Γ_{12} along with the errorbars, which can be compared with the input value of Γ_{12}. The result of the analysis for seven different input Γ_{12} values for $z \sim 0.3$ is shown in Fig. 5.5. The red dashed line indicates the case where there is perfect match between the input and the recovered Γ_{12}. The point with errorbars show the recovered Γ_{12} with the 1σ confidence interval for each input model. We can see that our analysis recovers the input value of the photoionization rate quite accurately. The typical statistical uncertainty in recovering Γ_{12} is ~ 0.015.

Finally, we test the effect of using the HST-COS LSF instead of the traditionally used Gaussian profile function. Fig. 5.6 shows the flux PDF (*left-hand panel*) and the flux PS (*right-hand panel*) obtained using the two LSFs. The Gaussian LSF used for making this plot has an FWHM = 17 km s^{-1} as shown by solid blue curve. The HST-COS LSF is asymmetric and has extended wings that do not go to zero as rapidly as Gaussian LSF. Hence, the number of pixels near zero are less in HST-COS LSF as compared to Gaussian LSF (*left-hand panel*). The LSF also affects the variance of flux field (σ_F^2) and hence the normalization of flux PS (*right-hand panel*). We found that one would overpredict the Γ_{12} by ~ 20 per cent if the Gaussian LSF is used in the simulated spectra instead of HST-COS LSF using flux PS.

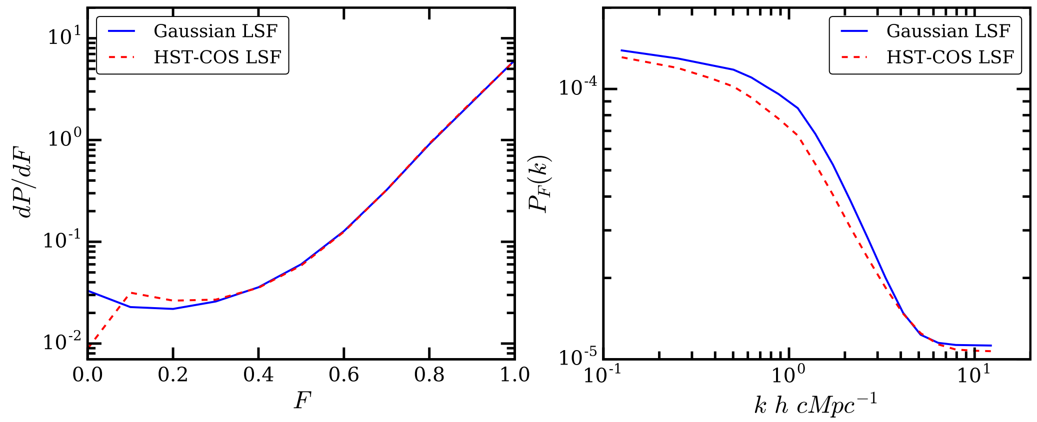

Figure 5.6: *Left-hand* and *right-hand panels* show the effect of LSF on flux PDF and PS respectively. In both panels results obtained with Gaussian LSF (FWHM ~ 17 km s^{-1}) are shown using solid blue curves and ones that are obtained using HST-COS LSF are shown by red dashed curves. In left-hand panel number of saturated pixels (i.e. $F \sim 0$) are smaller when we use the HST-COS LSF. Right-hand panel shows that the LSF affects the overall normalization (σ_F^2) of the flux PS below $k \sim 6\, h\, \mathrm{cMpc}^{-1}$.

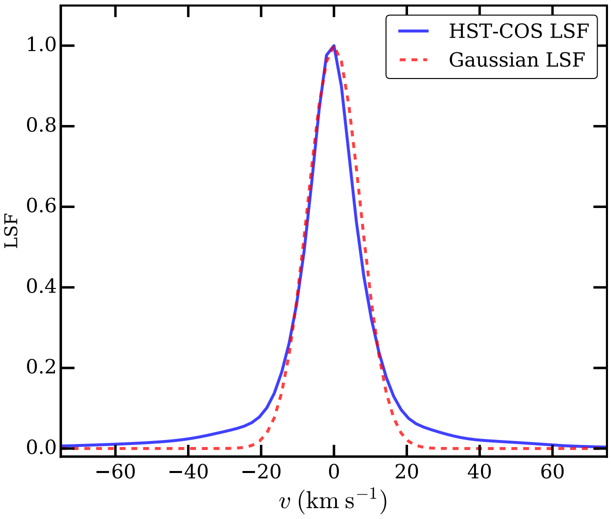

Figure 5.7: Figure shows comparison of HST-COS LSF (blue solid curve) with traditionally used Gaussian function (red dashed curve, FWHM $\sim$ 17 km s^{-1}). The HST-COS LSF is slightly asymmetric and has extended wings that do not go to zero (at $|v| > 21$ km s^{-1}) as rapidly as Gaussian LSF.

5.4.2 Voigt statistics

In this section we decomposed Lyα forest spectra generated from GADGET-2 + CITE simulations into multiple Voigt profiles using VIPER. We formed a line catalog for each mock sample and obtained three distributions (i) b parameter distribution, (ii) b vs $\log \mathrm{N_{HI}}$ distribution and (iii) CDDF. These distributions calculated from different mock samples are used to estimate the errors. We compare these distributions from simulations with those from observation. In particular, we used CDDF to constrain the Γ_{12} and its evolution in four different redshift bins.

5.4.2.1 b parameter distribution

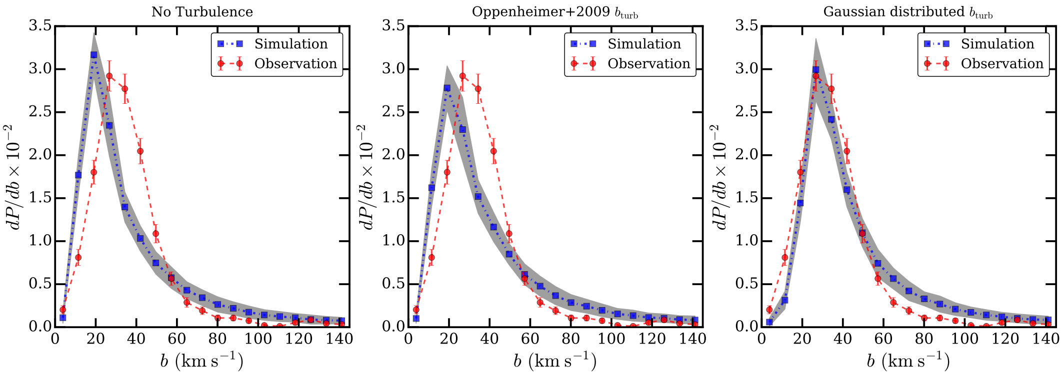

Figure 5.8: Comparison of b parameter distribution (at $0.075 \leq z \leq 0.45$) from observations (red dashed line with errorbar) and simulations (blue dotted line with 1σ shaded region) for 3 cases (see section 5.4.2.1): (i) when $b_{\rm turb}$ is not added in the simulation (*left-hand* panel), (ii) when density dependent $b_{\rm turb}$ at given $n_{\rm H}$ (Oppenheimer and Davé 2009) is added in the simulation (*middle panel*) and (iii) when Gaussian distributed $b_{\rm turb}$ is added in the simulation (*right-hand* panel). The errorbars on model b parameter distribution are calculated from mock sample whereas the errorbars on observed b parameter distribution are calculated assuming Poisson statistics. The b parameter distribution from models with Gaussian distributed $b_{\rm turb}$ qualitatively matches well with that from the observation.

The b parameter distribution calculated from Voigt profile fitting is sensitive to thermal history, the energy injected by various astrophysical processes in the form of heat and unknown turbulent motions in the IGM (Davé et al. 2001; McDonald et al. 2001; Schaye et al. 1999, 2000). Recently it has been found that the b parameters obtained from various hydrodynamical simulations are typically smaller compared to those from the observations at low-z (Viel et al. 2017) though the thermal history parameters self-consistently obtained in these simulations agree well with each other. The b parameters from our simulation (GADGET-2 + CITE) are also found to be significantly smaller than the observed b parameters (see left-hand panel of Fig. 5.8). Thus to match the b parameter distribution, additional thermally and/or non-thermally broadened b parameter is required.

We found that the observed b parameter distribution can be matched with simulation by increasing the temperature of each pixel along sightline by a factor of 3. This increase in temperature would correspond to injection of the energy in the form of heat into the IGM. Increasing the gas temperature can lead to two effects (i) reducing the recombination rate coefficient as it scales as $\sim T^{-0.7}$ and (ii) introduce additional ionization due to collisions in the high density gas. By artificially enhancing the heating rate by more than a factor of 3, Viel et al. (2017) have recently shown the simulated b distribution can be made consistent with the observed ones. However such a model tends to suggest lower $\Gamma_{\rm HI}$ compared to simulations without additional heating (see Table 1 in Viel et al. 2017). As we will show in section 5.6, a lower value of $\Gamma_{\rm HI}$ would imply that the number of ionizing photons in the IGM is less than that expected from only QSOs. Hence we rather focus on a scenario where the additional contribution to the line broadening arises from non-thermal motions. Here we explore this possibility by introducing turbulent motions that can simultaneously explain b parameter distributions and b vs $\log \rm N_{HI}$ distributions as well.

In this work we incorporate additional broadening by adding a *non-thermal* (micro-turbulence) component $b_{\rm turb}$ to thermal b parameter in quadrature ($b^2 = b^2_{\rm thermal} + b^2_{\rm turb}$) to mimic micro-turbulence that is missing in our simulation and see its effect on the derived Γ_{12} constraints. We refer to the non-thermal contribution to the b parameters as 'micro-turbulence', $b_{\rm turb}$. In general, this 'micro-turbulence' is due to any physical phenomenon affecting the width of the absorption line that is not captured properly in our simulations e.g. various feedback processes and/or numerical effects[1]. Using this new b parameter we compute the Lyα optical depth and fit each absorption line using VIPER to get b and $\log \mathrm{N_{HI}}$. The Γ_{12} is then constrained from the model with and without micro-turbulence. We used two different models to quantify the micro-turbulence in simulation as explained below.

1. Density dependent $b_{\rm turb}$: In order to match the observed line width distribution of O VI absorbers, Oppenheimer and Davé (2009) added density dependent turbulence in their simulation. Following Eq. 5 and 6 given in Oppenheimer and Davé (2009), we added (in quadrature) the $b_{\rm turb}(\mathrm{n_H})$ in simulated spectra where $\mathrm{n_H}$ is hydrogen number density. The form of these equations is such that the contribution of the $b_{\rm turb}$ is appreciable only at high column densities i.e. $\log \mathrm{N_{HI}} > 13.5$ (see middle panel in Fig. 5.9). The b parameter distribution for this case is shown in the middle panel of Fig. 5.8. It is clear that the simulated b values are smaller than the observed ones even in this case.

2. Gaussian random $b_{\rm turb}$: In this approach we generated $b_{\rm turb}$ from a Gaussian random variable with mean $\mu = 20\,\mathrm{km\,s^{-1}}$and standard deviation $\sigma = 10\,\mathrm{km\,s^{-1}}$at each grid point along a sightline in the simulation box[2]. These values are in agreement with the distribution of non-thermal broadening parameters (see Fig. 24 in Muzahid et al. 2012; Tripp et al. 2008) derived for the well-aligned O VI and H I absorbers. From Fig. 5.8 (right-hand panel), we see that the agreement between the observed and model b parameter distribution is better in the case of Gaussian distributed $b_{\rm turb}$ model than the other two models i.e., without any additional $b_{\rm turb}$ and density dependent $b_{\rm turb}$ models. The model results shown in Fig. 5.8 were based on simulations that use the best fitted redshift evolution of Γ_{12} as given in Paper-I. However, we also found that the b parameter distribution depends weakly on the assumed evolution of $\Gamma_{\rm HI}$.

---

[1] The physical phenomena occurring on scales below the resolution scale (i.e., below $\sim 50h^{-1}$ ckpc) of the simulation box may affect the scales that are resolved (Springel and Hernquist 2002). These physical phenomena may not be captured properly in the simulation box.

[2] We calculated b vs $\log \mathrm{N_{HI}}$ for 5 different $(\mu, \sigma) \equiv$ (10,10),(20,5),(20,10),(20,15),(30,10) combinations. The χ^2 is found to be minimum for $\mu = 20\,\mathrm{km\,s^{-1}}$and $\sigma = 10\,\mathrm{km\,s^{-1}}$.

5.4.2.2 b vs $\log \mathrm{N_{HI}}$ distribution

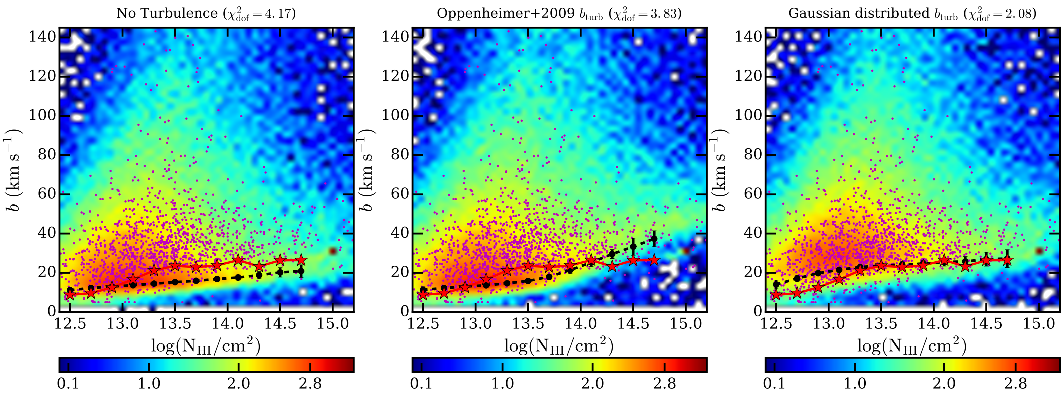

Figure 5.9: Comparison of b vs $\mathrm{N_{HI}}$ distribution (at $0.075 \leq z \leq 0.45$) from observation (magenta points) and simulation (color-coded diagram) for 3 cases (see section 5.4.2.1) (i) when $b_{\rm turb}$ is not added in the simulation (*left-hand* panel), (ii) when density dependent $b_{\rm turb}$ at given $n_{\rm H}$ (Oppenheimer and Davé 2009) is added in the simulation (*middle panel*) and (iii) when Gaussian distributed $b_{\rm turb}$ is added in the simulation (*right-hand* panel). The color scheme indicates density of points from the simulation in logarithmic units. The red solid line and black dashed line shows the lower envelope for observed and model data points in both panels. The lower envelope is obtained by calculating 10th percentile of b in $\log \mathrm{N_{HI}}$ bins. The lower envelope matches well in the case where Gaussian distributed $b_{\rm turb}$ is added. We calculated the χ^2 between model and observation by binning the data into 2D bins. These values are quoted on top of each panel. The $\chi^2_{\rm dof}$ is better for a model with Gaussian distributed $b_{\rm turb}$ ($\chi^2_{\rm dof} = 2.08$) than a model without turbulence ($\chi^2_{\rm dof} = 4.17$) and a model with density dependent $b_{\rm turb}$ ($\chi^2_{\rm dof} = 3.83$).

In this section we discuss the effect of adding $b_{\rm turb}$ on the b vs $\log \mathrm{N_{HI}}$ 2D distribution[1]. Fig. 5.9 shows comparison of b vs $\log \mathrm{N_{HI}}$ distribution ($0.075 \leq z \leq 0.45$) from observations (shown by magenta points) with that from 3 different models (i) without turbulence (left-hand panel), (ii) density dependent $b_{\rm turb}$ as suggested by Oppenheimer and Davé (2009, middle panel) and (iii) Gaussian distributed $b_{\rm turb}$ (right-hand panel). The color scheme indicates the density of

---

[1] We refer reader to Fernández-Soto et al. (1996); Webb and Carswell (1991) for discussions on Voigt profile fitting procedure influencing this correlation.

points in logarithmic units. To assess the goodness-of-fit, we calculated the reduced χ^2 for 2D distribution by binning the data along b axis in 21 bins (with bin width 7.5 $\rm km\,s^{-1}$) and along $\log \rm N_{HI}$ axis in 13 bins (with bin width 0.2). Let $P_{\rm sim,k}(i,j)$ be the value of 2D distribution of $k^{\rm th}$ mock sample in $i^{\rm th}$, $j^{\rm th}$ bin along $\log \rm N_{HI}$, b axis respectively. The mean and standard deviation of 2D distribution from mock samples can be calculated as,

$$\begin{aligned} \overline{P}_{\rm sim}(i,j) &= \frac{1}{N}\sum_{k=1}^{N} P_{{\rm sim},k}(i,j) \\ \sigma^2_{\rm sim}(i,j) &= \frac{1}{N-1}\sum_{k=1}^{N}[P_{{\rm sim},k}(i,j) - \overline{P}_{\rm sim}(i,j)]^2 \end{aligned} \tag{5.5}$$

where $N = 100$ is the number of mock samples. Let $P_{\rm obs}(i,j)$ be the value of observed 2D distribution in $i^{\rm th}$, $j^{\rm th}$ bin along $\log \rm N_{HI}$, b axis respectively. Note that both distributions i.e., $P_{\rm obs}(i,j)$ and $P_{\rm sim,k}(i,j)$ are normalized. The reduced χ^2 between the observed and model distribution is,

$$\chi^2_{\rm dof} = \frac{1}{N_{\rm y}\times N_{\rm x} - 1}\sum_{i=1}^{N_{\rm x}}\sum_{j=1}^{N_{\rm y}} \frac{[\overline{P}_{\rm sim}(i,j) - P_{\rm obs}(i,j)]^2}{\sigma^2_{\rm sim}(i,j)} \tag{5.6}$$

where $N_{\rm x} = 13$ and $N_{\rm y} = 21$ is number of bins along $\log \rm N_{HI}$ and b axis respectively. From Fig. 5.9 the $\chi^2_{\rm dof}$ for Gaussian distributed $b_{\rm turb}$ is better (~ 2.1) than the model without turbulence ($\chi^2_{\rm dof} \sim 4.2$) and model with density dependent $b_{\rm turb}$ ($\chi^2_{\rm dof} \sim 3.8$).

Another way to asses the goodness of the assumed form for $b_{\rm turb}$ is to match the lower-envelope in b vs $\log \rm N_{HI}$ distribution. At $z > 2$, the lower envelope is strongly correlated with thermal history parameters and has been used in the past to measure the effective equation of state of the IGM at high-z (Schaye et al. 1999, 2000). The red stars with solid line and black circles with dashed line in Fig. 5.9 shows the lower envelope for observed and model b vs $\log \rm N_{HI}$ distribution respectively. The lower envelope is obtained by calculating the $10^{\rm th}$ percentile of b values in each $\log \rm N_{HI}$ bin (Garzilli et al. 2015). In model b vs $\log \rm N_{HI}$ distribution, the lower envelope is calculated for all mock samples. The black circles with errorbars in Fig. 5.9 represents mean and standard deviation of lower envelope from mock samples. In the case of models without turbulence we see that the lower envelope obtained for the observed data is systematically higher than that from simulation for $\log \mathrm{N_{HI}} \geq 13.2$. It is also clear from the middle panel that density dependent turbulence overproduce b at $\log \mathrm{N_{HI}} > 14.2$ and under produce in the range $13 \leq \log \mathrm{N_{HI}} \leq 14$. It is evident from Fig. 5.9 that the lower envelope in Gaussian distributed b_{turb} model (right-hand panel) matches well with the observed lower envelope at $\log \mathrm{N_{HI}} \geq 13.2$[1]. At low column densities i.e., $\log \mathrm{N_{HI}} < 13.2$ the observed b parameters tend to be smaller than what is predicted in this case. This suggests that the actual b_{turb} could be smaller at $\log \mathrm{N_{HI}} < 13.2$ compared to mean value we have assumed. Note at these low $\log \mathrm{N_{HI}}$ values Lyα absorption are in the linear part of curve of growth and $\mathrm{N_{HI}}$ measurements are independent of b parameter.

In summary the Gaussian distributed b_{turb} model matches well with observation for b vs $\log \mathrm{N_{HI}}$ distribution and b parameter distribution. We reemphasize that this may not be the unique explanation for the additional broadening required in the simulation even-though it consistently reproduces the b parameter distribution and b vs $\log \mathrm{N_{HI}}$ scatter. In the next section, we calculate CDDF and constrain Γ_{12} from observation by comparing model with Gaussian distributed b_{turb} and model without any b_{turb}.

5.4.2.3 Column density distribution function (CDDF)

In this section we match model CDDF with the observed CDDF to constrain Γ_{12} in 4 redshift bins identified above. The CDDF is calculated for each mock sample. Γ_{12} is a free parameter in our analysis and the model CDDF depends on its value. The model CDDF is binned in a way identical to that of observed CDDF. Let $f_{\mathrm{sim},i,k}(\Gamma_{12})$ be value of the CDDF in k^{th} bin of i^{th} mock sample for a given Γ_{12}. The mean and variance of CDDF in k^{th} bin is calculated as follows.

$$
\begin{aligned}
\overline{f}_{\mathrm{sim},k}(\Gamma_{12}) &= \frac{1}{N}\sum_{i=1}^{N} f_{\mathrm{sim},i,k}(\Gamma_{12}) \\
\sigma^2_{\mathrm{sim},k} &= \frac{1}{N-1}\sum_{i=1}^{N}[f_{\mathrm{sim},i,k}(\Gamma_{12}) - \overline{f}_{\mathrm{sim},k}(\Gamma_{12})]^2
\end{aligned}
\tag{5.7}
$$

---

[1] The matching between observation and model with Gaussian distributed b_{turb} is also good when lower envelope is calculated using 5^{th} and 20^{th} percentile.

where $N = 100$ is number of mock samples.
The reduced χ^2 between observed CDDF and model CDDF is given as,

$$\chi^2_{\rm dof}(\Gamma_{12}) = \frac{1}{N_{bin} - 1} \sum_{k=1}^{N_{bin}} \frac{[\overline{f}_{{\rm sim},k}(\Gamma_{12}) - f_{{\rm obs},k}]^2}{\sigma^2_{{\rm sim},k}} \tag{5.8}$$

where $N_{bin} = 13$ is number of bins in CDDF and $f_{{\rm obs},k}$ is observed CDDF in $k^{\rm th}$ bin.

5.5 Constraints on Γ_{12}

In this section we constrain the Γ_{12} from two flux statistics namely flux PDF, flux PS and Voigt statistics CDDF.

5.5.1 Flux statistics: Flux PDF and PS

We obtain constraints on Γ_{12} by comparing the flux PDF and PS from the simulated Lyα forest with those from the HST-COS data using the χ^2-minimization technique discussed in section 3.3. Fig. 5.10 shows the results of our analysis for the four redshift bins identified in Fig. 5.1 and given in Table 5.1 i.e., for $z = 0.1125, 0.2, 0.3$ and 0.4. T_0 and γ at these redshifts are obtained for different model (see Table 4.2) by using CITE. For a particular model there is one free parameter Γ_{12} at each z.

The left-hand panels in the Fig. 5.10 (all results are shown for model $T20-\gamma1.8$) show the comparison of observed flux PDF (red dashed curve) with that of the best fit model (blue dotted curve). The blue shaded regions indicate the 1σ dispersion (i.e., corresponding to diagonal term of the covariance matrix in Eq. 5.1) in model flux PDF at each value of F calculated from the mock sample. Although we plot the flux PDF in the range 0 to 1, we use only the flux bins $0 \leq F \leq 0.8$ for the χ^2 analysis (as indicated by black dashed vertical line with arrow). We find that the match between the simulated spectra and the observed ones are quite good for all the four redshifts (typical χ^2 per degree of freedom i.e., $\chi^2_{\rm dof} \sim 1$). The only bin where the two do not agree is the bin with $F = 0$ where the observed spectra systematically predict less number of pixels. This mismatch could be due to uncertainty in background subtraction in HST-COS data.

The middle panels show the comparison of observed flux PS (red dashed curve) with best fit model (blue dotted curve). The blue shaded regions indicate the 1σ range on the model flux PS. It is interesting to see that the match again is quite good, and we can match both the flux PDF and PS for the same value of Γ_{12}.

The right-hand panels show the variation of the reduced χ^2 with Γ_{12} for the flux PDF (blue dotted curve), the flux PS (red dashed curve) and the combined (i.e., flux PDF and PS) statistics (black solid curve). The best fit Γ_{12} used in the left-hand and middle panels corresponds to the one which gives minimum χ^2 for the combined case. We see that all the three curves have the expected parabolic shape and the reduced χ^2 (i.e., $\chi^2_{\rm dof}$) at the minima have values ~ 1. The minimum of the χ^2 for both the flux PDF and PS occur at similar values of Γ_{12} and agree well within 1σ. We also note that the width of the curve is smaller for the flux PDF alone case than for the flux PS alone case which implies that the flux PDF is better at constraining $\Gamma_{\rm HI}$ than the flux PS. However, flux PDF is sensitive to continuum fitting uncertainty in the HST-COS data, whereas flux PS is less sensitive to it as we will discuss later.

The errors on Γ_{12} obtained using the above method would be an underestimation since it ignores various other effects. We account for possible statistical and systematics uncertainties in Γ_{12} in Table 5.7. The first of these is the uncertainty in the thermal history, i.e., the two initial parameters T_0 and γ. Ideally one should vary the three parameters simultaneously, obtain the joint likelihood and marginalize over all other parameters except Γ_{12}. However, since we need to solve the temperature evolution for each $T_0 - \gamma$ combination, the full analysis can be quite computationally expensive. Hence we take a slightly different approach where we choose the most extreme values of T_0 and γ at $z_1 = 2.1$ compared to observations Becker et al. (2011); Schaye et al. (2000) and Boera et al. (2014) (for compilation see Puchwein et al. 2015). For each combination of T_0 and γ (given in Table 4.2), we evolve the thermal history to lower redshifts, compute the Lyα forest and obtain Γ_{12}. The constraints for different thermal histories for the four redshift bins are summarized in Tables 5.2, 5.3, 5.4 and 5.5. It is clear from the tables that the constraints on Γ_{12} are relatively insensitive to the values of T_0 and γ at $z_1 = 2.1$. This is what we have seen in Fig. 5.4. In all cases the differences

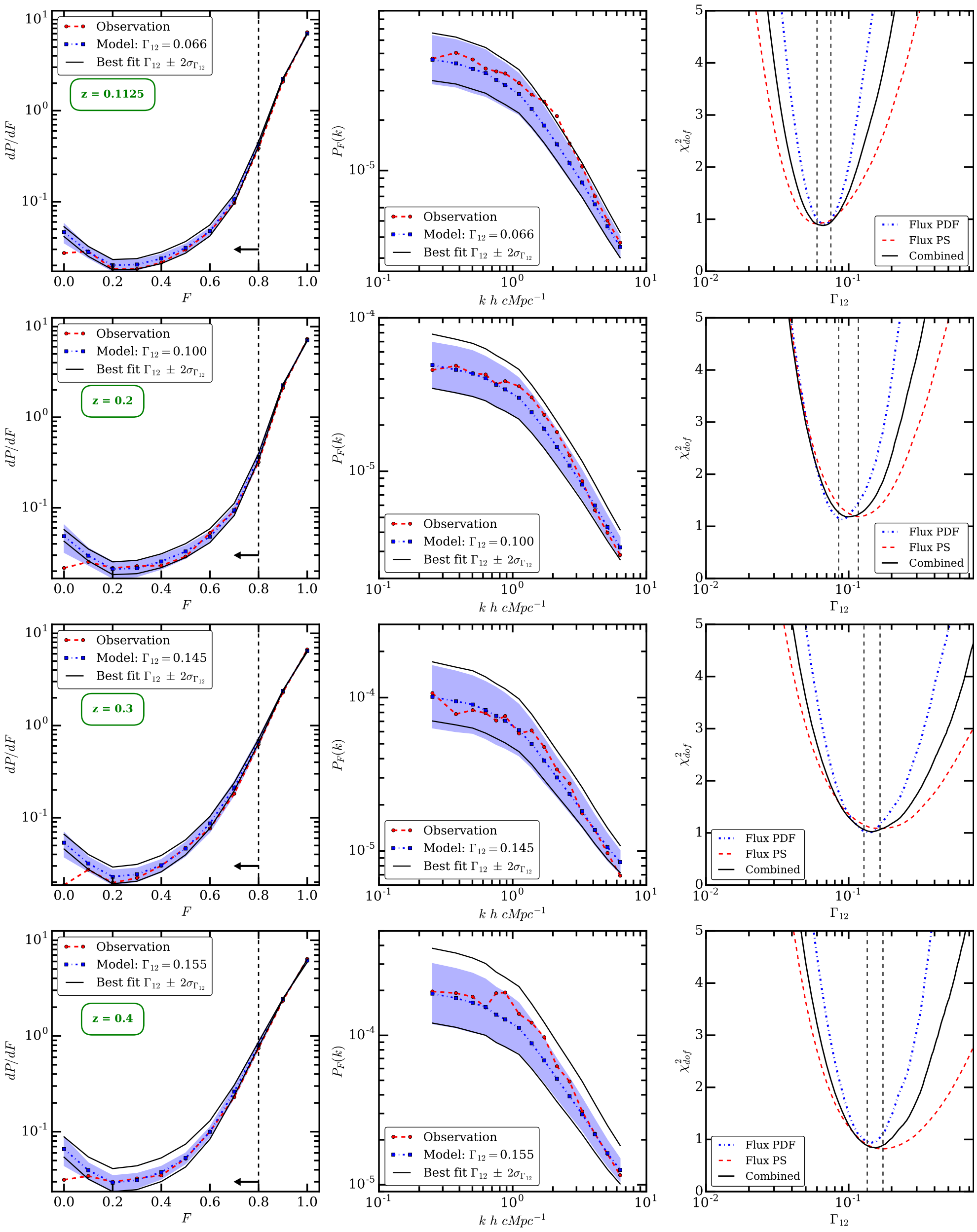

Figure 5.10: Constraints on Γ_{12} for $z = 0.1125, 0.2, 0.3, 0.4$ from top row to bottom row. The left-hand panels show the flux PDF of observed (red dashed line) and best fit model spectra (blue dotted line). The blue shaded regions show the 1σ range in flux PDF from mock samples covariance matrix (diagonal terms only). The middle panels show the flux PS of observed (red dashed line) and best fit model spectra (blue shaded region). The blue shaded regions show the 1σ range in flux PS from mock samples covariance matrix (diagonal terms only). The right-hand panels show the reduced χ^2 against Γ_{12} for flux PDF (blue dotted curve), flux PS (red dashed curve) and the combined statistics (i.e., flux PDF and PS). The black solid curve is obtained by adding the χ^2 of the flux PDF and PS. Note that the best fit model corresponds to minimum value of joint χ^2_{dof}. The vertical black dashed lines show the 1σ (statistical only) constraint on Γ_{12}. For visual purpose, the model flux PDF and PS obtained by shifting the best fit Γ_{12} by $\pm 2\sigma$ range is shown (black solid line) in the left-hand and middle panel respectively. All the plots are shown for model $T20 - \gamma1.8$ (see Table 4.2). For $z = 0.4$ redshift bin results are shown for the simulated spectra that are *not* contaminated by Lyβ forest (see text and Fig. 5.13 for details).

Table 5.2: Joint (Flux PDF + Flux PS) 1σ constraint on Γ_{12} for different thermal history (see Table 4.2) for redshift bin I (refer Table 5.1)

	$z = 0.1$		Joint constraint	
Model	T_0	γ	Γ_{12}[a]	$\chi^2_{\rm dof}$
$T10 - \gamma1.1$	4136	1.54	0.067 ± 0.008	0.83
$T10 - \gamma1.8$	4133	1.61	0.068 ± 0.007	0.83
$T20 - \gamma1.1$	4546	1.48	0.065 ± 0.006	0.81
$T20 - \gamma1.8$	4493	1.62	0.066 ± 0.007	0.83
$T15 - \gamma1.3$	4245	1.55	0.067 ± 0.008	0.82

[a] The quoted errors are purely statistical.

Table 5.3: Joint (Flux PDF + Flux PS) 1σ constraint on Γ_{12} for different thermal history (see Table 4.2) for redshift bin II (refer Table 5.1)

	$z = 0.2$		Joint constraint	
Model	T_0	γ	Γ_{12}[a]	$\chi^2_{\rm dof}$
$T10 - \gamma1.1$	4326	1.53	0.105 ± 0.015	1.22
$T10 - \gamma1.8$	4313	1.61	0.105 ± 0.015	1.22
$T20 - \gamma1.1$	4971	1.46	0.100 ± 0.014	1.13
$T20 - \gamma1.8$	4889	1.62	0.100 ± 0.013	1.11
$T15 - \gamma1.3$	4583	1.54	0.100 ± 0.015	1.18

[a] The quoted errors are purely statistical.

are small and within statistical uncertainty so for quoting best fit Γ_{12}, we used model $T20 - \gamma1.8$ which has minimum χ^2 in three out of four redshift bins.

The constraints on Γ_{12} can also depend on the cosmological parameters, e.g., any change in σ_8 can affect the overall normalization of flux PS since Lyα flux field and the matter density field are anti-correlated. In order to calculate uncertainty in Γ_{12}, we used Eq. 4.1 to propagate the error due to uncertainty in cosmological parameters. The uncertainty in cosmological parameters in this work is consistent with Planck Collaboration et al. (2016). We found that the change in Γ_{12} is ≤ 4 per cent due to uncertainty in σ_8 (in the range 0.820 to 0.848). Similarly, uncertainties in $\Omega_{\rm b}h^2$ and Ω_m would also affect the constraints on Γ_{12} (see Eq.

Table 5.4: Joint (Flux PDF + Flux PS) 1σ constraint on Γ_{12} for different thermal history (see Table 4.2) for redshift bin III (refer Table 5.1)

	$z = 0.3$		Joint constraint	
Model	T_0	γ	Γ_{12}[a]	$\chi^2_{\rm dof}$
$T10 - \gamma1.1$	4589	1.51	0.150 ± 0.021	1.05
$T10 - \gamma1.8$	4568	1.60	0.150 ± 0.020	1.04
$T20 - \gamma1.1$	5383	1.44	0.145 ± 0.023	0.99
$T20 - \gamma1.8$	5279	1.61	0.145 ± 0.022	0.99
$T15 - \gamma1.3$	4903	1.53	0.145 ± 0.023	1.02

[a] The quoted errors are purely statistical.

Table 5.5: Joint (Flux PDF + Flux PS) 1σ constraint on Γ_{12} for different thermal history (see Table 4.2) for redshift bin IV (refer Table 5.1). The Γ_{12} constraints are obtained from Lyβ contamination analysis (see section 5.5.1)

	$z = 0.4$		Joint constraint	
Model	T_0	γ	Γ_{12}[a]	$\chi^2_{\rm dof}$
$T10 - \gamma1.1$	4844	1.5	0.215 ± 0.025	1.09
$T10 - \gamma1.8$	4810	1.6	0.215 ± 0.025	1.08
$T20 - \gamma1.1$	5811	1.42	0.205 ± 0.027	1.01
$T20 - \gamma1.8$	5677	1.61	0.210 ± 0.030	0.96
$T15 - \gamma1.3$	5220	1.51	0.210 ± 0.030	1.06

[a] The quoted errors are purely statistical.

4.1). The combined uncertainty in Γ_{12} due to uncertainty in $\Omega_{\rm b}h^2$ (in the range 0.02184 to 0.02230) and Ω_m (in the range 0.297 to 0.323) (Planck Collaboration et al. 2016) is ~ 4 per cent. Thus we found that the uncertainties in parameters $\Omega_{\rm b}\,h^2$, Ω_m and σ_8 leads to ≤ 10 per cent change in Γ_{12} measurements. Note that we do not account for the correlation between these parameters. We also found that the derived value of Γ_{12} do not change with change in n_s from 0.96 to 1.0.[1]

One further source of error could come from the cosmic variance. We use

---

[1] We performed a GADGET-2 simulation with $n_s = 1.0$ and followed the method described in this paper to constraint Γ_{12}.

a box having a rather moderate size $50h^{-1}$ cMpc, which could in principle be smaller than the largest voids at $z \sim 0$ (see Fig. 2 in Mao et al. 2017). It is thus possible that our measurements of Γ_{12} may not be globally representative. To account for the effect, we simulate another box of identical size with identical parameters, however choosing a different set of initial conditions on the density and velocity fields. We perform the same statistical analysis on the second box and find that the constraints on Γ_{12} differ by ≤ 3 per cent. This is consistent with the finding of Smith et al. (2011) that simulations do converge for box sizes ≥ 50 cMpc for the number of particles considered in our simulations.

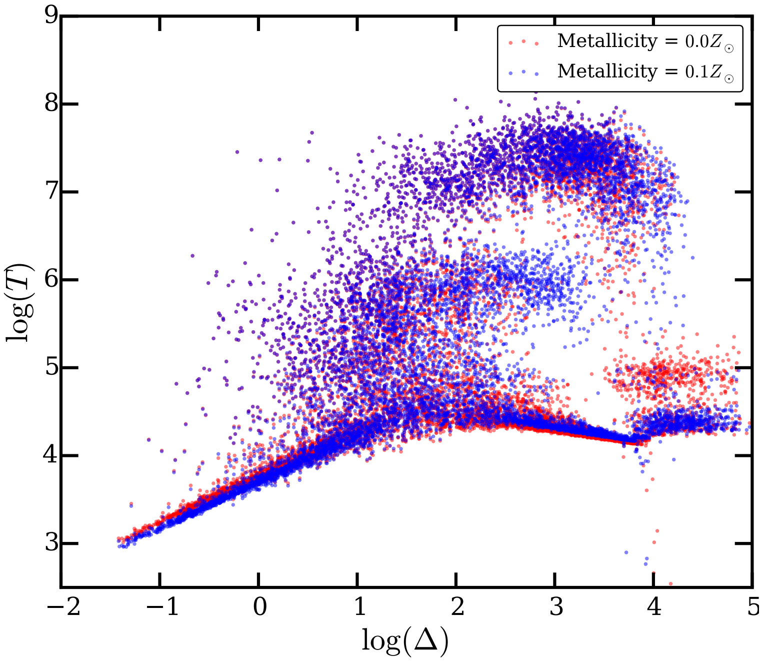

Figure 5.11: Effect of metallicity on $T-\Delta$ scatter plot for 10000 randomly selected SPH particles at $z = 0.3$. The temperature of SPH particles is calculated in the post-processing step of GADGET-2 using CITE. Red and blue points indicate the $T-\Delta$ scatter plot for $Z = 0.0Z_\odot$ and $Z = 0.1Z_\odot$ respectively. For $Z = 0.1Z_\odot$, the enhanced cooling rates due to metals are included from cooling tables given by Wiersma et al. (2009). As expected, the temperature of SPH particles for higher metallicity model is consistently lower by ~ 10 percent.

One possible source of uncertainty in $\Gamma_{\rm HI}$ could come from the metallicity of the IGM. At low-z, the IGM is enriched with metals (typical metallicity $\sim 0.1Z_\odot$

see Kulkarni et al. 2005; Shull et al. 2012). In presence of metals the cooling rates are enhanced which in turn can affect the $T - \Delta$ relation. To study the effect of metals on $T - \Delta$ relation, we included the cooling due to metals in CITE using cooling tables given by Wiersma et al. (2009)[1]. We enhance the cooling rates (due to metals) while calculating ionization fractions of species and hence temperature evolution. We do not account for the change in density due to metal cooling in CITE. We assume constant metallicity of $0.1 Z_{\odot}$ from redshift $z_1 = 2.1$ to $z = 0.0$. We found that due to metals the mean IGM temperature T_0 decreases by ~ 10 per cent whereas γ remains same (see Fig. 5.11). The fraction of baryons in the diffuse phase increase by ~ 2.5 per cent. Since Γ_{12} constraints are weakly dependent on T_0 and the fractional change in diffuse phase of baryons is small, the IGM metallicity has little effect on Γ_{12}.

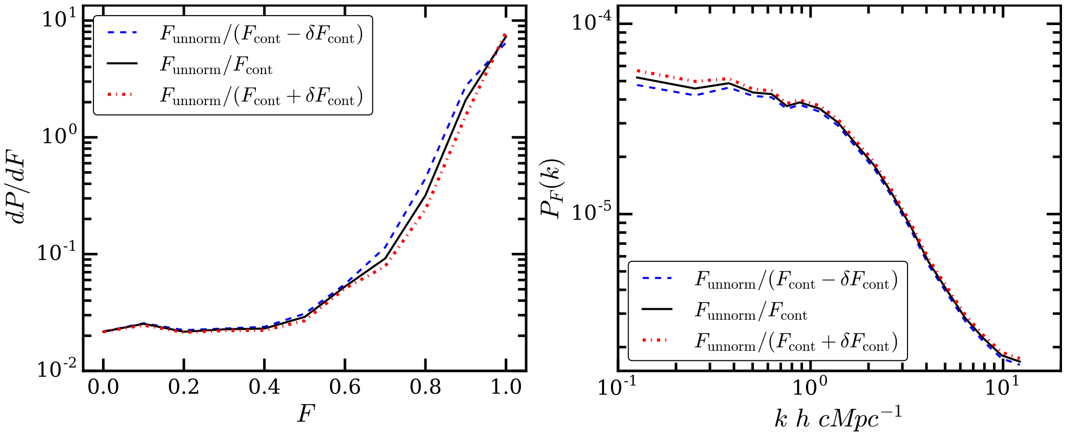

Figure 5.12: *Left* and *right-hand panels* show variation in flux PDF and PS (for observed spectra at $z = 0.2$) with uncertainty in continuum respectively. $F_{\rm unnorm}$ and $F_{\rm cont}$ are unnormalized and continuum flux respectively. $\delta F_{\rm cont}$ indicates 1σ uncertainty in continuum flux. Due to continuum uncertainty the flux PDF is affected more as compared to flux PS

The final source of errors on the measured Γ_{12} is the uncertainty in the continuum fitted to the observed spectra. As discussed earlier, each observed spectrum is fitted with the continuum by Danforth et al. (2016). However, because of the limited SNR the continuum fitting procedure is prone to uncertainties.

---

[1] http://www.strw.leidenuniv.nl/WSS08/

Table 5.6: Total error budget for our Γ_{12} measurements at different redshifts

Redshift bin ⇒ Type of simulated spectra ⇒	I Lyα forest	II Lyα forest	III Lyα forest	IV Lyα + Lyβ forest[1]
Best Fit Γ_{12}	0.066	0.100	0.145	0.210
Statistical Uncertainty[a]	± 0.007	± 0.013	± 0.022	± 0.030
Cosmological parameters (∼ 10 per cent) [b]	± 0.007	± 0.010	± 0.015	± 0.021
Cosmic Variance (∼ 3 per cent) [c]	± 0.002	± 0.003	± 0.004	± 0.006
Total statistical errors[d]	± 0.010	± 0.016	± 0.027	± 0.037
Continuum uncertainty (systematic)[e]	± 0.005	± 0.005	± 0.010	± 0.015
Total error[f]	± 0.015	± 0.021	± 0.037	± 0.052

[a] The best fit value and statistical uncertainty is given for the model $T20-\gamma1.8$ (see Table 4.2) since χ^2 is minimum as compared to other models in 3 out of 4 redshift bins.

[b] Cosmological parameters: Uncertainty due to ($\Omega_b h^2$, Ω_m, σ_8) is (∼ 2 per cent, ∼ 2 per cent, ∼ 4 per cent) respectively. The correlation between different parameters is not accounted for hence the uncertainty is conservative.

[c] To account for cosmic variance, we simulated two different boxes with identical cosmological parameters but with different initial conditions.

[d] All statistical errors are added in quadrature.

[e] This uncertainty arises due to continuum fitting uncertainty. This error is systematic in nature.

[f] Total error is obtained by adding total statistical error with total systematic error.

[1] Simulated Lyα forest at $z = 0.35$ to 0.45 is contaminated by Lyβ forest in the same wavelength range. The Lyβ forest is generated from simulation box at $z = 0.6$.

Conventionally, the observed flux is normalized as $F = F_{\rm unnorm}/F_{\rm cont}$, where $F_{\rm unnorm}$ and $F_{\rm cont}$ are the unnormalized and continuum flux, respectively. If $\delta F_{\rm cont}$ is the value of the uncertainty in the continuum, we can calculate the lower and upper bounds on the normalized flux as $F_{\rm lb} = F_{\rm unnorm}/(F_{\rm cont} + \delta F_{\rm cont})$ and $F_{\rm ub} = F_{\rm unnorm}/(F_{\rm cont} - \delta F_{\rm cont})$, respectively. Fig. 5.12 shows the effect of continuum fitting uncertainties on the observed flux PDF (left-hand panel) and flux PS (right-hand panel). When $\delta F_{\rm cont}$ is taken to be 1σ uncertainty in the continuum, the flux PDF is considerably affected by the continuum uncertainty, whereas the effect on the flux PS is milder and the changes are well within the errorbars. This is the main reason to constrain Γ_{12} from flux PS and flux PDF jointly. In this work, we obtained constraints on Γ_{12} using the three estimates of the transmitted flux F_{lb}, F, and F_{ub} and found that the Γ_{12} range changes systematically by ≤ 10 per cent.

We summarize the total error budget (statistical and systematic) in Table

5.7. The errors on Γ_{12} are thus calculated as follows: (i) We first estimate the error through the χ^2 minimization for the fiducial values of T_0 and γ. (ii) We then account for the uncertainties in T_0 and γ by obtaining constraints for models with extreme values of the two parameters. (iii) We add all statistical uncertainties in quadrature to account for uncertainties in the cosmological parameters, thermal history and cosmic variance. (iv) We finally add total statistical uncertainty with systematic (from the continuum fitting) uncertainty. The Γ_{12} constraints accounting for statistical and systematic uncertainties are given in Table 5.7.

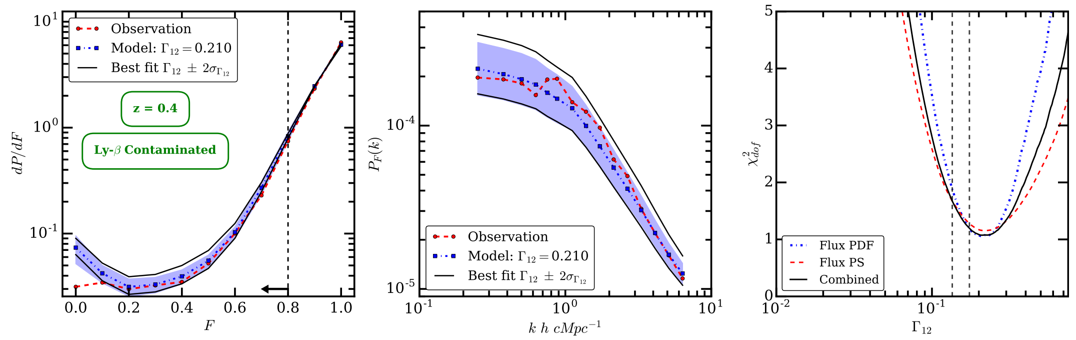

Figure 5.13: Each panel is same as explained in Fig. 5.10 except that the Lyα forest in the redshift range $z = 0.4 \pm 0.05$ is contaminated by Lyβ forest from high redshift ($z = 0.6$). Comparison of the bottom row in Fig. 5.10 (without Lyβ contamination) with this plot shows that the Γ_{12} constraints are underpredicted (at $z = 0.4$) when Lyβ contamination is not taken into account.

The redshift bin $z = 0.4$ is likely to be contaminated by Lyβ lines from H I interlopers (Danforth et al. 2016)[1]. The contaminated spectrum will have more absorption hence Γ_{12} would be underpredicted as compared to uncontaminated spectrum. To account for this we contaminated the Lyα forest (at $z = 0.4$) with Lyβ forest[2] from $z = 0.6$. We have contaminated the region after accounting for observed QSO emission redshift and avoiding proximity region. Fig. 5.13

---

[1] Note that one can still identify such lines using Ly-γ, Ly-δ transition if the Lyβ line is sufficiently saturated. However here we are concerned about the lines which have strong Lyα and Lyβ transition but weak Ly-γ, Ly-δ transition.

[2] We have only one free parameter Γ_{12} at $z = 0.4$ denoted by $\Gamma_{12,0.4}$ in this analysis. For a given $\Gamma_{12,0.4}$, we obtained $\Gamma_{12,0.6}$ at $z = 0.6$ by using scaling relation $\Gamma_{12,0.6} = \Gamma_{12,\rm norm}(1+z)^{4.4}$ where $\Gamma_{12,\rm norm} = \Gamma_{12,0.4}/1.4^{4.4}$ as found by Shull et al. (2015).

shows that Γ_{12} measurement obtained from contaminated spectra are higher as compared to those from uncontaminated spectra (Fig. 5.10 bottom row). This problem does not arise for other redshift bins because the Lyβ lines are identified based on the Lyα detected in the HST-COS spectrum and we removed these lines, higher H I Ly-series and metal lines in contaminating line removal process as illustrated in Fig. 5.2.

5.5.2 Voigt statistics: CDDF

Right-hand panels of Fig. 5.14 show the constraints on Γ_{12} from CDDF for a model without turbulence in the 4 different redshift bins. As expected the χ^2 curves are smooth parabolas with the minimum reduced ($\chi^2_{\mathrm{dof,min}}$) ~ 1.5 as given in Table 5.7. The statistical uncertainty in Γ_{12} is shown by black dashed vertical line. The statistical uncertainty in Γ_{12} is calculated by demanding $\chi^2 = \chi^2_{\mathrm{min}} + \Delta\chi^2$, where χ^2_{min} is minimum value of χ^2 and $\Delta\chi^2 = 1.0$ for 1 degree of freedom (corresponds to 1 free parameter in the problem i.e., Γ_{12}; Press et al. 1992). The shaded regions shown in the right-hand panels correspond to 1σ constraint on Γ_{12} from Paper-I. It is interesting to note that the Γ_{HI} constraints obtained using CDDF in this work (see Table 5.7) are consistent within $\sim 1\sigma$ (the best fit values differ by < 5.5 percent) with those obtained using flux statistics in Paper-I (see Table 8 in Paper-I). However, error range is smaller in the present study. It is because, contrary to Paper-I, here we do not consider the errors arising from uncertainties in cosmological parameters, continuum fitting and cosmic variance. Therefore, we caution the reader that errorbars on Γ_{12} from Paper-I are more realistic.

Left-hand panels in Fig. 5.14 shows the best fit (i.e., models with Γ_{12} corresponding to minimum χ^2_{dof}) model CDDF (blue square with shaded region) and observed CDDF (red circle with errorbar) in the 4 redshift bins. The gray shaded region represents 1σ range from the simulated mock sample (σ_{sim} given in Eq.5.7). The errorbars shown on observed CDDF are assumed to be poisson distributed and are not used while calculating χ^2.

The redshift bin IV is likely to be affected by Lyβ contamination from H I interlopers (refer to D16 and Paper-I). Note that it is difficult to remove the

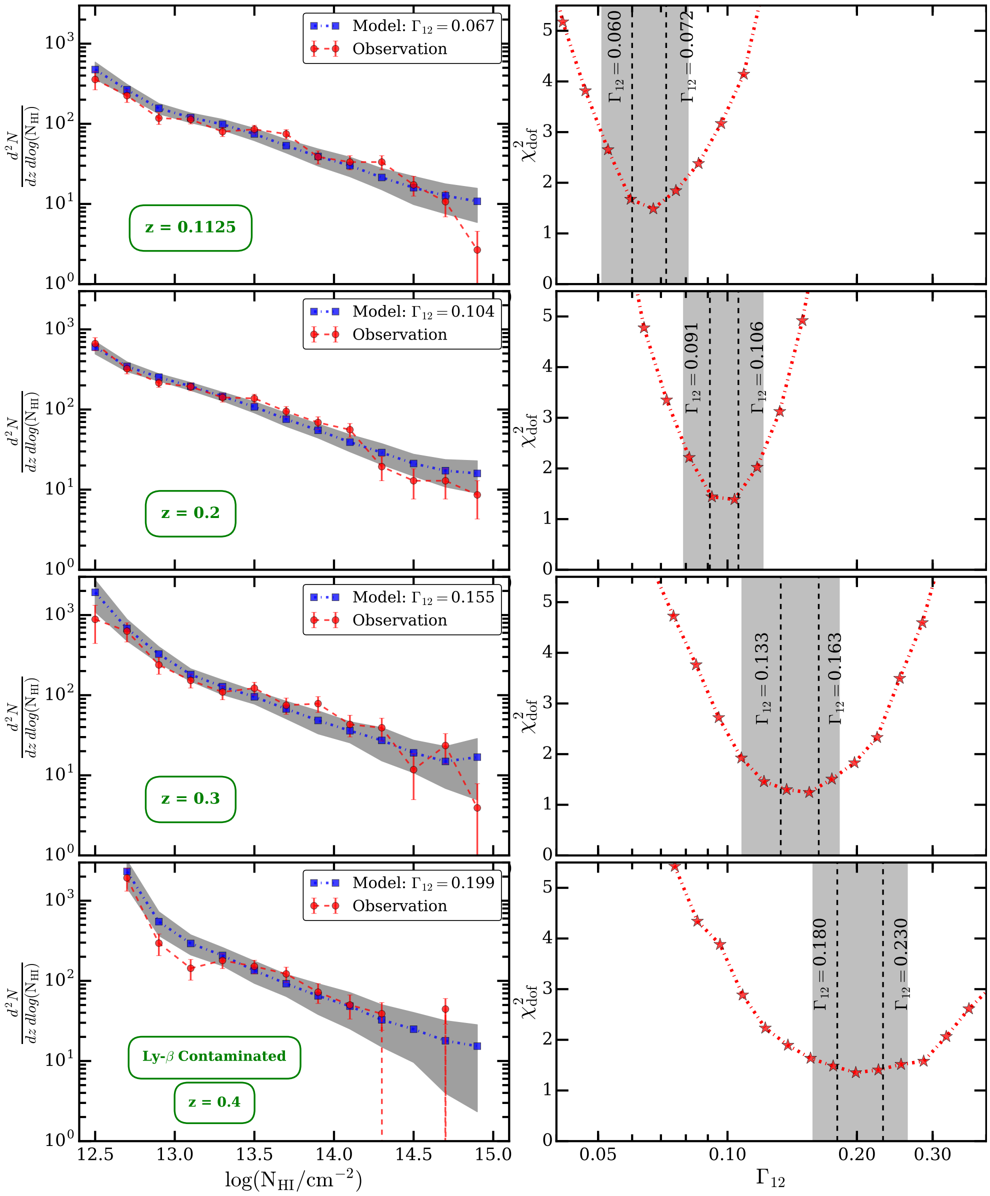

Figure 5.14: Constraints on Γ_{12} from CDDF in four different redshift bins. The *left-hand* panels show the observed CDDF and CDDF from the simulation using the best fitted Γ_{12}, i.e., Γ_{12} corresponding to the minimum reduced χ^2, in different redshift bins (given in green box). The gray shaded region shows 1σ errorbar on the model CDDF (calculated from mock samples). The errors on observed CDDF (red line with errorbar) are obtained assuming poisson distribution and are not considered for calculating the reduced χ^2. The *right-hand* panels show the reduced χ^2 as a function of the assumed Γ_{12}. The black dashed vertical lines indicate 1σ constraints on Γ_{12} around the Γ_{12} where reduced χ^2 is minimum. The shaded region indicates 1σ constraints on Γ_{12} from flux PDF and flux PS given in Paper-I. The 1σ constraints on Γ_{12} from CDDF are well within those obtained using flux PDF and flux PS. The simulated Lyα forest at $z = 0.4$ is contaminated by Lyβ forest from $z \sim 0.6$ in the same wavelength range to account for possible contamination due to intervening H I absorbers (see section 5.5 for more details).

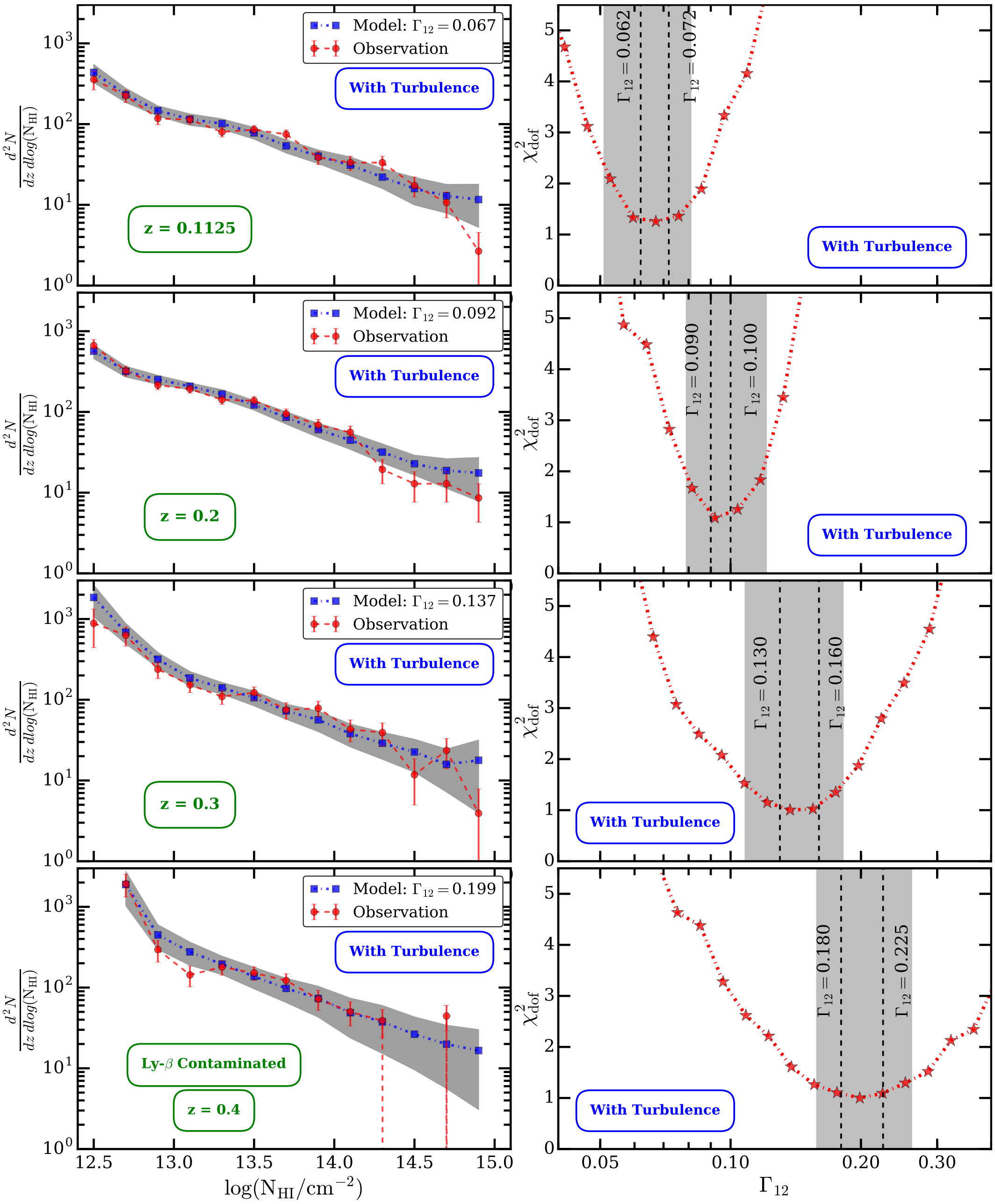

Figure 5.15: Same as Fig. 5.14 except Γ_{12} is constrained using a model with Gaussian distributed b_{turb} see section 5.4.2.1.

Table 5.7: Γ_{12} measurements at different redshifts

Redshift bin ⇒	I	II	III	IV
Type of simulated spectra ⇒	Lyα forest	Lyα forest	Lyα forest	Lyα + Lyβ forest[1]
Best Fit Γ_{12} (without turbulence)	0.066	0.104	0.137	0.199
Statistical uncertainty[2]	±0.006	±0.008	±0.015	±0.025
Reduced $\chi^2_{\rm min}$	1.49	1.39	1.24	1.35
Best Fit Γ_{12} (for Gaussian $b_{\rm turb}$)	0.067	0.095	0.145	0.200
Statistical uncertainty[2]	±0.005	±0.005	±0.015	±0.022
Reduced $\chi^2_{\rm min}$	1.26	1.09	1.04	1.00

[1] Following Paper-I, simulated Lyα forest at $z = 0.35$ to 0.45 is contaminated by Lyβ forest in the same wavelength range. The Lyβ forest is generated from simulation box at $z = 0.6$.

[2] The uncertainty in Γ_{12} due to uncertainty in thermal history parameters is well within statistical uncertainty.

Lyβ contamination in observations due to limited wavelength coverage of the spectrograph. We accounted for this effect by contaminating the simulated Lyα forest in the redshift $0.35 \leq z \leq 0.45$ with Lyβ forest from simulation box at $z = 0.6$. (for more detail we refer reader to section 5.1 in Paper-I). The lowest panels in Fig. 5.14 shows the Γ_{12} constraint at $z = 0.4$ where model CDDF is calculated from Lyα forest contaminated with Lyβ forest. The $\chi^2_{\rm dof,min} = 1.35$ and Γ_{12} constraint are consistent (within 1σ) with best fit Γ_{12} from Paper-I (0.210 ± 0.052) in the same redshift bin[1].

We have done similar analysis and constrained Γ_{12} for a model with Gaussian distributed $b_{\rm turb}$ (see section 5.4.2.1) as shown in Fig. 5.15 (see Table 5.7 also). It is interesting to note that the Γ_{12} constraints obtained from the model with Gaussian distributed $b_{\rm turb}$ are in good agreement with those from model without turbulence and Paper-I (see left-hand panel in Fig. 5.17). This suggests that the addition of Gaussian distributed $b_{\rm turb}$ has a mild effect on CDDF (and hence Γ_{12} constraints). The $\chi^2_{\rm dof,min}$ is also improved and close to 1 when Gaussian distributed $b_{\rm turb}$ is added to the model. We also find that our CDDF at $\log \rm N_{HI} > 13.6$ from simulation is steeper than that from S15 and Viel et al. (2017). This could be due to absorption along sightlines passing through the high density

---

[1] If we do not account for Lyβ contamination, the Γ_{12} constraint is underestimated, a result similar to Paper-I. For Lyα forest only simulation at $z = 0.4$, $\Gamma_{12} = 0.176 \pm 0.013$

environment (near the outskirts of galaxies) in our simulation. In addition, the number of high density regions are likely to be smaller in the simulations of S15 and Viel et al. (2017) as they incorporate various feedback processes and star formation criteria.

We summarize the best fit Γ_{12} (for the model with and without turbulence) along with the statistical errors for 4 redshift bins in Table 5.7. It is clear from left-hand panel of Fig. 5.17 that there is evolution in Γ_{12} with redshift and that follows $\Gamma_{\rm HI}(z) = (3.9 \pm 0.1) \times 10^{-14}\ (1+z)^{4.98\pm0.11}\ {\rm s}^{-1}$ up to $z = 0.5$.

5.6 Evolution of Γ_{12}

Having obtained the constraints on Γ_{12} at different redshift bins, we can now try to understand its redshift evolution. Left-hand panel in Fig. 5.16 shows the constraints on Γ_{12} using the combined χ^2 analysis of flux PDF and PS at the four redshift bins. Blue open circle at $z = 0.4$ in the left-hand panel shows that the Γ_{12} measurements after Lyβ contamination is accounted for properly. We find that there is a clear trend of photoionization rate increasing with increasing redshifts. The best fit values follow $\Gamma_{12} = 0.040 \pm 0.001(1+z)^{4.99\pm0.12}$.

In the left-hand panel of Fig. 5.17, we compare our derived $\Gamma_{\rm HI}$ from CDDF with that from flux PDF and PS. The Γ_{12} evolution using CDDF for models with and without turbulence is in good agreement with Γ_{12} evolution using flux PDF and flux PS). The best fit values follow $\Gamma_{12} = 0.040 \pm 0.001(1+z)^{4.99\pm0.12}$ for the constraints using CDDF. The uncertainty in Γ_{12} constraint from flux PDF and flux PS is *larger and more reliable* as it also account for other systematic and statistical uncertainties.

Right-hand panel in Fig. 5.17 shows that our derived Γ_{12} (0.066 ± 0.06) at $z = 0.1$ is lower than Γ_{12} from K14 (~ 0.178) by factor ~ 2.7 (see Appendix 5.8.2 for the reason of discrepancy) but is consistent with S15 (~ 0.070) within 1σ. Our Γ_{12} constraint at $z = 0.2$ is also in agreement with those obtained from modeling the observed metal abundances by Shull et al. (2014). Note previous studies do not get $\Gamma_{\rm HI}$ through χ^2 minimization and no errorbars are associated to the $\Gamma_{\rm HI}$ measurement. K14 compared the CDDF in the redshift range $0 \leq z \leq 0.75$ with models generated from simulation box at $z_{\rm sim} = 0.1$. This could be the possible reason for the discrepancy of Γ_{12} constraints obtained in this work and K14. We constrained Γ_{12} from CDDF in $0.075 \leq z \leq 0.45$ using models generated from simulation box at $z_{\rm sim} = 0.1$ ignoring evolution in Γ_{12} and large scale structures. The Γ_{12} constraint (0.092 ± 0.009) for this model is higher by factor ~ 1.4 as

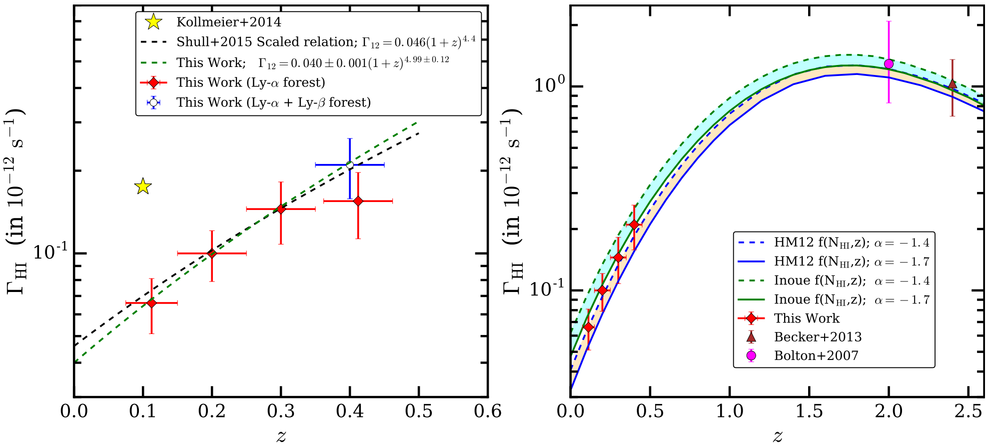

Figure 5.16: *Left-hand* panel shows the Γ_{12} constraint from joint χ^2 analysis of flux PS and flux PDF. The red diamonds show our Γ_{12} constraints using the simulated Lyα forest. The last $(z = 0.4)$ bin is likely to be affected by Lyβ forest from H I interlopers at high redshift. The blue open circle corresponds to the Γ_{12} constraint using simulated Lyα forest contaminated by Lyβ forest at $z = 0.6$. A best fit power-law to our measurements is also shown with green dashed line. The scaling relation used by Shull et al. (2015) (black dashed line), where they increased the Γ_{12} evolution of HM12 by a factor 2, is also consistent with our measurements. However our Γ_{12} at $z = 0.1$ is factor ~ 2.7 smaller than Kollmeier et al. (2014) (yellow star). *Right-hand* panel shows the Γ_{12} evolution from $z = 0$ to 2.5 from observations and different UVB models. The cyan and orange shaded regions show evolution of Γ_{12} from KS15 UVB for $f_{\rm esc} = 0$ using Inoue et al. (2014) and HM12 cloud distribution respectively. The shaded region accounts for uncertainty in UV spectral index $\alpha = -1.4$ to -1.7 at $\lambda < 912$ Å. Our results (shown by red diamonds) are consistent with $f_{\rm esc} = 0$ for HM12 and Inoue et al. (2014) cloud distribution allowing for the UV spectral index uncertainties. A constant $f_{\rm esc} = 0$ model (for different cloud distribution and FUV spectral index uncertainty) is sufficient to explain the evolution of Γ_{12} from $z = 0$ to $z = 2.5$ (high-z points are taken from Bolton and Haehnelt (2007) and Becker and Bolton (2013)). All of these predictions use the QSO emissivity of KS15, and no galaxy contribution.

compare to Γ_{12} constraint from the model at $z = 0.1$ where Γ_{12} evolution is accounted for (see Table 5.7). However, the Γ_{12} constraint from this model is in good ($\sim 1\sigma$) agreement with Γ_{12} constraint at $z = 0.2$ (0.104 ± 0.008, see Table 5.7) when Γ_{12} evolution is accounted for. This is still smaller than what was found by K14. Note that the median redshift of the observed sample is $z \sim 0.2$. Thus even if we do not account for Γ_{12} evolution, the Γ_{12} constraints (0.092 ± 0.009) are consistent with S15 (~ 0.103 at $z = 0.2$) and Paper-I (0.100 ± 0.021 at $z = 0.2$).

Recently, Gurvich et al. (2017) found that the Γ_{12} required to match observed CDDF (from D16) with simulated CDDF is lower than K14 by a factor ~ 3. They attributed this to the effect of AGN feedback and to the Faucher-Giguère et al. (2009) UVB model included in their simulation. The AGN feedback could be important but is not incorporated in our simulation. However our Γ_{12} constraints are in good agreement with those from Gurvich et al. (2017). Note that our Γ_{12} constraint at $z = 0$ ($\Gamma_{12} = 0.039 \pm 0.001$, from scaling relation) is consistent within 1σ with that from Faucher-Giguère et al. (2009) UVB model ($\Gamma_{12} = 0.0384$). Viel et al. (2017) have provided scaled Γ_{12} for different hydrodynamical simulations with and without additional heating that will reproduce the observed CDDF. The direct comparison between our Γ_{12} measurements and that of Viel et al. (2017) is difficult as goodness of the CDDF fits and error in Γ_{12} are not given in their work. However if we assign 10 percent uncertainty to their measurements, they are very much in agreement with our measurements. Cristiani et al. (2016) found a similar result where their estimated Γ_{12} evolution at $z < 0.5$ is in good agreement with Γ_{12} evolution from S15 and Paper-I. Cristiani et al. (2016) combined the information from measured QSO contribution to cosmic UVB in the range $3.6 < z < 4.0$ with QSO luminosity function and estimated the production of ionizing photons from QSOs at various epoch ($0 \leq z \leq 5$). To compare our results with the UVB models, we calculate the UVB as explained in Khaire and Srianand (2013) and KS15. The UVB estimate depends on the UV emissivity of QSOs and galaxies and the H I column density distribution, $f(N_{\rm HI}, z)$, of the IGM. We use the QSO emissivity from KS15 at 912 Å (their Eq. 6) obtained using the recent QSO luminosity functions. At $z < 3$, this ϵ_{912} is higher upto factor of 2.2 than the one used by previous UVB models such as Faucher-Giguère et al. (2009); Haardt and Madau (2012, hereafter HM12) which they obtained from the

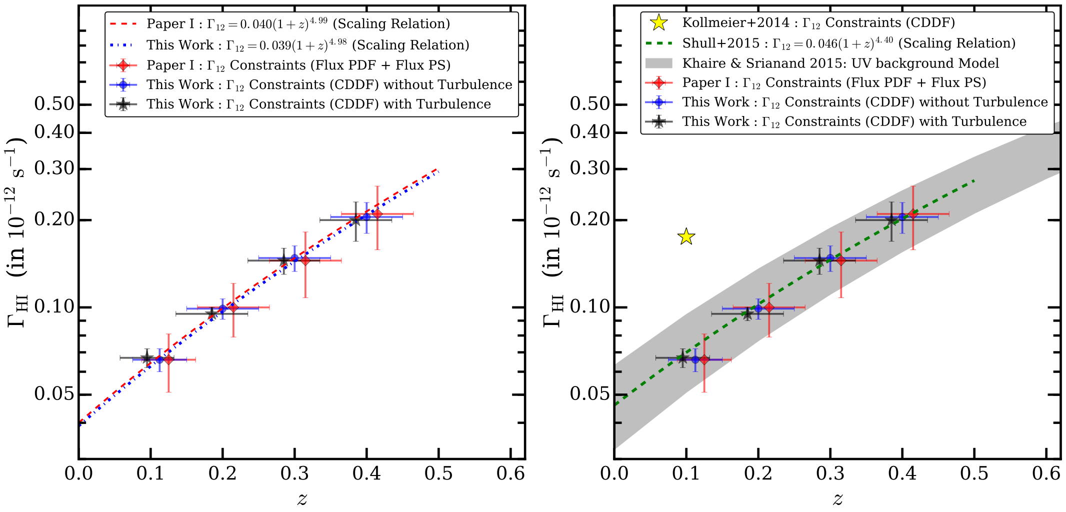

Figure 5.17: In the left-hand panel black stars and blue circles with errorbar show the evolution of Γ_{12} from this work (CDDF) for a model with turbulence (see section 5.4.2.1) and without turbulence respectively. Red diamonds with errorbars show the Γ_{12} evolution from Paper-I (using flux PDF and flux PS). For visual purpose the points are shifted along x axis. The blue dotted and red dashed line shows scaling relation for Γ_{12} evolution from this work (for model without turbulence) and Paper-I. Within 1σ uncertainty the Γ_{12} evolution from this work (for both models with and without turbulence) is consistent with that from Paper-I. The errorbars on Γ_{12} evolution given in Paper-I are more realistic as they account for cosmic variance, cosmological parameter uncertainty, continuum fitting uncertainty. In the right-hand panel black stars, blue circles and red diamonds with errorbars are same as given in left-hand panel. Our derived Γ_{12} at $z \sim 0.1$ is lower by factor ~ 2.7 than that of K14 (shown by yellow star) but is in 1σ agreement with S15 (green dashed line). The derived $\Gamma_{\rm HI}$ evolution is consistent with Khaire and Srianand (2015a,b) UVB model (shown by gray shaded region) where the UVB is contributed only by QSOs.

compiled luminosity functions of Hopkins et al. (2007). It is because the recent QSO luminosity function compiled by KS15 from Croom et al. (2009); Palanque-Delabrouille et al. (2013); Schulze et al. (2009) have more bright QSOs. The ratio of the ϵ_{912} from KS15 to the one used by HM12 at $z = 0.1, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$ is $1.16, 1.61, 2.03, 2.11, 1.86, 1.54, 1.28$ respectively (see Fig. 2 of KS15). The QSO emissivity at $\lambda < 912$ Å depends on the assumed spectral energy distribution (SED) of QSOs. The SED can be approximated by a power-law, $L_\nu \propto \nu^\alpha$, for $\lambda < 912$ Å. We refer to α as UV spectral index. Here, we use $\alpha = -1.4$ from Stevans et al. (2014) as well as $\alpha = -1.7$ from Lusso et al. (2014). Stevans et al. (2014) obtained it by stacking the FUV spectra of 159 QSOs at $0.001 < z < 1.476$ observed with HST-COS which probes rest frame wavelengths down to 475 Å. Whereas, Lusso et al. (2014) obtain it by stacking spectra of a sample of 53 QSOs at $z \sim 2.4$ observed using Wide Field Camera 3 on HST which probes rest wavelength down to 600Å. To model the UVB, we take the galaxy emissivity using a fiducial self-consistent combination of star formation rate density (SFRD) and dust attenuation from Khaire and Srianand (2015b) (as summarized in section 3 of Khaire et al. 2016) with $f_{\rm esc}$ being a free parameter. In general, the UVB at $z < 0.5$ has significant contribution coming from high-z sources upto $z \sim 2$ due to steep rise in QSO emissivity as well as SFRD with z. Therefore, we need to constrain the $f_{\rm esc}$ at $z < 2$. For simplicity, we assume a constant $f_{\rm esc}(z)$ over this redshift range.

The $f(N_{\rm HI}, z)$ for $\log N_{\rm HI} > 16$ affects the UVB significantly. It is not well constrained at $z < 2$ due to the small number statistics and because one needs to use spectra obtained with space based observatories. Here, we use recently updated $f(N_{\rm HI}, z)$ from Inoue et al. (2014) which is different than the one used by HM12. For comparison, we calculate UVB using $f(N_{\rm HI}, z)$ of both Inoue et al. (2014) as well as HM12.

The $\Gamma_{\rm HI}(z)$ obtained using these UVB models calculated for two UV spectral indexes and two $f(N_{\rm HI}, z)$ are shown in the right-hand panel of Fig. 5.16. Here, we assumed $f_{\rm esc}(z)$=0 for $z < 3$ and at $z > 3$ $f_{\rm esc}(z)$ is taken from Khaire et al. (2016). As shown in Fig. 5.16, all four UVB models are consistent with our low-z $\Gamma_{\rm HI}$ measurements. As can be seen from the figure, QSOs alone are sufficient to explain the $\Gamma_{\rm HI}$ measurements upto $z = 2.5$ (i.e with $f_{\rm esc}(z) = 0$).

At higher-z a rapid evolution in $f_{\rm esc}$ is needed as shown in Khaire et al. (2016). We calculate the reduced χ^2 for these UVB models obtained with $f_{\rm esc} = 0$. The UVB with $f(N_{\rm HI}, z)$ of HM12 gives reduced $\chi^2_{\rm dof} = 0.14$ and 1.26 for $\alpha = -1.4$ and $\alpha = -1.7$, respectively. The UVB with $f(N_{\rm HI}, z)$ of Inoue et al. (2014) gives reduced $\chi^2_{\rm dof} = 3.06$ and 0.22 for $\alpha = -1.4$ and $\alpha = -1.7$, respectively.

For UVB model obtained using $f(N_{\rm HI},\, z)$ of Inoue et al. (2014) and $\alpha = -1.7$, we find that $f_{\rm esc}(z < 2) = 0.008$ is a conservative 3σ upper limit (with $\Delta\chi^2 \sim 9$). It is consistent with the 3σ upper limits on average $f_{\rm esc} \leq 0.02$ obtained by stacking samples of galaxies (Bridge et al. 2010; Cowie et al. 2009; Rutkowski et al. 2016; Siana et al. 2010). In these observations, lowest average mass of galaxies is $\sim 10^{9.3}$ M$_\odot$ (Rutkowski et al. 2016). It was believed that the UV escape from the lower mass galaxies could be appreciable at all z (Ferrara and Loeb 2013; Fujita et al. 2003; Razoumov and Sommer-Larsen 2006; Wise et al. 2014) which may solve the problem of higher $f_{\rm esc}$ ($\sim$ 0.15 to 0.2 Khaire et al. 2016; Mitra et al. 2015) at $z > 6$ required for H I reionization. However our derived 3-σ upper limit of $f_{\rm esc}$ is in conjunction with observations of Rutkowski et al. (2016), suggests that galaxies with mass lower than $10^{9.3}$ M$_\odot$ may also have very low $f_{\rm esc}$ providing negligible contributions to the UVB at low-z. It is possible that some additional heating as suggested recently by Viel et al. (2017) may be present and if included that will reduce the derived $\Gamma_{\rm HI}$ further. This will further strengthen our conclusion that QSOs alone are sufficient to provide necessary H I ionizing photons.

5.7 Summary

In this work we measure the H I photoionization rate, $\Gamma_{\rm HI}$, at $z \leq 0.45$ using a sample of QSO spectra obtained with Cosmic Origins Spectrograph onboard the Hubble Space Telescope (HST-COS) and hydrodynamical simulations GADGET-2 post-processed with CITE, GLASS and VIPER.

For fair comparison, we mimic the simulated Lyα forest as close to observations as possible in terms of SNR, resolution and line spread function. The spectra generated using our method are remarkably similar to the observed spectra. We constrain $\Gamma_{\rm HI}$ using three statistics namely, flux PDF, flux PS and CDDF. We perform χ^2 minimization using appropriate covariance matrices to compare the observations with the model predictions. The main results of our work are as follows,

1. We measured Γ_{HI} in four different redshift bins (of $\Delta z = 0.1$) centered at $z = 0.1125, 0.2, 0.3, 0.4$ using joint constraints from the two statistics flux PDF and PS. We estimated the associated errors by varying thermal history parameters, cosmological parameters and continuum fitted to the observed spectrum. Due to limited wavelength range covered in the HST-COS spectrum used in this study, the Γ_{HI} measurement for the highest redshift bin (i.e $z = 0.4$) is likely to be affected by the contamination of Lyβ forest absorption from higher-z. We contaminated our simulated Lyα forest at $z = 0.4$ by Lyβ forest from $z = 0.6$ and corrected for the effect of Lyβ contamination in our Γ_{HI} measurement for this z bin. The measured Γ_{12} values at redshift bins $z = 0.1125, 0.2, 0.3, 0.4$ are 0.066 ± 0.015, 0.100 ± 0.021, 0.145 ± 0.037, 0.210 ± 0.052, respectively.

2. Our final quoted errors in the Γ_{HI} measurements include possible uncertainties coming from the statistical uncertainty[1] ($\sim$ 14 per cent), cosmic variance ($\sim$ 3 per cent), cosmological parameters uncertainty ($\sim$ 10 per cent) and continuum uncertainty (systematic uncertainty $\sim$ 7 per cent). Uncertainty in Γ_{HI} due to uncertainty in thermal history parameters, over the range considered here, is small and within statistical uncertainty.

3. We fitted all the observed Lyα forest spectra using VIPER and compiled a Lyα line catalog called "VIPER line catalog". We found that the b parameters of Voigt profile components from simulations are typically underestimated as compared to observations. This difference can be rectified by including the Gaussian distributed line width parameter b_{turb} ($\mu = 20$ km s^{-1}and $\sigma = 10$ km s^{-1}) at each pixel in the simulation. The resulting line width distribution from simulations matches roughly with observed line width distribution, scatter and lower envelope of the b vs $\log \mathrm{N_{HI}}$ distribution.

---

[1] The percentage values given in parenthesis are quoted for redshift bin IV i.e., $z = 0.4$ (see Table 5.7)

However, the CDDF has little effect of additional $b_{\rm turb}$ (<7 percent) and the $\Gamma_{\rm HI}$ constraints are mildly affected (<9 percent). However, if we consider additional heating effect for the excess broadening then the $\Gamma_{\rm HI}$ obtained will be slightly reduced (roughly scale as $T^{-0.7}$).

4. We obtained CDDF at four different z bins and matched with simulated CDDF at each mean z. This allowed us to measure the $\Gamma_{\rm HI}$ in four redshift bins (of $\Delta z = 0.1$) centered around $\overline{z} = 0.1125, 0.2, 0.3, 0.4$. We estimated the associated statistical error using χ^2 statistics. When additional turbulent broadening are not included measured Γ_{12} values at the redshift bins $\overline{z} = 0.1125, 0.2, 0.3, 0.4$ are $\Gamma_{12} = 0.066 \pm 0.006, 0.104 \pm 0.008, 0.137 \pm 0.015, 0.199 \pm 0.025$ respectively. The corresponding values after inclusion of $b_{\rm turb}$ are $\Gamma_{12} = 0.067 \pm 0.005, 0.095 \pm 0.005, 0.145 \pm 0.015, 0.200 \pm 0.022$. Thus the uncertainties in the velocity broadening seem to have little effect on the derived $\Gamma_{\rm HI}$. Our measured Γ_{12} values are in good agreement with Γ_{12} measurement that are obtained with two different statistics namely flux PDF and flux PS. However, the errorbars on $\Gamma_{\rm HI}$ measurements from flux PDF and flux PS analysis (Gaikwad et al. 2017b) are more reliable as they account for cosmic variance, continuum fitting uncertainty and cosmological parameter uncertainty in their calculation.

5. As expected based on UVB models, even in the small redshift range covered in our study the measured $\Gamma_{\rm HI}$ shows a rapid evolution with z. We fit the redshift evolution of Γ_{12} obtained from flux PDF and PS as $\Gamma_{12} = 0.040 \pm 0.001\,(1+z)^{4.99 \pm 0.12}$ and that obtained from CDDF as $\Gamma_{12} = 0.039 \pm 0.001\ (1+z)^{4.98 \pm 0.11}$ at $0.075 \leq z \leq 0.45$. The $\Gamma_{\rm HI}(z)$ obtained here are consistent with the measurement of Shull et al. (2015) however our $\Gamma_{\rm HI}$ measurement at $z = 0.1$ is factor ~ 2.7 smaller than Kollmeier et al. (2014). Note these two earlier measurements used H I column density distribution to constrain $\Gamma_{\rm HI}(z)$.

6. The $\Gamma_{\rm HI}$ measurement at any z_1 depends on the emissivities of the ionizing sources at $z \geq z_1$ and Lyman continuum opacity of the IGM. We considered the updated emissivities of QSOs and galaxies (with $f_{\rm esc}$ as a free parameter)

and two different H I column density distribution as a function of z obtained by HM12 and Inoue et al. (2014) and obtained $\Gamma_{\rm HI}$ using KS15 UVB code. We find that for, both H I distributions, our derived $\Gamma_{\rm HI}(z)$ is consistent with being contributed only by QSOs. This is true even if we allow for variations in the UV spectral index of QSOs. We also find the maximum 3σ upper limit on $f_{\rm esc}$ at $z < 2$, allowing for uncertainty in FUV spectral index and cloud distribution $f(N_{\rm HI}, z)$ of Inoue et al. (2014), is 0.008. This is consistent with 3σ upper limits on average $f_{\rm esc}$ (i.e ≤ 0.02) obtained by stacking samples of galaxies probing average galaxy mass $M \geq 10^{9.3} M_{\odot}$.

Our measurements suggest that the contribution of low mass galaxies to average $f_{\rm esc}$ will also be small. *Our study confirms that there is no crisis at low redshift in accounting for the observed Lyman continuum photons using standard known luminous astronomical sources.* Thus our $\Gamma_{\rm HI}(z)$ measurement can in turn be used to place a strong constraint on the contributions of decaying dark matter to the low-z UVB.

5.8 Appendix

5.8.1 Flux PDF and PS in our simulations

We tabulate the flux PDF, the PS and mean flux decrement (DA) obtained from the observational data used in this paper, and also the values corresponding to our best-fit model. The errors on flux PDF and flux PS are obtained from the simulated mock samples.

Mean flux decrement: Following the standard definition the mean flux decrement along a sightline is given by,

$$\mathrm{DA} = \langle 1 - \mathrm{e}^{-\tau} \rangle \tag{5.9}$$

where $\mathrm{e}^{-\tau}$ is normalized flux and angle brackets represent average along the wavelength. The total mean flux decrement for the N sightlines is given by,

$$\mathrm{DA}_{\rm sample} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{DA}_i \tag{5.10}$$

where DA_i is mean flux decrement along $i^{\rm th}$ sightline. The variance ($\sigma^2_{\rm sample}$) of the total mean flux decrement for the N sightlines is given by,

$$\sigma^2_{\rm sample} = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{DA}_i - \mathrm{DA}_{\rm sample})^2 \tag{5.11}$$

Table 5.10 shows the observed ($\mathrm{DA}_{\rm data}$) and best fit flux decrement ($\mathrm{DA}_{\rm model}$) from our simulation for different redshift bins (calculated using Eq. 5.10). The uncertainty in $\mathrm{DA}_{\rm data}$ is calculated using Eq. 5.11 whereas the uncertainty in $\mathrm{DA}_{\rm model}$ corresponds to 1σ uncertainty in best fit Γ_{12}. The uncertainty in $\mathrm{DA}_{\rm data}$ due to continuum fitting uncertainty is not accounted for. Note that for χ^2 analysis, we used flux PDF in the range $F \leq 0.85$.

5.8.2 Discrepancy in Γ_{12} at $z = 0.1$

In this section, we show that the discrepancy in our Γ_{12} and Kollmeier et al. (2014) is due to observational systematics (and not due to differences in simulation). To illustrate this, first we compare CDDF from VIPER, Danforth et al. (2016) and Kollmeier et al. (2014) (see Fig. 5.18). It is clear from Fig. 5.18 that above completeness limit (i.e., $\log \mathrm{N_{HI}} \geq 13.6$), the observed CDDF from Kollmeier et al. (2014) is consistently lower than that from VIPER and Danforth et al. (2016). Since $\Gamma_{\rm HI}$ scales inversely with CDDF, one would expect to get higher $\Gamma_{\rm HI}$ from Kollmeier et al. (2014) observational data.

To quantify the difference in Γ_{12}, we followed the procedure given by Kollmeier et al. (2014) and generated the Lyα forest spectra from our simulation. We used the Kollmeier et al. (2014) observational CDDF and constrain the $\Gamma_{\rm HI}$ using χ^2 minimization as discussed in section 5.5. The results are shown in Fig. 5.19. It is clear from the figure that if we assume observed CDDF along with its errorbar from Kollmeier et al. (2014) is correct then the Γ_{12} constraints from our simulations are 0.138 ± 0.01. This is ~ 3.9 times smaller than HM12 Γ_{12} ($= 0.0354$) at $z = 0.1$.

We now account for the differences in cosmological and astrophysical parameters between Kollmeier et al. (2014) and our simulation. Following the standard

Table 5.8: The observed and best fit Γ_{12} (see Table 5.7) flux PDF from our simulation for redshift bins. The errorbars on model flux PDF indicate diagonal terms of the covariance matrix. The covariance matrix can be available on request.

Spectra →	Lyα forest		Lyα forest		Lyα forest		Lyα + Lyβ forest	
	$z = 0.1125$	$z = 0.1125$	$z = 0.2$	$z = 0.2$	$z = 0.3$	$z = 0.3$	$z = 0.4$	$z = 0.4$
F	$P_{\rm data}$	$P_{\rm model} \pm dP_{\rm model}$	$P_{\rm data}$	$P_{\rm model} \pm dP_{\rm model}$	$P_{\rm data}$	$P_{\rm model} \pm dP_{\rm model}$	$P_{\rm data}$	$P_{\rm model} \pm dP_{\rm model}$
0	0.027	0.046 ± 0.011	0.022	0.049 ± 0.016	0.019	0.054 ± 0.016	0.031	0.074 ± 0.022
0.1	0.028	0.028 ± 0.004	0.025	0.030 ± 0.006	0.027	0.032 ± 0.006	0.034	0.042 ± 0.008
0.2	0.018	0.020 ± 0.003	0.022	0.021 ± 0.005	0.020	0.023 ± 0.004	0.030	0.031 ± 0.006
0.3	0.018	0.020 ± 0.003	0.023	0.022 ± 0.004	0.022	0.024 ± 0.004	0.032	0.033 ± 0.006
0.4	0.021	0.024 ± 0.003	0.023	0.026 ± 0.005	0.030	0.030 ± 0.005	0.035	0.039 ± 0.006
0.5	0.030	0.031 ± 0.003	0.029	0.033 ± 0.005	0.047	0.046 ± 0.008	0.052	0.056 ± 0.007
0.6	0.047	0.048 ± 0.004	0.053	0.048 ± 0.006	0.077	0.087 ± 0.017	0.100	0.103 ± 0.013
0.7	0.097	0.106 ± 0.010	0.092	0.094 ± 0.011	0.182	0.212 ± 0.038	0.229	0.271 ± 0.039
0.8	0.386	0.429 ± 0.037	0.319	0.362 ± 0.050	0.626	0.674 ± 0.073	0.745	0.828 ± 0.086
0.9	2.079	2.229 ± 0.057	2.093	2.250 ± 0.083	2.282	2.379 ± 0.052	2.334	2.458 ± 0.051
1	7.247	7.018 ± 0.094	7.300	7.066 ± 0.134	6.668	6.439 ± 0.112	6.375	6.065 ± 0.116

Table 5.9: The observed and best fit Γ_{12} (see Table 5.7) flux PS from our simulation for redshift bins. The errorbars on model flux PS indicate diagonal terms of the covariance matrix. The covariance matrix can be available on request. The wavenumber k is expressed in units of h cMpc^{-1}.

Spectra →	Lyα forest		Lyα forest		Lyα forest		Lyα + Lyβ forest	
	$z = 0.1125$	$z = 0.1125$	$z = 0.2$	$z = 0.2$	$z = 0.3$	$z = 0.3$	$z = 0.4$	$z = 0.4$
$\log k$	Data $\log P_F$	Model $\log P_F \pm$ dlog P_F	Data $\log P_F$	Model $\log P_F \pm$ dlog P_F	Data $\log P_F$	Model $\log P_F \pm$ dlog P_F	Data $\log P_F$	Model $\log P_F \pm$ dlog P_F
−0.601	−4.332	-4.338 ± 0.145	−4.340	-4.308 ± 0.149	−3.971	-3.995 ± 0.204	−3.706	-3.652 ± 0.162
−0.425	−4.297	-4.360 ± 0.142	−4.312	-4.339 ± 0.153	−4.107	-4.025 ± 0.198	−3.717	-3.685 ± 0.158
−0.300	−4.337	-4.394 ± 0.143	−4.361	-4.364 ± 0.151	−4.082	-4.045 ± 0.189	−3.741	-3.716 ± 0.158
−0.203	−4.391	-4.418 ± 0.140	−4.369	-4.395 ± 0.144	−4.102	-4.082 ± 0.187	−3.813	-3.749 ± 0.155
−0.124	−4.408	-4.459 ± 0.133	−4.433	-4.436 ± 0.145	−4.150	-4.119 ± 0.187	−3.718	-3.799 ± 0.151
−0.057	−4.421	-4.490 ± 0.132	−4.413	-4.466 ± 0.141	−4.120	-4.151 ± 0.180	−3.713	-3.835 ± 0.148
0.047	−4.477	-4.543 ± 0.123	−4.446	-4.521 ± 0.131	−4.234	-4.212 ± 0.169	−3.857	-3.894 ± 0.135
0.142	−4.547	-4.632 ± 0.119	−4.518	-4.616 ± 0.123	−4.214	-4.302 ± 0.157	−3.915	-4.002 ± 0.128
0.236	−4.589	-4.732 ± 0.114	−4.633	-4.724 ± 0.120	−4.323	-4.411 ± 0.146	−4.012	-4.127 ± 0.117
0.331	−4.675	-4.843 ± 0.108	−4.745	-4.842 ± 0.111	−4.469	-4.521 ± 0.130	−4.207	-4.257 ± 0.110
0.425	−4.839	-4.957 ± 0.100	−4.897	-4.963 ± 0.101	−4.560	-4.628 ± 0.123	−4.309	-4.386 ± 0.099
0.520	−4.977	-5.072 ± 0.092	−5.064	-5.087 ± 0.095	−4.756	-4.741 ± 0.109	−4.509	-4.517 ± 0.089
0.615	−5.154	-5.202 ± 0.083	−5.254	-5.224 ± 0.085	−4.864	-4.863 ± 0.098	−4.658	-4.657 ± 0.081
0.709	−5.299	-5.333 ± 0.075	−5.401	-5.364 ± 0.077	−5.013	-4.976 ± 0.089	−4.794	-4.791 ± 0.071
0.804	−5.431	-5.456 ± 0.065	−5.546	-5.496 ± 0.067	−5.161	-5.073 ± 0.077	−4.936	-4.906 ± 0.063

Table 5.10: The observed ($\mathrm{DA_{data}}$) and best fit flux decrement ($\mathrm{DA_{model}}$) from our simulation for different redshift bins.

Redshift bin ⇒	I	II	III	IV
Type of simulated spectra ⇒	Lyα forest	Lyα forest	Lyα forest	Lyα + Lyβ forest[a]
Best Fit Γ_{12}	0.066 ± 0.015	0.100 ± 0.021	0.145 ± 0.037	0.210 ± 0.052
$\mathrm{DA_{data}}$[b]	0.021 ± 0.002	0.024 ± 0.002	0.025 ± 0.002	0.031 ± 0.003
$\mathrm{DA_{model}}$[c]	0.028 ± 0.005	0.030 ± 0.004	0.032 ± 0.006	0.033 ± 0.006

[a] Simulated Lyα forest at $z = 0.35$ to 0.45 is contaminated by Lyβ forest in the same wavelength range. The Lyβ forest is generated from simulation box at $z = 0.6$.

[b] The uncertainty in $\mathrm{DA_{data}}$ is calculated using Eq. 5.11. However, the uncertainty in $\mathrm{DA_{data}}$ due to continuum fitting uncertainty is not accounted for. Note that for χ^2 analysis, we used flux PDF in the range $F \leq 0.85$.

[c] The uncertainty in $\mathrm{DA_{model}}$ corresponds to 1σ uncertainty in best fit Γ_{12}.

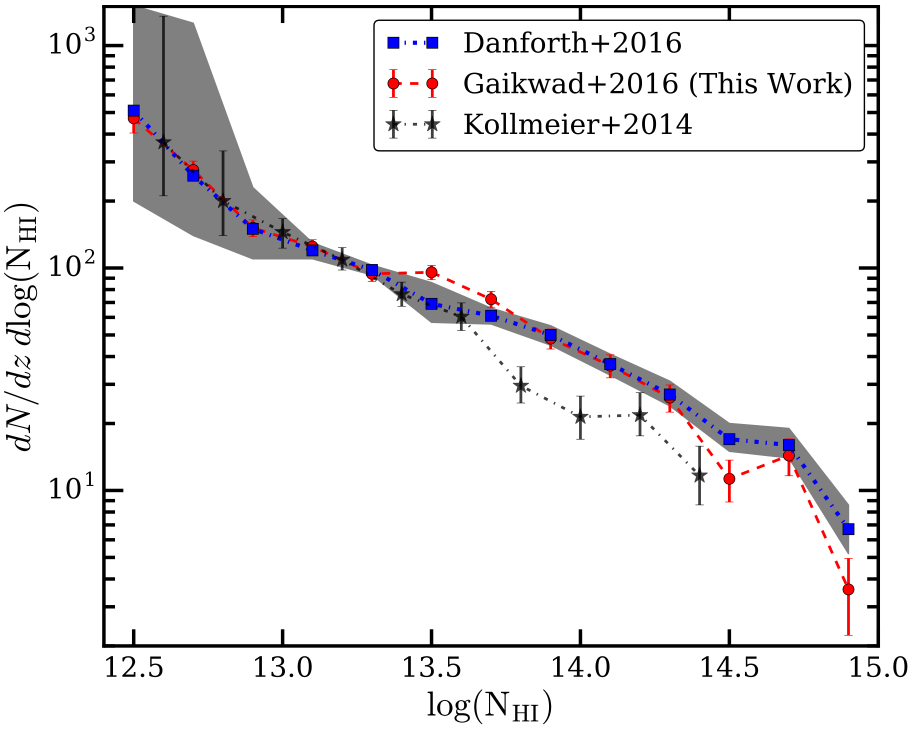

Figure 5.18: Comparison of observed CDDF obtained using VIPER (red circles), from Danforth et al. (2016, blue squares) and Kollmeier et al. (2014, black stars) at $0.075 \leq z \leq 0.5$. At $\log \mathrm{N_{HI}} \geq 13.6$ (which is also completeness limit of the sample), the CDDF from Kollmeier et al. (2014) is consistently lower than our CDDF and Danforth et al. (2016)

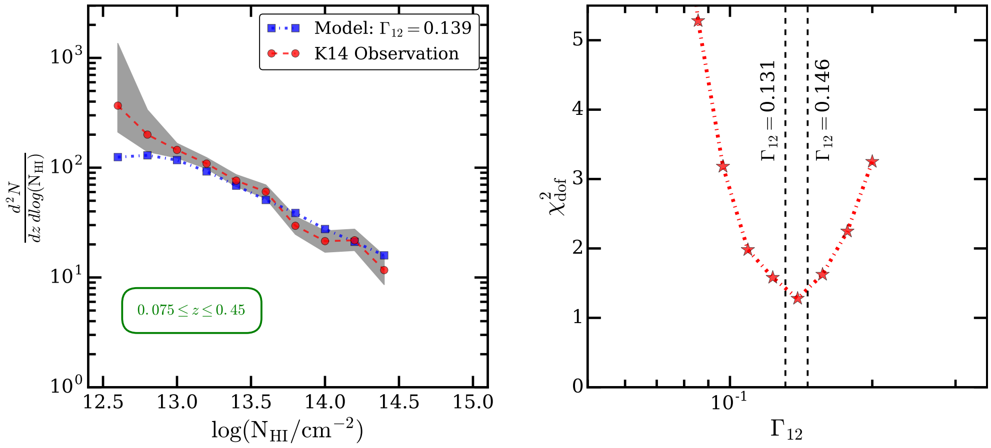

Figure 5.19: Constraints on Γ_{12} using observed K14 CDDF. Left-hand panel shows the best fit model CDDF (blue squares) and observed K14 CDDF (red circles) in the redshift range $0.075 \leq z \leq 0.45$. The shaded region corresponds to 1σ error on observed CDDF. The model CDDF is calculated from $z = 0.1$ simulation box using 5000 Ly-α forest spectra. Right-hand panel shows variation of $\chi^2_{\rm dof}$ with Γ_{12}. The minimum value $\chi^2_{\rm dof}$ (~ 1.38) occurs at $\Gamma_{12} = 0.138$. The black dashed vertical lines represent 1σ statistical uncertainty in Γ_{12}.

definition the flux decrement is given by $D_A = 1 - e^{-\tau_{\rm eff}}$. For a small value of $\tau_{\rm eff}$ we can write,

$$\begin{aligned} D_A &\approx 1 - (1 - \tau_{\rm eff}) \\ D_A &\approx \tau_{\rm eff} \end{aligned} \tag{5.12}$$

As given in Kollmeier et al. (2014, K14), $D_A = 0.05$ for HM12 background at $z = 0.1$.

In our simulation (G16), $D_A = 0.04$ for HM12 background at $z = 0.1$.

According to Fluctuating Gunn-Peterson Approximation (FGPA) the Ly-α optical depth is given as,

$$\tau \propto \frac{(f_{\rm d}\, \Omega_b\, h^2)^2}{\Gamma_{12}\, \Omega_m^{0.5}} \tag{5.13}$$

where $f_{\rm d}$ is fraction of the baryons in the cold IGM. We noticed that Kollmeier et al. (2014) used simulation with momentum-driven wind formalism ('vzw' model). The fraction of baryons in diffused phase ($f_{\rm d}$) for this model is $\sim 41\%$ whereas in our case $f_{\rm d} \sim 34\%$ (from Table. 2). The scaling relation between Γ_{12} of K14 and G16 is given as,

$$\begin{aligned}
\frac{[\Gamma_{12}]_{\rm G16}}{[\Gamma_{12}]_{\rm K14}} &= \left(\frac{[f_{\rm d}]_{\rm G16}}{[f_{\rm d}]_{\rm K14}}\right)^{2} \times \left(\frac{[\Omega_b]_{\rm G16}}{[\Omega_b]_{\rm K14}}\right)^{2} \times \left(\frac{[h^2]_{\rm G16}}{[h^2]_{\rm K14}}\right)^{2} \times \left(\frac{[\Omega_m]_{\rm K14}}{[\Omega_m]_{\rm G16}}\right)^{0.5} \times \frac{[\tau_{\rm eff}]_{\rm K14}}{[\tau_{\rm eff}]_{\rm G16}} \\
\frac{[\Gamma_{12}]_{\rm G16}}{[\Gamma_{12}]_{\rm K14}} &= \left(\frac{0.34}{0.41}\right)^{2} \times \left(\frac{0.0486}{0.0440}\right)^{2} \times \left(\frac{0.674^2}{0.700^2}\right)^{2} \times \left(\frac{0.25}{0.31}\right)^{0.5} \times \frac{0.05}{0.04} \\
\frac{[\Gamma_{12}]_{\rm G16}}{[\Gamma_{12}]_{\rm K14}} &= 0.6877 \times 1.2200 \times 0.8595 \times 0.8980 \times 1.2500 \\
\frac{[\Gamma_{12}]_{\rm G16}}{[\Gamma_{12}]_{\rm K14}} &= 0.8095
\end{aligned} \tag{5.14}$$

Expected Γ_{12} from K14 observation data

$$\begin{aligned}
\Big[\Gamma_{12}]_{\rm K14} &= 1.2353 \times [\Gamma_{12}]_{\rm G16} \\
[\Gamma_{12}]_{\rm K14} &= 1.2353 \times (0.138 \pm 0.007) \\
[\Gamma_{12}]_{\rm K14} &= 0.170 \pm 0.009
\end{aligned}$$

This is ~ 4.8 times HM12 and is consistent within 1σ with Γ_{12} constraints quoted by K14 (0.174) at $z = 0.1$. Thus the discrepancy in our Γ_{12} and Kollmeier et al. (2014) is due to observational systematics (and not due to differences in simulation).

6 | Conclusions and Future outlook

6.1 Conclusions

The primary objective of the thesis is: (i) To develop an efficient, flexible and sufficiently accurate method to simulate the IGM in a post-processing step of the cosmological N-body hydrodynamical simulation GADGET-2 and (ii) to apply these models to constrain astrophysical parameters from observations using various statistics. For this purpose, we have developed following post-processing tools that captures the effect of various astrophysical processes important in IGM studies:

(i) A module that accounts for the effect of thermal, ionization evolution and dynamic pressure of the gas.

(ii) A tool to generate the Lyα forest by shooting random sightlines through the simulation box.

(iii) A module that automatically decomposes the Lyα forest in multi-component Voigt profiles.

(iv) Different codes to derive the various Lyα flux statistics such as flux probability distribution function (PDF), flux power spectrum (PS), wavelet statistics (Lidz et al. 2010) and curvature statistics (Becker et al. 2011).

Using these tools, we model various observed properties of the IGM and compare with observations to constrain astrophysical parameters with appropriate errorbars.

6.1.1 Hydrodynamical simulations of Lyα forest

The Lyα forest is one of the most sensitive tools to study the physics of IGM. The observed properties of the Lyα forest are sensitive to fluctuations in the cosmic density fields ($\Delta \leq 10$ at $z \sim 2-4$ and $\Delta \leq 50$ at $z \leq 0.5$), velocity fields and physical conditions such as the temperature, turbulence and ionizing radiation prevailing in the IGM. We use the smoothed particle hydrodynamic (SPH) code GADGET-2 (Springel 2005) to generate the density and velocity field in a periodic box. The modelling of the Lyα optical depth involves shooting of random sightlines through simulation box and the calculation of three fields, namely, (i) the overdensity Δ, (ii) the peculiar velocity v and (iii) the temperature T along these sighlines. We found that the temperature obtained from GADGET-2 is not realistic because radiative heating and cooling processes are not incorporated in GADGET-2.

In this thesis, we have developed a "Code for Ionization and Temperature Evolution" (CITE) to model the thermal and ionization evolution of particles in the post-processing step of GADGET-2. In CITE, we solve the equilibrium (or non-equilibrium, if desired) ionization evolution equation for a given UVB to calculate various ionization fractions. For a given temperature and density of each SPH particle, we calculate cooling and heating rate in presence of surrounding radiation field set by UVB. We add (or subtract) the temperature corresponding to these processes in temperature evolution equation. We found that the resulting temperature from GADGET-2 + CITE is realistic and obeys power law $T-\Delta$ relation consistent with that from the analytic calculation and full hydrodynamic simulations. CITE offers many advantages for the IGM studies as given below:

(1) Because CITE is based on post-processing the GADGET-2 output, the method is computationally less expensive.

(2) CITE allows us to explore large thermal history parameter space without performing the full SPH simulation from high-z.

(3) CITE provides flexibility to (i) simulate the non-equilibrium evolution of thermal and ionization state of the gas, (ii) incorporate the radiative cooling for

a wide range of metallicities and (iii) include heating due to non-standard sources like cosmic rays and high energy γ-rays from Blazars.

Despite offering the flexibility, the obvious shortcoming of CITE is that the diffuse gas is evolved dynamically at effectively zero pressure (because of its low temperature), rather than the pressure it would have if it were at $T \sim 10^4$ K typical of photoionized gas. Thus the dynamical impact of the diffuse IGM pressure (especially at small scales) is not modelled self-consistently in CITE. However, we found that the effect of dynamic pressure is not important for moderate to low resolution (gas particle mass $\delta m \sim 1.26 \times 10^7 \, h^{-1} \, \mathrm{M}_\odot$, pixel size $\delta x \sim 48.8 h^{-1}$ ckpc) Lyα forest studies typically achieved with HST-COS at low-z ($z < 0.5$). On the other hand, for studying the high-z ($2 \leq z \leq 4$) Lyα forest one usually uses higher resolution echelle data. When we use appropriate high resolution (gas particle mass $\delta m \sim 1.01 \times 10^5 \, h^{-1} \, \mathrm{M}_\odot$, pixel size $\delta x \sim 9.77$ ckpc) simulation boxes, we notice that the density (Δ) and velocity (v) fields are smoother for GADGET-3[1] as compared to those from GADGET-2 (known as "pressure smoothing"). This is because the temperature of the SPH particles in GADGET-2 is not calculated self-consistently, the unshocked gas is effectively evolved at zero pressure in GADGET-2, and the local Jeans length of the particles, responsible for pressure smoothing in GADGET-3, is comparable to the resolution of the simulation box. We show that the effect of dynamical pressure can be captured by running GADGET-2 at an elevated temperature floor $\sim 10^4$ K, post-processing with CITE, and smoothing (in 3 dimensions) the density and velocity fields over a local Jeans scale (over which pressure smoothing effects are important) with a modified SPH kernel.

We further develop a module for "Generating Lyman Alpha forest Spectra in Simulation" (GLASS). Using GLASS, we calculated the Lyα transmitted flux that has signal-to-noise ratio (SNR), spectral resolution and line spread function (LSF) effects similar to the observational data. We show that the line-of-sight density, velocity, temperature field and Lyα transmitted flux is remarkably similar to those obtained from self-consistent simulations like GADGET-3. Our method to

---

[1] Radiative heating and cooling is incorporated internally in GADGET-3

simulate the Lyα forest is computationally less expensive, flexible and accurate to within 10 percent.

6.1.2 Voigt decomposition of Lyα forest

The observations of the Lyα forest have regularly been used to constrain cosmological and astrophysical parameters by comparing different statistics of the Lyα forest derived from observed spectra with those from the simulated ones. An important class of statistics requires to decompose the Lyα forest into multiple Voigt profiles. The line width distribution function calculated from Voigt profile fitting is sensitive to the thermal history and the energy injected by various astrophysical processes in the form of heat and turbulent motions in the IGM (Davé et al. 2001; Schaye et al. 1999, 2000). Similarly, the column density distribution function (CDDF) calculated from Voigt profile decomposition is sensitive to $\Gamma_{\rm HI}$ (Gurvich et al. 2017; Kollmeier et al. 2014; Shull et al. 2015) and cosmological parameters (Shull et al. 2012). While statistics based on parameters obtained using Voigt profile fitting are useful in deriving thermal history and equation of state of the IGM, the Voigt profile decomposition is usually subjective, laborious and time consuming process. Therefore, a large parameter space exploration in simulations is usually difficult.

We have developed a parallel processing module "VoIgt profile Parameter Estimation Routine" (VIPER) to fit the Lyα forest with multiple Voigt profiles automatically. In VIPER, the blended and saturated features are fitted simultaneously with multi-component Voigt profiles. An objective criteria based on information theory is used to find the number of Voigt profiles needed to describe the Lyα forest. Using VIPER, we fitted 82 HST-COS Lyα forest spectra using VIPER and compiled a Lyα line catalog called "VIPER line catalog". We show that the Voigt profiles fitted to observed spectra using VIPER match well, in terms of number of components and the values of fitted parameter along with the errorbar, with those fitted using semi-numerical method in Danforth et al. (2016). The median b parameter from VIPER (32.9 ± 20.8 km s^{-1}) is consistent with that from Danforth et al. (2016, 33.9 ± 18.3 km s^{-1}). Also, the median $\log \mathrm{N_{HI}}$ from VIPER (13.39 ± 0.61) is in good agreement with that from Danforth et al. (2016, 13.38 ± 0.63). We calculate the appropriate redshift path length $\Delta z(\mathrm{N_{HI}})$ and the sensitivity curve from the HST-COS data. Subsequently, we calculate the CDDF after accounting for the incompleteness of the sample. We show that our calculated CDDF in the redshift range ($0.075 \leq z \leq 0.45$) is consistent (KS test *p*-value is 0.83) with that of Danforth et al. (2016) CDDF in the redshift range ($0 \leq z \leq 0.75$). The parallel and automated nature of VIPER allows us to simultaneously fit large number of simulated spectra and to explore a wide parameter space efficiently.

6.1.3 Validation of our simulations

We validated our method of evolving thermal and ionization state of the IGM using CITE with other simulations in the past and full hydrodynamic simulations like GADGET-3. For low resolution (similar to that achieved in HST-COS data) simulation, we show the consistency of our method with other simulations in the literature by comparing with three metrics:

(i) thermal history parameters: our simulation predicts $T_0 \sim 5000$ K and $\gamma \sim 1.6$ in the redshift range $z = 0.1$ to 0.45. These values are shown to be insensitive to our choice of T_0 and γ at an initial redshift, $z_1 = 2.1$;

(ii) distribution of baryons in phase diagram at $z = 0$: We find ∼ 34 per cent of baryons are in diffuse phase, ∼ 29 per cent in warm hot intergalactic medium (WHIM), ∼ 18 per cent in hot halo and ∼ 19 per cent in condensed phase and

(iii) the correlation between baryon overdensity Δ vs H I column density, $\mathrm{N_{HI}}$, in the redshift range $0.2 < z < 0.3$: we find $\Delta = 34.8 \pm 5.9 (\mathrm{N_{HI}}/10^{14})^{0.770 \pm 0.022}$.

We show that all these predictions compare well with those of low-z simulations in the literature (Davé et al. 2010; Smith et al. 2011) that include different feedback processes at varied levels.

On the other hand, for high resolution (typically achieved in echelle data) simulations (at $2 \leq z \leq 4$), the dynamical evolution of SPH particles at finite pressure is an important effect. It is well known that the pressure smoothing in

GADGET-3 is not only decided by the instantaneous density and temperature of the particles but also to some extent by the thermal history (Gnedin and Hui 1998; Kulkarni et al. 2015). To illustrate this, we perform three high resolution simulations (gas particle mass $\delta m \sim 1.01 \times 10^5\ h^{-1}\ \mathrm{M}_\odot$, pixel size $\delta x \sim 9.77$ ckpc) with same initial conditions (i) G2-LTF: GADGET-2 with low temperature ($T \sim 100$ K) floor in which local Jeans length is decided by instantaneous density and temperature and (ii) G2-HTF: GADGET-2 with high temperature ($T \sim 10^4$ K) floor in which even the unshocked gas is evolved at a pressure appropriate for a photoionized gas at $T = 10^4$ K and (iii) GADGET-3: a reference model for comparison with G2-LTF and G2-HTF model.

First, we obtain the evolution of thermal history parameters T_0 and γ by estimating the temperature of the SPH particles from CITE. The T_0 and γ evolution from our model is in very good agreement with that from GADGET-3. CITE also provides us with enough flexibility to solve the non-equilibrium ionization evolution equation. The T_0 and γ evolution for non-equilibrium case is considerably different (T_0 is larger by ~ 60 percent and γ is smaller by 15 percent at $z = 3.7$) than that for equilibrium case. Using our module GLASS, we generate the Lyα forest spectra by shooting random sightlines through simulation box in all the 3 models. The resulting Lyα forest spectra along sightline are remarkably similar in the G2-HTF and GADGET-3 methods. We compare the G2-LTF and G2-HTF with the GADGET-3 model using 8 different statistics, namely: (i) 1D density field PS, (ii) flux PDF, (iii) flux PS, (iv) wavelet PDF (Lidz et al. 2010), (v) curvature PDF (Becker et al. 2011), (vi) CDDF, (vii) linewidth (b) distribution and (viii) b vs $\log \mathrm{N}_{\mathrm{HI}}$ correlation, at four different redshift $z = 2.5, 3.0, 3.5$ and 4.0. We have developed a statistical module to calculate the 8 statistics mentioned above. Treating the GADGET-3 model as the reference, we demonstrate that the Γ_{HI} can be recovered, using flux PDF and flux PS statistics, within 1σ statistical uncertainty using the G2-HTF model. We find that the G2-HTF model is in general very good agreement (within 1σ) with GADGET-3 model at all redshifts. Using enhanced Haardt and Madau (2012) photo-heating rates, we obtain a thermal history such that T_0 is increased by a factor of ~ 2. We show that our method for such significantly different thermal history is consistent (in 1σ) with GADGET-3 simulation.

6.1.4 Parameter estimation: H I photoionization rate at low-z

The highly ionized state of the IGM at $z < 0.5$ is maintained by UVB radiation ($\lambda \leq 912$ Å). The accurate characterization UVB is important for modeling the thermal and ionization state of the IGM (Becker and Bolton 2013; Becker et al. 2011; Boera et al. 2014; Lidz et al. 2010), measuring the baryon content in the IGM (Shull et al. 2012), and deriving metal abundances in the IGM using ionization corrections (Carswell et al. 2002; Peeples et al. 2014; Shull et al. 2014; Songaila 2001). This UVB is contributed by radiation from blackhole accretion in QSOs and stellar light escaping from galaxies. The stellar contribution to the UVB depends crucially on the average fraction of ionizing photons escaping the galaxies known as the escape fraction ($f_{\rm esc}$). The theoretical modelling of $f_{\rm esc}$ is difficult as it depends on various physical factors such as the galaxy mass, morphology, composition of the interstellar medium (ISM), spatial distribution of gas and supernova rates (Cen and Kimm 2015; Ricotti and Shull 2000; Roy et al. 2015). One way of constraining $f_{\rm esc}$ is by using measured $\Gamma_{\rm HI}$ (Inoue et al. 2006; Khaire et al. 2016).

We measure $\Gamma_{\rm HI}$ at $z \leq 0.45$ from a sample of 82 QSO spectra obtained with HST-COS and hydrodynamical simulations using GADGET-2 post-processed with CITE and GLASS. For a fair comparison, we mimic the simulated Lyα forest as close to observations as possible in terms of the SNR, resolution and LSF. The spectra generated using our method are remarkably similar to the observed spectra. We use three statistics (i.e., flux PDF, flux PS and CDDF) and χ^2 minimization using appropriate covariance matrices to compare the observations with the model predictions. We measured $\Gamma_{\rm HI}$ in four different redshift bins (of $\Delta z = 0.1$) centered at $z = 0.1125, 0.2, 0.3, 0.4$ using constraints from these statistics mentioned above. We estimated the associated errors by varying thermal history parameters, cosmological parameters and continuum fitted to the observed spectrum. Due to limited wavelength range covered in the HST-COS spectrum used in this study, the $\Gamma_{\rm HI}$ measurement for the highest redshift bin (i.e $z = 0.4$) is likely to be affected by the contamination of Lyβ forest absorption from higher-z. We contaminated our simulated Lyα forest at $z = 0.4$ by Lyβ forest from $z = 0.6$

and corrected for the effect of Lyβ contamination in our $\Gamma_{\rm HI}$ measurement for this z bin. The measured $\Gamma_{\rm HI}$ values at redshift bins $z = 0.1125, 0.2, 0.3, 0.4$ are $0.066 \pm 0.015,\ 0.100 \pm 0.021,\ 0.145 \pm 0.037,\ 0.210 \pm 0.052$, respectively.

We find that the b parameters of Voigt profile components from simulations are typically underestimated as compared to observations. This difference can be rectified by including the Gaussian distributed line width parameter $b_{\rm turb}$ ($\mu = 20$ $\rm km\,s^{-1}$and $\sigma = 10$ $\rm km\,s^{-1}$) at each pixel in the simulation (Muzahid et al. 2012; Tripp et al. 2008). The resulting line width distribution from simulations matches roughly with observed line width distribution, scatter and lower envelope of the b vs $\log \rm N_{HI}$ distribution. However, the CDDF has little effect of additional $b_{\rm turb}$ ($<$ 7 percent) and the $\Gamma_{\rm HI}$ constraints are mildly affected ($<$ 9 percent). On the other hand if we consider additional heating effect (Viel et al. 2017) for the excess broadening then the $\Gamma_{\rm HI}$ obtained will be slightly reduced (roughly scale as $T^{-0.7}$).

Our final quoted errors in the $\Gamma_{\rm HI}$ measurements include possible uncertainties coming from the statistical uncertainty ($\sim$ 14 per cent), cosmic variance ($\sim$ 3 per cent), cosmological parameters uncertainty ($\sim$ 10 per cent) and continuum uncertainty (systematic uncertainty $\sim$ 7 per cent). Uncertainty in $\Gamma_{\rm HI}$ due to uncertainty in thermal history parameters, over the range considered here, is small and within statistical uncertainty. As expected based on UVB models, even in the small redshift range covered in our study the measured $\Gamma_{\rm HI}$ shows a rapid evolution with z. We fit the redshift evolution of $\Gamma_{12} \equiv \Gamma_{\rm HI}/10^{-12}$ $\rm s^{-1}$ as $\Gamma_{12} = 0.040 \pm 0.001\ (1+z)^{4.99\pm0.12}$ at $0.075 \leq z \leq 0.45$. The $\Gamma_{\rm HI}(z)$ obtained here are consistent with the measurement of Shull et al. (2015) however our $\Gamma_{\rm HI}$ measurement at $z = 0.1$ is factor $\sim$ 2.7 smaller than Kollmeier et al. (2014).

The $\Gamma_{\rm HI}$ measurement at any z_1 depends on the emissivities of the ionizing sources at $z \geq z_1$ and Lyman continuum opacity of the IGM. We considered the updated emissivities of QSOs and galaxies (with $f_{\rm esc}$ as a free parameter) and two different H I column density distribution as a function of z obtained by Haardt and Madau (2012) and Inoue et al. (2014) and obtained $\Gamma_{\rm HI}$ using Khaire and Srianand (2015a) UVB code. We find that for, both H I distributions, our derived $\Gamma_{\rm HI}(z)$ is consistent with being contributed only by QSOs. This is true even if we allow for variations in the UV spectral index of QSOs. We also find the maximum

3σ upper limit on $f_{\rm esc}$ at $z < 2$, allowing for uncertainty in far UV spectral index and cloud distribution $f(N_{\rm HI}, z)$ of Inoue et al. (2014), is 0.008. This is consistent with 3σ upper limits on average $f_{\rm esc}$ (i.e ≤ 0.02) obtained by stacking samples of galaxies probing average galaxy mass $M \geq 10^{9.3} M_{\odot}$.

Our measurements suggest that the contribution of low mass galaxies to average $f_{\rm esc}$ will also be small. *Our study confirms that there is no crisis at low redshift in accounting for the observed Lyman continuum photons using standard known luminous astronomical sources as claimed by* Kollmeier et al. (2014). Thus our $\Gamma_{\rm HI}(z)$ measurement can in turn be used to place a strong constraint on the contributions of decaying dark matter to the low-z UVB. For more details we refer the reader to Gaikwad et al. (2017b,c).

6.2 Future Outlook

With the rapid advancement in observations and simulations, IGM science is entering a new phase that will provide new opportunities for other areas of astrophysics and cosmology to exploit the IGM as a tool for testing models of interest. The upcoming surveys with ground based telescopes like Thirty Meter Telescope[1] (TMT), Giant Magellan Telescope[2] (GMT), Giant Segmented Mirror Telescope[3] (GSMT), Extremely Large Telescope[4] (ELT), Sloan Digital Sky Survey[5] (SDSS) and space telescope like James Webb Space Telescope[6] (JWST), promises to settle a number of outstanding questions in IGM science such as

(i) **Ionization of IGM:** The opacity of IGM rapidly increases with redshift ($z > 6$) and it becomes difficult to probe the IGM properties with Lyα (or even Lyβ) absorption lines. The ongoing and upcoming radio experiments with Square Kilometer Array[7] (SKA), Low-Frequency Array[8] (LOFAR),

---

[1] `http://www.tmt.sorg`

[2] `http://www.gmto.org`

[3] `http://www.gsmt.noao.edu`

[4] `http://www.eso.org/projects/e-elt/`

[5] `http://www.sdss.org`

[6] `http://www.jwst.nasa.gov/`

[7] `http://skatelescope.org/`

[8] `http://www.lofar.org/`

Murchison Widefield Array[1] (MWA), Precision Array for Probing the Epoch of Reionization[2] (PAPER) and Giant Meterwave Radio Telescope[3] (GMRT), are aiming to detect the redshifted 21cm emission line of hydrogen around $z \sim 6-15$ which will help to understand the reionization process. Powerful radio telescopes and interferometers such as SKA should be able to trace the topology and growth of the ionized regions around the first sources by mapping the redshifted 21cm emission line of hydrogen in a certain region of sky (Datta et al. 2016). However, although these observation will improve our knowledge of epoch of cosmic reionization, they may not detect the sources responsible it. Thus our physical understanding of the process will crucially rely on observations with telescopes like JWST and TMT. JWST should be able to detect the brightest sources lying within each ionized bubble (Gardner et al. 2006). On the other hand TMT will be able to detect and study fainter (by an order of magnitude) objects (Skidmore et al. 2015) . The high resolution and high SNR absorption spectra from the fainter objects will allow one to study, (i) the properties and influence of earlier sources on the IGM. (ii) the thermal and ionization state of the IGM during the epoch of He II reionization ($2 \leq z \leq 4$) (iii) the nature of ionizing sources and intensity of ionizing background during epoch of He II reionization and (iv) interaction of galaxies with environment in the form of outflows or inflows.

(ii) **Galaxy-IGM connection:** Structure formation models predict the interaction between galaxies and IGM over the cosmic time. The sensitivity of 30 m class telescopes (Skidmore et al. 2015) allows one to obtain spectra faint galaxies as background source. Since number density of galaxies is large, it will be possible to sample the region densely enough in the longitudinal direction to study the absorption systems as a function of impact parameter from galaxies. The metal line systems frequently seen in QSO absorption spectra are also suitable for studying the galaxy-IGM interaction. The ionization potential of the species found in typical QSO spectra

---

[1] `http://www.mwatelescope.org/`

[2] `http://eor.berkeley.edu/`

[3] `http://gmrt.ncra.tifr.res.in/`

are quite different e.g., 15 eV for Mg II, 65 eV for C IV and 138 eV for O VI. For a photoionization models (since densities are low collisional ionization is assumed to be sub-dominant), Mg II systems traces the IGM gas at $T \sim 10^4$ K (similar to H I gas) on the other hand C IV and O VI systems are likely associated with hotter gas at $T \sim 10^5 - 10^6$ K. Hence metal absorption line systems are useful to probe the different phases of the IGM. Furthermore the metal enriched gas in the ISM of galaxies can be ejected into the large-scale environment by galactic winds, stellar and AGN feedback processes thereby polluting the IGM. Thus the study of metal lines will also allow one to quantify the effect of galactic winds and various feedback processes on IGM (Meiksin 2009).

(iii) **IGM tomography:** With upcoming 30 m class telescopes (Skidmore et al. 2015), it would be possible to obtain spectra from faint QSO and faint galaxies as background source. Since number density of galaxies is much more than QSO, it will also be possible to sample the region densely enough in the longitudinal direction to study the absorption systems in the global context. This will allow one to construct the 3D H I distribution map (known as IGM tomography) from Lyα forest. The Lyα forest tomography will allow one (i) to study galaxy properties as a function of environment, (ii) to quantify the IGM overdensities associated with protoclusters (Lee et al. 2016), (iii) to study alignment between galaxy spin and the direction of filaments or sheets which is predicted by galaxy formation models (Krolewski et al. 2017) and (iv) to improve the photometric redshifts of the galaxies (Schmittfull and White 2016).

(iv) **Baryon Acoustic Oscillations (BAO):** BAO are the oscillations in primordial plasma before the decoupling of baryons from radiation. The BAO produces the distinct peak in the correlation function or oscillation pattern in power spectrum (McDonald and Eisenstein 2007). Till now, galaxies are used as a tracers for BAO. However, the current measurements of the BAO with galaxies are restricted to redshifts $z \leq 1$. Many dark energy models usually introduce an early dark energy component to circumvent tuning problems. Such a component can only be measured by a technique sensitive to the expansion history at high redshift. The Lyα forest (i) traces mildly non-linear densities of H I in the IGM, (ii) covers large redshift path length, and (iii) traces the dark matter density field on large scales. This makes Lyα forest a suitable probe for measuring BAO and constraining expansion history of the universe at high redshift (Busca et al. 2013; McDonald and Eisenstein 2007; Slosar et al. 2013).

(v) **Parameter estimation:** The free-streaming of neutrinos or presence of warm dark matter particles in cosmological models results in a scale-dependent suppression of power spectrum of the total matter distribution at scales probed by Ly-α forest (Viel et al. 2010). The effect of neutrinos or warm dark matter can be quantified by measuring the flux PDF and flux PS in Lyα forest. The upcoming surveys using instruments like SDSS will allow one to put tight constraints on the mass of neutrino or warm dark matter particles. This technique is complementary to the cosmic microwave background data or other large scale structure probes. Note that the above effects are degenerate with thermal history parameters T_0 and γ. Hence one need to probe the large parameter space and marginalize over the nuisance parameters T_0, γ. The tools developed in this thesis will allow one to efficiently explore $T_0 - \gamma$ space.

In order to synchronize with flow of observational data from these surveys, IGM simulations needs to be efficient, flexible and sufficiently accurate to probe large parameter space. *The tools and techniques developed in this thesis provide a starting point to efficiently simulate the IGM.* With large amount of observational data together with efficient simulations, the IGM science will continue to flourish for many years to come.

Bibliography

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions.* 85

Adams, J. J., Uson, J. M., Hill, G. J., and MacQueen, P. J. (2011). A New z = 0 Metagalactic Ultraviolet Background Limit. ApJ, 728:107. 138

Agertz, O., Moore, B., Stadel, J., Potter, D., Miniati, F., Read, J., Mayer, L., Gawryszczak, A., Kravtsov, A., Nordlund, Å., Pearce, F., Quilis, V., Rudd, D., Springel, V., Stone, J., Tasker, E., Teyssier, R., Wadsley, J., and Walder, R. (2007). Fundamental differences between SPH and grid methods. MNRAS, 380:963–978. 33

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19:716–723. 93

Almgren, A. S., Bell, J. B., Lijewski, M. J., Lukić, Z., and Van Andel, E. (2013). Nyx: A Massively Parallel AMR Code for Computational Cosmology. ApJ, 765:39. 34

Arinyo-i-Prats, A., Miralda-Escudé, J., Viel, M., and Cen, R. (2015). The non-linear power spectrum of the Lyman alpha forest. *JCAP*, 12:017. 123, 149

Armstrong, B. (1967). Spectrum line profiles: The voigt unction. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 7(1):61 – 88. 91

Bainbridge, M. B. and Webb, J. K. (2016). Artificial intelligence applied to the automatic analysis of absorption spectra. Objective measurement of the fine structure constant. *ArXiv e-prints.* 88

Bajtlik, S., Duncan, R. C., and Ostriker, J. P. (1988). Quasar ionization of Lyman-alpha clouds - The proximity effect, a probe of the ultraviolet background at high redshift. ApJ, 327:570–583. 138

Barkana, R. and Loeb, A. (2001). In the beginning: the first sources of light and the reionization of the universe. Phys. Rep., 349:125–238. 3

Becker, G. D. and Bolton, J. S. (2013). New measurements of the ionizing ultraviolet background over $2 < z < 5$ and implications for hydrogen reionization. MNRAS, 436:1023–1039. xxiii, lv, 25, 36, 42, 53, 87, 174, 195

Becker, G. D., Bolton, J. S., Haehnelt, M. G., and Sargent, W. L. W. (2011). Detection of extended He II reionization in the temperature evolution of the intergalactic medium. MNRAS, 410:1096–1112. xv, xix, xxii, xxiii, 3, 25, 36, 42, 53, 60, 87, 109, 128, 133, 161, 189, 194, 195

Becker, R. H., Fan, X., White, R. L., Strauss, M. A., Narayanan, V. K., Lupton, R. H., Gunn, J. E., Annis, J., Bahcall, N. A., Brinkmann, J., Connolly, A. J., Csabai, I., Czarapata, P. C., Doi, M., Heckman, T. M., Hennessy, G. S., Ivezić, Ž., Knapp, G. R., Lamb, D. Q., McKay, T. A., Munn, J. A., Nash, T., Nichol, R., Pier, J. R., Richards, G. T., Schneider, D. P., Stoughton, C., Szalay, A. S., Thakar, A. R., and York, D. G. (2001). Evidence for Reionization at z˜6: Detection of a Gunn-Peterson Trough in a z=6.28 Quasar. AJ, 122:2850–2857. 3, 137

Benson, A., Venkatesan, A., and Shull, J. M. (2013). The Escape Fraction of Ionizing Radiation from Galaxies. ApJ, 770:76. 137

Berger, M. J. and Colella, P. (1989). Local adaptive mesh refinement for shock hydrodynamics. *Journal of Computational Physics*, 82:64–84. 34

Bergvall, N., Zackrisson, E., Andersson, B.-G., Arnberg, D., Masegosa, J., and Östlin, G. (2006). First detection of Lyman continuum escape from a local starburst galaxy. I. Observations of the luminous blue compact galaxy Haro 11 with the Far Ultraviolet Spectroscopic Explorer (FUSE). A&A, 448:513–524. xxiii, 138

Bi, H. (1993). Lyman-alpha absorption spectrum of the primordial intergalactic medium. ApJ, 405:479–490. 4, 14, 27, 28

Bi, H. and Davidsen, A. F. (1997). Evolution of Structure in the Intergalactic Medium and the Nature of the Lyα Forest. ApJ, 479:523–542. xiv, 4, 14, 27, 35, 37, 139

Bi, H. G., Boerner, G., and Chu, Y. (1992). An alternative model for the Ly-alpha absorption forest. A&A, 266:1–5. 4, 14, 27, 28, 35, 37

Black, J. H. (1981). The physical state of primordial intergalactic clouds. MNRAS, 197:553–563. 14

Boera, E., Murphy, M. T., Becker, G. D., and Bolton, J. S. (2014). The thermal history of the intergalactic medium down to redshift z = 1.5: a new curvature measurement. MNRAS, 441:1916–1933. xxiii, 42, 53, 109, 128, 161, 195

Bolton, J. S. and Becker, G. D. (2009). Resolving the high redshift Lyα forest in smoothed particle hydrodynamics simulations. MNRAS, 398:L26–L30. 55, 139

Bolton, J. S. and Haehnelt, M. G. (2007). The observed ionization rate of the intergalactic medium and the ionizing emissivity at z >= 5: evidence for a photon-starved and extended epoch of reionization. MNRAS, 382:325–341. xv, lv, 25, 36, 87, 139, 174

Bolton, J. S., Haehnelt, M. G., Viel, M., and Carswell, R. F. (2006). Spatial fluctuations in the spectral shape of the ultraviolet background at 2 < z < 3 and the reionization of helium. MNRAS, 366:1378–1390. xvi, 57, 65, 118

Bolton, J. S., Haehnelt, M. G., Viel, M., and Springel, V. (2005). The Lyman α forest opacity and the metagalactic hydrogen ionization rate at z˜ 2-4. MNRAS, 357:1178–1188. xv, 139

Bolton, J. S., Haehnelt, M. G., Warren, S. J., Hewett, P. C., Mortlock, D. J., Venemans, B. P., McMahon, R. G., and Simpson, C. (2011). How neutral is the intergalactic medium surrounding the redshift z = 7.085 quasar ULAS J1120+0641? MNRAS, 416:L70–L74. 3, 137

Bolton, J. S., Viel, M., Kim, T.-S., Haehnelt, M. G., and Carswell, R. F. (2008). Possible evidence for an inverted temperature-density relation in the intergalactic medium from the flux distribution of the Lyα forest. MNRAS, 386:1131–1144. 25, 36, 60, 87, 145

Borthakur, S., Heckman, T. M., Leitherer, C., and Overzier, R. A. (2014). A local clue to the reionization of the universe. *Science*, 346:216–219. xxiii, 138

Boutsia, K., Grazian, A., Giallongo, E., Fontana, A., Pentericci, L., Castellano, M., Zamorani, G., Mignoli, M., Vanzella, E., Fiore, F., Lilly, S. J., Gallozzi, S., Testa, V., Paris, D., and Santini, P. (2011). A Low Escape Fraction of Ionizing Photons of L > L* Lyman Break Galaxies at z = 3.3. ApJ, 736:41. 137

Bridge, C. R., Teplitz, H. I., Siana, B., Scarlata, C., Conselice, C. J., Ferguson, H. C., Brown, T. M., Salvato, M., Rudie, G. C., de Mello, D. F., Colbert, J., Gardner, J. P., Giavalisco, M., and Armus, L. (2010). A Spectroscopic Search for Leaking Lyman Continuum at z ˜ 0.7. ApJ, 720:465–479. 138, 178

Bromm, V. and Larson, R. B. (2004). The First Stars. ARA&A, 42:79–118. 2

Bryan, G. L., Norman, M. L., O'Shea, B. W., Abel, T., Wise, J. H., Turk, M. J., Reynolds, D. R., Collins, D. C., Wang, P., Skillman, S. W., Smith, B., Harkness, R. P., Bordner, J., Kim, J.-h., Kuhlen, M., Xu, H., Goldbaum, N., Hummels, C., Kritsuk, A. G., Tasker, E., Skory, S., Simpson, C. M., Hahn, O., Oishi, J. S., So, G. C., Zhao, F., Cen, R., Li, Y., and Enzo Collaboration (2014). ENZO: An Adaptive Mesh Refinement Code for Astrophysics. ApJS, 211:19. 37

Bryan, G. L., Norman, M. L., Stone, J. M., Cen, R., and Ostriker, J. P. (1995). A piecewise parabolic method for cosmological hydrodynamics. *Computer Physics Communications*, 89:149–168. 34

Busca, N. G., Delubac, T., Rich, J., Bailey, S., Font-Ribera, A., Kirkby, D., Le Goff, J.-M., Pieri, M. M., Slosar, A., Aubourg, É., Bautista, J. E., Bizyaev, D., Blomqvist, M., Bolton, A. S., Bovy, J., Brewington, H., Borde, A., Brinkmann, J., Carithers, B., Croft, R. A. C., Dawson, K. S., Ebelke, G., Eisenstein, D. J., Hamilton, J.-C., Ho, S., Hogg, D. W., Honscheid, K., Lee, K.-G., Lundgren, B., Malanushenko, E., Malanushenko, V., Margala, D., Maraston, C., Mehta, K., Miralda-Escudé, J., Myers, A. D., Nichol, R. C., Noterdaeme, P., Olmstead, M. D., Oravetz, D., Palanque-Delabrouille, N., Pan, K., Pâris, I., Percival, W. J., Petitjean, P., Roe, N. A., Rollinde, E., Ross, N. P., Rossi, G., Schlegel, D. J., Schneider, D. P., Shelden, A., Sheldon, E. S., Simmons, A., Snedden, S., Tinker, J. L., Viel, M., Weaver, B. A., Weinberg, D. H., White, M., Yèche, C., and York, D. G. (2013). Baryon acoustic oscillations in the Lyα forest of BOSS quasars. A&A, 552:A96. 200

Calura, F., Tescari, E., D'Odorico, V., Viel, M., Cristiani, S., Kim, T.-S., and Bolton, J. S. (2012). The Lyman α forest flux probability distribution at z>3. MNRAS, 422:3019–3036. 25, 36, 87

Calverley, A. P., Becker, G. D., Haehnelt, M. G., and Bolton, J. S. (2011). Measurements of the ultraviolet background at 4.6 < z < 6.4 using the quasar proximity effect. MNRAS, 412:2543–2562. 138

Carswell, B., Schaye, J., and Kim, T.-S. (2002). The Enrichment History of the Intergalactic Medium: O VI in Lyα Forest Systems at Redshift z˜2. ApJ, 578:43–59. xxiii, 42, 195

Cen, R. and Fang, T. (2006). Where Are the Baryons? III. Nonequilibrium Effects and Observables. ApJ, 650:573–591. 55, 139

Cen, R. and Kimm, T. (2015). Quantifying Distributions of the Lyman Continuum Escape Fraction. ApJ, 801:L25. xxiii, 137, 195

Cen, R., Miralda-Escudé, J., Ostriker, J. P., and Rauch, M. (1994). Gravitational collapse of small-scale structure as the origin of the Lyman-alpha forest. ApJ, 437:L9–L12. xiv, 35, 53, 55, 139

Cen, R. and Ostriker, J. P. (1999). Where Are the Baryons? ApJ, 514:1–6. 55, 139

Cen, R. and Ostriker, J. P. (2006). Where Are the Baryons? II. Feedback Effects. ApJ, 650:560–572. 55, 139

Charlton, J. and Churchill, C. (2000). *Quasistellar Objects: Intervening Absorption Lines*. xxxvi, 8

Choudhury, T. R. and Ferrara, A. (2006). Physics of Cosmic Reionization. *ArXiv Astrophysics e-prints*. 3

Choudhury, T. R., Srianand, R., and Padmanabhan, T. (2001). Semianalytic Approach to Understanding the Distribution of Neutral Hydrogen in the Universe: Comparison of Simulations with Observations. ApJ, 559:29–40. xiv, xix, 4, 14, 15, 27, 28, 32, 35, 37, 63, 67, 76, 87, 139

Ciardi, B. and Ferrara, A. (2005). The First Cosmic Structures and Their Effects. Space Sci. Rev., 116:625–705. 2, 3

Coles, P., Melott, A. L., and Shandarin, S. F. (1993). Testing approximations for non-linear gravitational clustering. MNRAS, 260:765–776. 28

Cooke, A. J., Espey, B., and Carswell, R. F. (1997). Evolution of the ionizing background at high redshifts. MNRAS, 284:552–568. 25, 36, 88

Cooke, J., Ryan-Weber, E. V., Garel, T., and Díaz, C. G. (2014). Lyman-continuum galaxies and the escape fraction of Lyman-break galaxies. MNRAS, 441:837–851. 88, 137

Couchman, H. M. P., Thomas, P. A., and Pearce, F. R. (1995). Hydra: an Adaptive-Mesh Implementation of P 3M-SPH. ApJ, 452:797. 34

Cowie, L. L., Barger, A. J., and Trouille, L. (2009). Measuring the Sources of the Intergalactic Ionizing Flux. ApJ, 692:1476–1488. xxiii, 138, 178

Cristiani, S., Serrano, L. M., Fontanot, F., Vanzella, E., and Monaco, P. (2016). The spectral slope and escape fraction of bright quasars at z $\sim$ 3.8: the contribution to the cosmic UV background. MNRAS, 462:2478–2485. 175

Croft, R. A. C., Weinberg, D. H., Katz, N., and Hernquist, L. (1997). Intergalactic Helium Absorption in Cold Dark Matter Models. ApJ, 488:532–549. xiv, 4, 14

Croft, R. A. C., Weinberg, D. H., Katz, N., and Hernquist, L. (1998). Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Lyα Forest. ApJ, 495:44–62. 4, 14, 28, 29, 123, 139

Croom, S. M., Richards, G. T., Shanks, T., Boyle, B. J., Strauss, M. A., Myers, A. D., Nichol, R. C., Pimbblet, K. A., Ross, N. P., Schneider, D. P., Sharp, R. G., and Wake, D. A. (2009). The 2dF-SDSS LRG and QSO survey: the QSO luminosity function at $0.4 < z < 2.6$. MNRAS, 399:1755–1772. 177

Dall'Aglio, A., Wisotzki, L., and Worseck, G. (2009). The UV background photoionization rate at $2.3 < z < 4.6$ as measured from the Sloan Digital Sky Survey. *ArXiv e-prints*. 138

Danforth, C. W., Keeney, B. A., Tilton, E. M., Shull, J. M., Stocke, J. T., Stevans, M., Pieri, M. M., Savage, B. D., France, K., Syphers, D., Smith, B. D., Green, J. C., Froning, C., Penton, S. V., and Osterman, S. N. (2016). An HST/COS Survey of the Low-redshift Intergalactic Medium. I. Survey, Methodology, and Overall Results. ApJ, 817:111. xx, xxi, xxiii, xxxviii, xliii, xliv, xlix, lvi, 63, 64, 89, 90, 91, 95, 96, 97, 98, 99, 100, 101, 102, 140, 141, 142, 143, 166, 168, 169, 175, 182, 186, 192, 193

Danforth, C. W. and Shull, J. M. (2008). The Low-z Intergalactic Medium. III. H I and Metal Absorbers at $z < 0.4$. ApJ, 679:194–219. 110

Datta, K. K., Ghara, R., Majumdar, S., Choudhury, T. R., Bharadwaj, S., Roy, H., and Datta, A. (2016). Probing Individual Sources during Reionization and Cosmic Dawn using Square Kilometre Array HI 21-cm Observations. *Journal of Astrophysics and Astronomy*, 37:27. 198

Davé, R., Cen, R., Ostriker, J. P., Bryan, G. L., Hernquist, L., Katz, N., Weinberg, D. H., Norman, M. L., and O'Shea, B. (2001). Baryons in the Warm-Hot Intergalactic Medium. ApJ, 552:473–483. xix, 55, 59, 88, 139, 155, 192

Davé, R., Oppenheimer, B. D., Katz, N., Kollmeier, J. A., and Weinberg, D. H. (2010). The intergalactic medium over the last 10 billion years - I. Lyα absorption and physical conditions. MNRAS, 408:2051–2070. xiv, xxi, xlv, 55, 62, 108, 109, 110, 111, 116, 117, 139, 193

Davé, R. and Tripp, T. M. (2001). The Statistical and Physical Properties of the Low-Redshift LYα Forest Observed with the Hubble Space Telescope/STIS. ApJ, 553:528–537. 55, 59, 62, 131, 139

Deharveng, J.-M., Buat, V., Le Brun, V., Milliard, B., Kunth, D., Shull, J. M., and Gry, C. (2001). Constraints on the Lyman continuum radiation from galaxies: First results with FUSE on Mrk 54. A&A, 375:805–813. 138

Desjacques, V., Nusser, A., and Sheth, R. K. (2007). The probability distribution function of the Lyman α transmitted flux from a sample of Sloan Digital Sky Survey quasars. MNRAS, 374:206–219. 121, 145

Dicke, R. H., Peebles, P. J. E., Roll, P. G., and Wilkinson, D. T. (1965). Cosmic Black-Body Radiation. ApJ, 142:414–419. 1

Doroshkevich, A. G. and Shandarin, S. F. (1977). Structure of the intergalactic medium. MNRAS, 179:95P–99P. 25, 26

Draine, B. (2010). *Physics of the Interstellar and Intergalactic Medium*. Princeton Series in Astrophysics. Princeton University Press. 8

Fan, X., Narayanan, V. K., Lupton, R. H., Strauss, M. A., Knapp, G. R., Becker, R. H., White, R. L., Pentericci, L., Leggett, S. K., Haiman, Z., Gunn, J. E., Ivezić, Ž., Schneider, D. P., Anderson, S. F., Brinkmann, J., Bahcall, N. A., Connolly, A. J., Csabai, I., Doi, M., Fukugita, M., Geballe, T., Grebel, E. K., Harbeck, D., Hennessy, G., Lamb, D. Q., Miknaitis, G., Munn, J. A., Nichol, R., Okamura, S., Pier, J. R., Prada, F., Richards, G. T., Szalay, A., and York, D. G. (2001). A Survey of z>5.8 Quasars in the Sloan Digital Sky Survey. I. Discovery of Three New Quasars and the Spatial Density of Luminous Quasars at z˜6. AJ, 122:2833–2849. 3, 137

Fan, X., Strauss, M. A., Becker, R. H., White, R. L., Gunn, J. E., Knapp, G. R., Richards, G. T., Schneider, D. P., Brinkmann, J., and Fukugita, M. (2006). Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with z˜6 Quasars. II. A Sample of 19 Quasars. AJ, 132:117–136. 3, 12, 137

Fardal, M. A., Giroux, M. L., and Shull, J. M. (1998). The High-Redshift He II Gunn-Peterson Effect: Implications and Future Prospects. AJ, 115:2206–2230. xiv, 4, 137

Faucher-Giguère, C.-A., Lidz, A., and Hernquist, L. (2008a). Numerical Simulations Unravel the Cosmic Web. *Science*, 319:52. xxxv, 2, 55, 139

Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., and Zaldarriaga, M. (2008b). A Flat Photoionization Rate at 2 <= z <= 4.2: Evidence for a Stellar-Dominated UV Background and against a Decline of Cosmic Star Formation beyond z ˜ 3. ApJ, 682:L9. xv, 139

Faucher-Giguère, C.-A., Lidz, A., Zaldarriaga, M., and Hernquist, L. (2008c). The Line-of-Sight Proximity Effect and the Mass of Quasar Host Halos. ApJ, 673:39–61. 138

Faucher-Giguère, C.-A., Lidz, A., Zaldarriaga, M., and Hernquist, L. (2009). A New Calculation of the Ionizing Background Spectrum and the Effects of He II Reionization. ApJ, 703:1416–1443. 134, 175

Ferland, G. J., Korista, K. T., Verner, D. A., Ferguson, J. W., Kingdon, J. B., and Verner, E. M. (1998). CLOUDY 90: Numerical Simulation of Plasmas and Their Spectra. PASP, 110:761–778. 140

Fernandez, E. R. and Shull, J. M. (2011). The Effect of Galactic Properties on the Escape Fraction of Ionizing Photons. ApJ, 731:20. 137

Fernández-Soto, A., Lanzetta, K. M., Barcons, X., Carswell, R. F., Webb, J. K., and Yahil, A. (1996). Strong Clustering of High-Redshift LY alpha Forest Absorption Systems. ApJ, 460:L85. 157

Ferrara, A. and Loeb, A. (2013). Escape fraction of the ionizing radiation from starburst galaxies at high redshifts. MNRAS, 431:2826–2833. 178

Finn, C. W., Morris, S. L., Tejos, N., Crighton, N. H. M., Perry, R., Fumagalli, M., Bielby, R., Theuns, T., Schaye, J., Shanks, T., Liske, J., Gunawardhana, M. L. P., and Bartle, S. (2016). On the connection between the metal-enriched intergalactic medium and galaxies: an O VI-galaxy cross-correlation study at z < 1. MNRAS, 460:590–616. 138

Fujita, A., Martin, C. L., Mac Low, M.-M., and Abel, T. (2003). The Influence of Supershells and Galactic Outflows on the Escape of Ionizing Radiation from Dwarf Starburst Galaxies. ApJ, 599:50–69. 178

Furlanetto, S. R., Oh, S. P., and Briggs, F. H. (2006). Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. Phys. Rep., 433:181–301. 3

Gaikwad, P., Choudhury, T. R., Srianand, R., and Khaire, V. (2017a). Efficient hydrodynamical simulations of the high-redshift intergalactic medium. *ArXiv e-prints*. xix, xxii, 53, 107

Gaikwad, P., Khaire, V., Choudhury, T. R., and Srianand, R. (2017b). Intergalactic Lyman continuum photon budget in the past 5 billion years. MNRAS, 466:838–860. xix, xxii, xxvi, liv, lvi, 53, 67, 87, 90, 107, 118, 119, 121, 124, 134, 137, 156, 169, 170, 172, 175, 176, 180, 197

Gaikwad, P., Srianand, R., Choudhury, T. R., and Khaire, V. (2017c). VoIgt profile Parameter Estimation Routine (viper): H I photoionization rate at z < 0.5. MNRAS, 467:3172–3187. xxi, xxvi, 67, 87, 118, 129, 131, 137, 197

Gardner, J. P., Mather, J. C., Clampin, M., Doyon, R., Greenhouse, M. A., Hammel, H. B., Hutchings, J. B., Jakobsen, P., Lilly, S. J., Long, K. S., Lunine, J. I., McCaughrean, M. J., Mountain, M., Nella, J., Rieke, G. H., Rieke, M. J., Rix, H.-W., Smith, E. P., Sonneborn, G., Stiavelli, M., Stockman, H. S., Windhorst, R. A., and Wright, G. S. (2006). The James Webb Space Telescope. Space Sci. Rev., 123:485–606. 198

Garzilli, A., Bolton, J. S., Kim, T.-S., Leach, S., and Viel, M. (2012). The intergalactic medium thermal history at redshift z = 1.7-3.2 from the Lyα forest: a comparison of measurements using wavelets and the flux distribution. MNRAS, 424:1723–1736. 124

Garzilli, A., Theuns, T., and Schaye, J. (2015). The broadening of Lyman-α forest absorption lines. MNRAS, 450:1465–1476. 87, 131, 158

Gingold, R. A. and Monaghan, J. J. (1977). Smoothed particle hydrodynamics - Theory and application to non-spherical stars. MNRAS, 181:375–389. 30

Gnedin, N. Y. and Hui, L. (1996). The Low Column Density Lyman-alpha Forest. ApJ, 472:L73. 27

Gnedin, N. Y. and Hui, L. (1998). Probing the Universe with the Lyalpha forest - I. Hydrodynamics of the low-density intergalactic medium. MNRAS, 296:44–55. xxi, 17, 26, 32, 36, 37, 74, 194

Gnedin, N. Y., Kravtsov, A. V., and Chen, H.-W. (2008). Escape of Ionizing Radiation from High-Redshift Galaxies. ApJ, 672:765–775. 137

Grazian, A., Giallongo, E., Gerbasi, R., Fiore, F., Fontana, A., Le Fèvre, O., Pentericci, L., Vanzella, E., Zamorani, G., Cassata, P., Garilli, B., Le Brun, V., Maccagni, D., Tasca, L. A. M., Thomas, R., Zucca, E., Amorín, R., Bardelli, S., Cassarà, L. P., Castellano, M., Cimatti, A., Cucciati, O., Durkalec, A., Giavalisco, M., Hathi, N. P., Ilbert, O., Lemaux, B. C., Paltani, S., Ribeiro, B., Schaerer, D., Scodeggio, M., Sommariva, V., Talia, M., Tresse, L., Vergani, D., Bonchi, A., Boutsia, K., Capak, P., Charlot, S., Contini, T., de la Torre, S., Dunlop, J., Fotopoulou, S., Guaita, L., Koekemoer, A., López-Sanjuan, C., Mellier, Y., Merlin, E., Paris, D., Pforr, J., Pilo, S., Santini, P., Scoville, N., Taniguchi, Y., and Wang, P. W. (2016). The Lyman continuum escape fraction of galaxies at z = 3.3 in the VUDS-LBC/COSMOS field. A&A, 585:A48. 137

Guimarães, R., Petitjean, P., Rollinde, E., de Carvalho, R. R., Djorgovski, S. G., Srianand, R., Aghaee, A., and Castro, S. (2007). Evidence for overdensity around $z_{em} > 4$ quasars from the proximity effect. MNRAS, 377:657–666. 138

Gunn, J. E. and Peterson, B. A. (1965). On the Density of Neutral Hydrogen in Intergalactic Space. ApJ, 142:1633–1641. 4, 24, 139

Gurvich, A., Burkhart, B., and Bird, S. (2017). The Effect of AGN Heating on the Low-redshift Lyα Forest. ApJ, 835:175. xix, 25, 36, 53, 88, 129, 175, 192

Haardt, F. and Madau, P. (1996). Radiative Transfer in a Clumpy Universe. II. The Ultraviolet Extragalactic Background. ApJ, 461:20. xiv, 4, 137

Haardt, F. and Madau, P. (2012). Radiative Transfer in a Clumpy Universe. IV. New Synthesis Models of the Cosmic UV/X-Ray Background. ApJ, 746:125. xxii, xxv, xxxvii, 20, 43, 66, 175, 194, 196

Hernquist, L. (1993). Some cautionary remarks about smoothed particle hydrodynamics. ApJ, 404:717–722. 30

Hernquist, L. and Katz, N. (1989). TREESPH - A unification of SPH with the hierarchical tree method. ApJS, 70:419–446. 30

Hernquist, L., Katz, N., Weinberg, D. H., and Miralda-Escudé, J. (1996). The Lyman-Alpha Forest in the Cold Dark Matter Model. ApJ, 457:L51. xiv, 35, 53, 55, 139

Hinshaw, G., Larson, D., Komatsu, E., Spergel, D. N., Bennett, C. L., Dunkley, J., Nolta, M. R., Halpern, M., Hill, R. S., Odegard, N., Page, L., Smith, K. M., Weiland, J. L., Gold, B., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., Wollack, E., and Wright, E. L. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. ApJS, 208:19. 1, 42

Hopkins, P. F., Richards, G. T., and Hernquist, L. (2007). An Observational Determination of the Bolometric Quasar Luminosity Function. ApJ, 654:731–753. 177

Hui, L. and Gnedin, N. Y. (1997). Equation of state of the photoionized intergalactic medium. MNRAS, 292:27. xvi, xvii, 3, 17, 22, 29, 57, 58, 67, 69, 76

Hui, L., Gnedin, N. Y., and Zhang, Y. (1997). The Statistics of Density Peaks and the Column Density Distribution of the Lyα Forest. ApJ, 486:599–622. 25, 27, 35

Ikeuchi, S. (1986). The baryon clump within an extended dark matter region. Ap&SS, 118:509–514. 14

Ikeuchi, S. and Ostriker, J. P. (1986). Evolution of the intergalactic medium - What happened during the epoch Z = 3-10? ApJ, 301:522–543. 13

Inoue, A. K., Iwata, I., and Deharveng, J.-M. (2006). The escape fraction of ionizing photons from galaxies at z = 0-6. MNRAS, 371:L1–L5. xxiii, 138, 195

Inoue, A. K., Shimizu, I., Iwata, I., and Tanaka, M. (2014). An updated analytic model for attenuation by the intergalactic medium. MNRAS, 442:1805–1820. xxv, xxvi, lv, 174, 177, 178, 181, 196, 197

Iwata, I., Inoue, A. K., Matsuda, Y., Furusawa, H., Hayashino, T., Kousai, K., Akiyama, M., Yamada, T., Burgarella, D., and Deharveng, J.-M. (2009). Detections of Lyman Continuum from Star-Forming Galaxies at z ˜ 3 through Subaru/Suprime-Cam Narrow-Band Imaging. ApJ, 692:1287–1293. 137

Izotov, Y. I., Orlitová, I., Schaerer, D., Thuan, T. X., Verhamme, A., Guseva, N. G., and Worseck, G. (2016a). Eight per cent leakage of Lyman continuum photons from a compact, star-forming dwarf galaxy. Nature, 529:178–180. xxiii

Izotov, Y. I., Orlitová, I., Schaerer, D., Thuan, T. X., Verhamme, A., Guseva, N. G., and Worseck, G. (2016b). Eight per cent leakage of Lyman continuum photons from a compact, star-forming dwarf galaxy. Nature, 529:178–180. 138

Izotov, Y. I., Schaerer, D., Thuan, T. X., Worseck, G., Guseva, N. G., Orlitová, I., and Verhamme, A. (2016c). Detection of high Lyman continuum leakage from four low-redshift compact star-forming galaxies. MNRAS, 461:3683–3701. xxiii

Jeffreys, H. (1961). *Theory of Probability.* 3rd edn. Oxford Univ. Press, Oxford. 93

Jenkins, E. B. and Ostriker, J. P. (1991). Lyman-alpha depression of the continuum from high-redshift quasars - A new technique applied in search of the Gunn-Peterson effect. ApJ, 376:33–42. 121, 145

Katz, N., Weinberg, D. H., and Hernquist, L. (1996). Cosmological Simulations with TreeSPH. ApJS, 105:19. 18, 60

Keeney, B. A., Danforth, C. W., Stocke, J. T., France, K., and Green, J. C. (2012). On the Significance of Absorption Features in HST/COS Data. PASP, 124:830. 95, 103

Khaire, V. and Srianand, R. (2013). He II optical depth and ultraviolet escape fraction of galaxies. MNRAS, 431:L53–L57. 59, 138, 175

Khaire, V. and Srianand, R. (2015a). Photon underproduction crisis: Are QSOs sufficient to resolve it? MNRAS, 451:L30–L34. xiv, xxv, lvi, 4, 60, 133, 134, 176, 196

Khaire, V. and Srianand, R. (2015b). Star Formation History, Dust Attenuation, and Extragalactic Background Light. ApJ, 805:33. lvi, 43, 133, 134, 141, 176, 177

Khaire, V., Srianand, R., Choudhury, T. R., and Gaikwad, P. (2016). The redshift evolution of escape fraction of hydrogen ionizing photons from galaxies. MNRAS, 457:4051–4062. xxiii, 3, 137, 138, 177, 178, 195

Khare, P. (2013). Quasar absorption lines: an overview. *Bulletin of the Astronomical Society of India*, 41:41. 4

Kim, T.-S., Bolton, J. S., Viel, M., Haehnelt, M. G., and Carswell, R. F. (2007). An improved measurement of the flux distribution of the Lyα forest in QSO absorption spectra: the effect of continuum fitting, metal contamination and noise properties. MNRAS, 382:1657–1674. 121, 145

Kim, T.-S., Partl, A. M., Carswell, R. F., and Müller, V. (2013). The evolution of H I and C IV quasar absorption line systems at $1.9 < z < 3.2$. A&A, 552:A77. 137

Kim, T.-S., Viel, M., Haehnelt, M. G., Carswell, R. F., and Cristiani, S. (2004). The power spectrum of the flux distribution in the Lyman α forest of a large sample of UVES QSO absorption spectra (LUQAS). MNRAS, 347:355–366. 119, 123, 149

King, J. A., Murphy, M. T., Ubachs, W., and Webb, J. K. (2011). New constraint on cosmological variation of the proton-to-electron mass ratio from Q0528-250. MNRAS, 417:3010–3024. 93

Kirkman, D. and Tytler, D. (2008). The transverse proximity effect in the z ˜ 2 Lyman α forest suggests quasi-stellar object episodic lifetimes of ˜1 Myr. MNRAS, 391:1457–1471. 138

Knebe, A., Green, A., and Binney, J. (2001). Multi-level adaptive particle mesh (MLAPM): a c code for cosmological simulations. MNRAS, 325:845–864. 34

Kollmeier, J. A., Miralda-Escudé, J., Cen, R., and Ostriker, J. P. (2006). Galactic Wind Effects on the Lyα Absorption in the Vicinity of Galaxies. ApJ, 638:52–71. xiv

Kollmeier, J. A., Weinberg, D. H., Oppenheimer, B. D., Haardt, F., Katz, N., Davé, R., Fardal, M., Madau, P., Danforth, C., Ford, A. B., Peeples, M. S., and McEwen, J. (2014). The Photon Underproduction Crisis. ApJ, 789:L32. xix, xxiii, xxv, xxvi, lv, lvi, 25, 36, 53, 88, 109, 116, 117, 129, 139, 140, 173, 174, 175, 176, 180, 182, 186, 187, 188, 192, 196, 197

Komatsu, E., Smith, K. M., Dunkley, J., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Nolta, M. R., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L. (2011). Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. ApJS, 192:18. 1, 42

Kriss, G. A., Shull, J. M., Oegerle, W., Zheng, W., Davidsen, A. F., Songaila, A., Tumlinson, J., Cowie, L. L., Deharveng, J.-M., Friedman, S. D., Giroux, M. L., Green, R. F., Hutchings, J. B., Jenkins, E. B., Kruk, J. W., Moos, H. W., Morton, D. C., Sembach, K. R., and Tripp, T. M. (2001). Resolving the Structure of Ionized Helium in the Intergalactic Medium with the Far Ultraviolet Spectroscopic Explorer. *Science*, 293:1112–1116. 3, 59

Krolewski, A., Lee, K.-G., Lukić, Z., and White, M. (2017). Measuring Alignments between Galaxies and the Cosmic Web at z $\sim$ 2-3 Using IGM Tomography. ApJ, 837:31. 199

Kulkarni, G., Hennawi, J. F., Oñorbe, J., Rorai, A., and Springel, V. (2015). Characterizing the Pressure Smoothing Scale of the Intergalactic Medium. ApJ, 812:30. xxi, 17, 74, 75, 194

Kulkarni, V. P. and Fall, S. M. (1993). The proximity effect and the mean intensity of ionizing radiation at low redshifts. ApJ, 413:L63–L66. 138

Kulkarni, V. P., Fall, S. M., Lauroesch, J. T., York, D. G., Welty, D. E., Khare, P., and Truran, J. W. (2005). Hubble Space Telescope Observations of Element Abundances in Low-Redshift Damped Lyα Galaxies and Implications for the Global Metallicity-Redshift Relation. ApJ, 618:68–90. 166

Kutyrev, A. S. and Reynolds, R. J. (1989). Probable detection of H-alpha emission from a very high velocity cloud in Cetus. ApJ, 344:L9–L11. 138

Lee, K.-G., Hennawi, J. F., White, M., Prochaska, J. X., Font-Ribera, A., Schlegel, D. J., Rich, R. M., Suzuki, N., Stark, C. W., Le Fèvre, O., Nugent, P. E., Salvato, M., and Zamorani, G. (2016). Shadow of a Colossus: A z = 2.44 Galaxy Protocluster Detected in 3D Lyα Forest Tomographic Mapping of the COSMOS Field. ApJ, 817:160. 199

Lehner, N., Savage, B. D., Richter, P., Sembach, K. R., Tripp, T. M., and Wakker, B. P. (2007). Physical Properties, Baryon Content, and Evolution of the Lyα Forest: New Insights from High-Resolution Observations at z <˜ 0.4. ApJ, 658:680–709. 55, 139

Leitet, E., Bergvall, N., Hayes, M., Linné, S., and Zackrisson, E. (2013). Escape of Lyman continuum radiation from local galaxies. Detection of leakage from the young starburst Tol 1247-232. A&A, 553:A106. xxiii, 138

Leitherer, C., Hernandez, S., Lee, J. C., and Oey, M. S. (2016). Direct Detection of Lyman Continuum Escape from Local Starburst Galaxies with the Cosmic Origins Spectrograph. ApJ, 823:64. 138

Liddle, A. R. (2007). Information criteria for astrophysical model selection. MNRAS, 377:L74–L78. 93

Lidz, A., Faucher-Giguère, C.-A., Dall'Aglio, A., McQuinn, M., Fechner, C., Zaldarriaga, M., Hernquist, L., and Dutta, S. (2010). A Measurement of Small-scale Structure in the 2.2 <= z <= 4.2 Lyα Forest. ApJ, 718:199–230. xv, xix, xxii, xxiii, 3, 25, 36, 42, 53, 87, 109, 124, 127, 189, 194, 195

Lidz, A., McQuinn, M., Zaldarriaga, M., Hernquist, L., and Dutta, S. (2007). Quasar Proximity Zones and Patchy Reionization. ApJ, 670:39–59. 142

Loeb, A. and Barkana, R. (2001). The Reionization of the Universe by the First Stars and Quasars. ARA&A, 39:19–66. 2, 3

Lucy, L. B. (1977). A numerical approach to the testing of the fission hypothesis. AJ, 82:1013–1024. 30

Lukić, Z., Stark, C. W., Nugent, P., White, M., Meiksin, A. A., and Almgren, A. (2015). The Lyman α forest in optically thin hydrodynamical simulations. MNRAS, 446:3697–3724. 75

Lusso, E., Hennawi, J. F., Comastri, A., Zamorani, G., Richards, G. T., Vignali, C., Treister, E., Schawinski, K., Salvato, M., and Gilli, R. (2014). Erratum: "The Obscured Fraction of Active Galactic Nuclei in the XMM-COSMOS Survey: A Spectral Energy Distribution Perspective" <A href="/abs/2013ApJ...777...86L">(2013, ApJ, 777, 86)</A>. ApJ, 784:176. 177

Madau, P., Haardt, F., and Rees, M. J. (1999). Radiative Transfer in a Clumpy Universe. III. The Nature of Cosmological Ionizing Sources. ApJ, 514:648–659. 3, 137

Madsen, G. J., Reynolds, R. J., Haffner, L. M., Tufte, S. L., and Maloney, P. R. (2001). Observations of the Extended Distribution of Ionized Hydrogen in the Plane of M31. ApJ, 560:L135–L138. 138

Mao, Q., Berlind, A. A., Scherrer, R. J., Neyrinck, M. C., Scoccimarro, R., Tinker, J. L., McBride, C. K., Schneider, D. P., Pan, K., Bizyaev, D., Malanushenko, E., and Malanushenko, V. (2017). A Cosmic Void Catalog of SDSS DR12 BOSS Galaxies. ApJ, 835:161. 165

Mather, J. C., Cheng, E. S., Cottingham, D. A., Eplee, Jr., R. E., Fixsen, D. J., Hewagama, T., Isaacman, R. B., Jensen, K. A., Meyer, S. S., Noerdlinger, P. D., Read, S. M., Rosen, L. P., Shafer, R. A., Wright, E. L., Bennett, C. L., Boggess, N. W., Hauser, M. G., Kelsall, T., Moseley, Jr., S. H., Silverberg, R. F., Smoot, G. F., Weiss, R., and Wilkinson, D. T. (1994). Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument. ApJ, 420:439–444. 1, 42

McDonald, P. (2003). Toward a Measurement of the Cosmological Geometry at z ˜ 2: Predicting Lyα Forest Correlation in Three Dimensions and the Potential of Future Data Sets. ApJ, 585:34–51. 123

McDonald, P. and Eisenstein, D. J. (2007). Dark energy and curvature from a future baryonic acoustic oscillation survey using the Lyman-α forest. Phys. Rev. D, 76(6):063009. 199, 200

McDonald, P., Miralda-Escudé, J., Rauch, M., Sargent, W. L. W., Barlow, T. A., and Cen, R. (2001). A Measurement of the Temperature-Density Relation in the Intergalactic Medium Using a New Lyα Absorption-Line Fitting Method. ApJ, 562:52–75. 57, 88, 131, 155

McDonald, P., Miralda-Escudé, J., Rauch, M., Sargent, W. L. W., Barlow, T. A., Cen, R., and Ostriker, J. P. (2000). The Observed Probability Distribution Function, Power Spectrum, and Correlation Function of the Transmitted Flux in the Lyα Forest. ApJ, 543:1–23. 4, 87, 121, 123, 146, 149, 150

McDonald, P., Seljak, U., Burles, S., Schlegel, D. J., Weinberg, D. H., Cen, R., Shih, D., Schaye, J., Schneider, D. P., Bahcall, N. A., Briggs, J. W., Brinkmann, J., Brunner, R. J., Fukugita, M., Gunn, J. E., Ivezić, Ž., Kent, S., Lupton, R. H., and Vanden Berk, D. E. (2006). The Lyα Forest Power Spectrum from the Sloan Digital Sky Survey. ApJS, 163:80–109. xiv

McDonald, P., Seljak, U., Cen, R., Shih, D., Weinberg, D. H., Burles, S., Schneider, D. P., Schlegel, D. J., Bahcall, N. A., Briggs, J. W., Brinkmann, J., Fukugita, M., Ivezić, Ž., Kent, S., and Vanden Berk, D. E. (2005). The Linear Theory Power Spectrum from the Lyα Forest in the Sloan Digital Sky Survey. ApJ, 635:761–783. xv, 4, 35, 53, 69, 87

McGill, C. (1990). The redshift projection. II - Caustics and the Lyman alpha forest. MNRAS, 242:544–554. 25, 37

McQuinn, M., Oh, S. P., and Faucher-Giguère, C.-A. (2011). On Lyman-limit Systems and the Evolution of the Intergalactic Ionizing Background. ApJ, 743:82. 25, 36, 87

Meiksin, A. and White, M. (2004). The effects of ultraviolet background correlations on Lyα forest flux statistics. MNRAS, 350:1107–1126. 25, 36, 87

Meiksin, A. A. (2009). The physics of the intergalactic medium. *Reviews of Modern Physics*, 81:1405–1469. 14, 199

Melott, A. L. (1980). Quasar L-alpha absorbers - Are precise conclusions possible. ApJ, 241:889–893. 14

Micheva, G., Iwata, I., Inoue, A. K., Matsuda, Y., Yamada, T., and Hayashino, T. (2015). Searching for candidates of Lyman continuum sources - revisiting the SSA22 field. *ArXiv e-prints*. 138

Miralda-Escudé, J., Cen, R., Ostriker, J. P., and Rauch, M. (1996). The Ly alpha Forest from Gravitational Collapse in the Cold Dark Matter + Lambda Model. ApJ, 471:582. xiv, 4, 14, 35, 53, 55, 139

Miralda-Escude, J. and Ostriker, J. P. (1990). What produces the ionizing background at large redshift? ApJ, 350:1–22. xiv, 4, 137

Mitra, S., Choudhury, T. R., and Ferrara, A. (2015). Cosmic reionization after Planck. MNRAS, 454:L76–L80. 178

Mo, H., van den Bosch, F., and White, S. (2010). *Galaxy Formation and Evolution*. Galaxy Formation and Evolution. Cambridge University Press. 8, 13, 40

Mo, H. J. and Miralda-Escude, J. (1996). Gaseous Galactic Halos and Quasistellar Object Absorption-Line Systems. ApJ, 469:589. 13

Monaghan, J. J. (1992). Smoothed particle hydrodynamics. ARA&A, 30:543–574. 30, 31, 73

Morales, M. F. and Wyithe, J. S. B. (2010). Reionization and Cosmology with 21-cm Fluctuations. ARA&A, 48:127–171. 3

Mostardi, R. E., Shapley, A. E., Steidel, C. C., Trainor, R. F., Reddy, N. A., and Siana, B. (2015). A High-Resolution Hubble Space Telescope Study of Apparent Lyman Continuum Leakers at z˜3. ApJ, 810:107. 138

Muecket, J. P., Petitjean, P., Kates, R. E., and Riediger, R. (1996). Evolution of the Lyα forest: a consistent picture. A&A, 308:17–26. 28, 29, 37

Muzahid, S., Srianand, R., Bergeron, J., and Petitjean, P. (2012). A high-resolution study of intergalactic O VI absorbers at z$\sim$ 2.3. MNRAS, 421:446–467. xxv, 156, 196

Narlikar, J. (2002). *An Introduction to Cosmology*. Cambridge University Press. 40

Nelson, D., Pillepich, A., Genel, S., Vogelsberger, M., Springel, V., Torrey, P., Rodriguez-Gomez, V., Sijacki, D., Snyder, G. F., Griffen, B., Marinacci, F., Blecha, L., Sales, L., Xu, D., and Hernquist, L. (2015). The illustris simulation: Public data release. *Astronomy and Computing*, 13:12–37. 32

Nestor, D. B., Shapley, A. E., Kornei, K. A., Steidel, C. C., and Siana, B. (2013). A Refined Estimate of the Ionizing Emissivity from Galaxies at z ˜= 3: Spectroscopic Follow-up in the SSA22a Field. ApJ, 765:47. 138

Oppenheimer, B. D., Crain, R. A., Schaye, J., Rahmati, A., Richings, A. J., Trayford, J. W., Tumlinson, J., Bower, R. G., Schaller, M., and Theuns, T. (2016). Bimodality of low-redshift circumgalactic O VI in non-equilibrium EAGLE zoom simulations. MNRAS, 460:2157–2179. 138

Oppenheimer, B. D. and Davé, R. (2009). The nature and origin of low-redshift OVI absorbers. MNRAS, 395:1875–1904. li, lii, 154, 156, 157

O'Shea, B. W., Bryan, G., Bordner, J., Norman, M. L., Abel, T., Harkness, R., and Kritsuk, A. (2004). Introducing Enzo, an AMR Cosmology Application. *ArXiv Astrophysics e-prints*. 34, 37

Ostriker, J. P. and Cowie, L. L. (1981). Galaxy formation in an intergalactic medium dominated by explosions. ApJ, 243:L127–L131. 13

Ostriker, J. P. and Ikeuchi, S. (1983). Physical properties of the intergalactic medium and the Lyman-alpha absorbing clouds. ApJ, 268:L63–L68. 13

Padmanabhan, H., Choudhury, T. R., and Srianand, R. (2014). Probing reionization using quasar near-zones at redshift z $\sim$ 6. MNRAS, 443:3761–3779. 32, 128

Padmanabhan, H., Srianand, R., and Choudhury, T. R. (2015). Measuring the equation of state of the high-z intergalactic medium using curvature statistics. MNRAS, 450:L29–L33. 32, 63, 128

Padmanabhan, T. (2000). *Theoretical Astrophysics: Volume 3, Galaxies and Cosmology*. Theoretical Astrophysics. Cambridge University Press. 40

Padmanabhan, T. (2010). *Gravitation: Foundations and Frontiers*. Cambridge University Press. 40

Palanque-Delabrouille, N., Magneville, C., Yèche, C., Eftekharzadeh, S., Myers, A. D., Petitjean, P., Pâris, I., Aubourg, E., McGreer, I., Fan, X., Dey, A., Schlegel, D., Bailey, S., Bizayev, D., Bolton, A., Dawson, K., Ebelke, G., Ge, J., Malanushenko, E., Malanushenko, V., Oravetz, D., Pan, K., Ross, N. P., Schneider, D. P., Sheldon, E., Simmons, A., Tinker, J., White, M., and

Willmer, C. (2013). Luminosity function from dedicated SDSS-III and MMT data of quasars in 0.7 < z < 4.0 selected with a new approach. A&A, 551:A29. 177

Paschos, P., Jena, T., Tytler, D., Kirkman, D., and Norman, M. L. (2009). The Lyα forest at redshifts 0.1-1.6: good agreement between a large hydrodynamic simulation and HST spectra. MNRAS, 399:1934–1953. 110, 116

Paschos, P. and Norman, M. L. (2005). A Statistical Analysis of Intergalactic Medium Transmission Approaching Reionization. ApJ, 631:59–84. 55, 139

Peebles, P. (1993). *Principles of Physical Cosmology*. Princeton series in physics. Princeton University Press. 40

Peeples, M. S., Weinberg, D. H., Davé, R., Fardal, M. A., and Katz, N. (2010). Pressure support versus thermal broadening in the Lyman α forest - I. Effects of the equation of state on longitudinal structure. MNRAS, 404:1281–1294. 17

Peeples, M. S., Werk, J. K., Tumlinson, J., Oppenheimer, B. D., Prochaska, J. X., Katz, N., and Weinberg, D. H. (2014). A Budget and Accounting of Metals at z ˜ 0: Results from the COS-Halos Survey. ApJ, 786:54. xxiii, 42, 195

Penton, S. V., Shull, J. M., and Stocke, J. T. (2000). The Local Lyα Forest. II. Distribution of H I Absorbers,Doppler Widths, and Baryon Content. ApJ, 544:150–175. 4, 88

Penzias, A. A. and Wilson, R. W. (1965). A Measurement of Excess Antenna Temperature at 4080 Mc/s. ApJ, 142:419–421. 1

Petitjean, P., Mueket, J. P., and Kates, R. E. (1995). The LY alpha forest at low redshift: Tracing the dark matter filaments. A&A, 295:L9–L12. 28, 29, 37

Phillips, J., Weinberg, D. H., Croft, R. A. C., Hernquist, L., Katz, N., and Pettini, M. (2001). Constraints on Cosmological Parameters from the Lyα Forest Power Spectrum and COBE DMR. ApJ, 560:15–27. 4, 87

Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2014). Planck 2013 results. XVI. Cosmological parameters. A&A, 571:A16. 1, 42

Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., Bartlett, J. G., and et al. (2016). Planck 2015 results. XIII. Cosmological parameters. A&A, 594:A13. 1, 3, 42, 54, 137, 163, 164

Pontzen, A., Bird, S., Peiris, H., and Verde, L. (2014). Constraints on Ionizing Photon Production from the Large-scale Lyα Forest. ApJ, 792:L34. 25, 36, 87

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical recipes in FORTRAN. The art of scientific computing.* 124, 145, 148, 169

Puchwein, E., Bolton, J. S., Haehnelt, M. G., Madau, P., Becker, G. D., and Haardt, F. (2015). The photoheating of the intergalactic medium in synthesis models of the UV background. MNRAS, 450:4081–4097. xxii, xl, 23, 71, 72, 81, 161

Rahmati, A., Schaye, J., Crain, R. A., Oppenheimer, B. D., Schaller, M., and Theuns, T. (2016). Cosmic distribution of highly ionized metals and their physical conditions in the EAGLE simulations. MNRAS. 138

Rauch, M., Miralda-Escudé, J., Sargent, W. L. W., Barlow, T. A., Weinberg, D. H., Hernquist, L., Katz, N., Cen, R., and Ostriker, J. P. (1997). The Opacity of the Lyα Forest and Implications for Ω_b and the Ionizing Background. ApJ, 489:7–20. 25, 36, 53, 55, 87, 139

Razoumov, A. O. and Sommer-Larsen, J. (2006). Escape of Ionizing Radiation from Star-forming Regions in Young Galaxies. ApJ, 651:L89–L92. 178

Rees, M. J. (1986). Lyman absorption lines in quasar spectra - Evidence for gravitationally-confined gas in dark minihaloes. MNRAS, 218:25P–30P. 14

Regan, J. A., Haehnelt, M. G., and Viel, M. (2007). Numerical simulations of the Lyman α forest - a comparison of GADGET-2 and ENZO. MNRAS, 374:196–205. 36

Ricotti, M. and Shull, J. M. (2000). Feedback from Galaxy Formation: Escaping Ionizing Radiation from Galaxies at High Redshift. ApJ, 542:548–558. xxiii, 137, 195

Rindler, W. (2001). *Relativity: Special, General, and Cosmological*. Relativity: Special, General, and Cosmological. Oxford University Press. 40

Robertson, B. E., Ellis, R. S., Dunlop, J. S., McLure, R. J., and Stark, D. P. (2010). Early star-forming galaxies and the reionization of the Universe. Nature, 468:49–55. 3, 137

Rollinde, E., Srianand, R., Theuns, T., Petitjean, P., and Chand, H. (2005). The density structure around quasars from optical depth statistics*. MNRAS, 361:1015–1029. 138

Rollinde, E., Theuns, T., Schaye, J., Pâris, I., and Petitjean, P. (2013). Sample variance and Lyman α forest transmission statistics. MNRAS, 428:540–550. 118, 119, 121, 140, 146, 147

Roy, A., Nath, B. B., and Sharma, P. (2015). Narrow escape: how ionizing photons escape from disc galaxies. MNRAS, 451:1939–1954. xxiii, 137, 195

Rutkowski, M. J., Scarlata, C., Haardt, F., Siana, B., Henry, A., Rafelski, M., Hayes, M., Salvato, M., Pahl, A. J., Mehta, V., Beck, M., Malkan, M., and Teplitz, H. I. (2016). Lyman Continuum Escape Fraction of Star-forming Dwarf Galaxies at z $\sim$ 1. ApJ, 819:81. 138, 178

Sargent, W. L. W., Young, P. J., Boksenberg, A., and Tytler, D. (1980). The distribution of Lyman-alpha absorption lines in the spectra of six QSOs - Evidence for an intergalactic origin. ApJS, 42:41–81. 13

Schaye, J. (2001). Model-independent Insights into the Nature of the Lyα Forest and the Distribution of Matter in the Universe. ApJ, 559:507–515. 67, 72, 96, 110, 113, 129

Schaye, J., Crain, R. A., Bower, R. G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C. S., McCarthy, I. G., Helly, J. C., Jenkins, A., Rosas-Guevara, Y. M., White, S. D. M., Baes, M., Booth, C. M., Camps, P., Navarro, J. F., Qu, Y., Rahmati, A., Sawala, T., Thomas, P. A., and Trayford, J. (2015). The EAGLE project: simulating the evolution and assembly of galaxies and their environments. MNRAS, 446:521–554. 32

Schaye, J., Dalla Vecchia, C., Booth, C. M., Wiersma, R. P. C., Theuns, T., Haas, M. R., Bertone, S., Duffy, A. R., McCarthy, I. G., and van de Voort, F. (2010). The physics driving the cosmic star formation history. MNRAS, 402:1536–1560. xiv

Schaye, J., Theuns, T., Leonard, A., and Efstathiou, G. (1999). Measuring the equation of state of the intergalactic medium. MNRAS, 310:57–70. xix, 53, 88, 114, 131, 155, 158, 192

Schaye, J., Theuns, T., Rauch, M., Efstathiou, G., and Sargent, W. L. W. (2000). The thermal history of the intergalactic medium*. MNRAS, 318:817–826. xix, 3, 53, 88, 109, 131, 155, 158, 161, 192

Schirber, M., Miralda-Escudé, J., and McDonald, P. (2004). The Transverse Proximity Effect: A Probe to the Environment, Anisotropy, and Megayear Variability of QSOs. ApJ, 610:105–116. 138

Schmittfull, M. and White, M. (2016). Improving photometric redshifts with Lyα tomography. MNRAS, 463:332–337. 199

Schneider, D. P., Hartig, G. F., Jannuzi, B. T., Kirhakos, S., Saxe, D. H., Weymann, R. J., Bahcall, J. N., Bergeron, J., Boksenberg, A., Sargent, W. L. W., Savage, B. D., Turnshek, D. A., and Wolfe, A. M. (1993). The Hubble Space Telescope quasar absorption line key project. II - Data calibration and absorption-line selection. ApJS, 87:45–62. 90

Schulze, A., Wisotzki, L., and Husemann, B. (2009). Low redshift AGN in the Hamburg/ESO survey. I. The local AGN luminosity function. A&A, 507:781–793. 177

Scoccimarro, R., Hui, L., Manera, M., and Chan, K. C. (2012). Large-scale bias and efficient generation of initial conditions for nonlocal primordial non-Gaussianity. Phys. Rev. D, 85(8):083002. 55, 65

Scott, J., Bechtold, J., Dobrzycki, A., and Kulkarni, V. P. (2000). A Uniform Analysis of the Lyα Forest at z=0-5. II. Measuring the Mean Intensity of the Extragalactic Ionizing Background Using the Proximity Effect. ApJS, 130:67–89. 138

Seljak, U., Slosar, A., and McDonald, P. (2006). Cosmological parameters from combining the Lyman-α forest with CMB, galaxy clustering and SN constraints. *JCAP*, 10:014. 4, 87

Shapley, A. E., Steidel, C. C., Strom, A. L., Bogosavljević, M., Reddy, N. A., Siana, B., Mostardi, R. E., and Rudie, G. C. (2016). Q1549-C25: A Clean Source of Lyman-Continuum Emission at z = 3.15. ApJ, 826:L24. xxiii

Shull, J. M. (2004). He II Reionization and Sources of Metagalactic Ionization. *ArXiv Astrophysics e-prints*. 59

Shull, J. M., Danforth, C. W., and Tilton, E. M. (2014). Tracing the Cosmic Metal Evolution in the Low-redshift Intergalactic Medium. ApJ, 796:49. xxiii, 42, 138, 140, 173, 195

Shull, J. M., France, K., Danforth, C. W., Smith, B., and Tumlinson, J. (2010). HST/COS Observations of the Quasar HE 2347-4342: Probing the Epoch of He II Patchy Reionization at Redshifts z = 2.4-2.9. ApJ, 722:1312–1324. 3

Shull, J. M., Moloney, J., Danforth, C. W., and Tilton, E. M. (2015). The Metagalactic Ionizing Background: A Crisis in UV Photon Production or Incorrect Galaxy Escape Fractions? ApJ, 811:3. xix, xxiii, xxv, lv, lvi, 25, 36, 53, 55, 62, 88, 99, 109, 116, 117, 129, 136, 139, 140, 168, 172, 173, 174, 175, 176, 180, 192, 196

Shull, J. M., Roberts, D., Giroux, M. L., Penton, S. V., and Fardal, M. A. (1999). The Metagalactic Ionizing Radiation Field at Low Redshift. AJ, 118:1450–1460. xiv, 4, 137

Shull, J. M., Smith, B. D., and Danforth, C. W. (2012). The Baryon Census in a Multiphase Intergalactic Medium: 30% of the Baryons May Still be Missing. ApJ, 759:23. xix, xxiii, 4, 42, 55, 62, 88, 96, 129, 139, 166, 192, 195

Shull, J. M., Tumlinson, J., Giroux, M. L., Kriss, G. A., and Reimers, D. (2004). The Fluctuating Intergalactic Radiation Field at Redshifts z = 2.3-2.9 from He II and H I Absorption toward HE 2347-4342. ApJ, 600:570–579. 3

Siana, B., Shapley, A. E., Kulas, K. R., Nestor, D. B., Steidel, C. C., Teplitz, H. I., Alavi, A., Brown, T. M., Conselice, C. J., Ferguson, H. C., Dickinson, M., Giavalisco, M., Colbert, J. W., Bridge, C. R., Gardner, J. P., and de Mello, D. F. (2015). A Deep Hubble Space Telescope and Keck Search for Definitive Identification of Lyman Continuum Emitters at z˜3.1. ApJ, 804:17. 138

Siana, B., Teplitz, H. I., Ferguson, H. C., Brown, T. M., Giavalisco, M., Dickinson, M., Chary, R.-R., de Mello, D. F., Conselice, C. J., Bridge, C. R., Gardner, J. P., Colbert, J. W., and Scarlata, C. (2010). A Deep Hubble Space Telescope Search for Escaping Lyman Continuum Flux at z ˜ 1.3: Evidence for an Evolving Ionizing Emissivity. ApJ, 723:241–250. xxiii, 138, 178

Skidmore, W., TMT International Science Development Teams, and Science Advisory Committee, T. (2015). Thirty Meter Telescope Detailed Science Case: 2015. *Research in Astronomy and Astrophysics*, 15:1945. 5, 198, 199

Slosar, A., Iršič, V., Kirkby, D., Bailey, S., Busca, N. G., Delubac, T., Rich, J., Aubourg, É., Bautista, J. E., Bhardwaj, V., Blomqvist, M., Bolton, A. S., Bovy, J., Brownstein, J., Carithers, B., Croft, R. A. C., Dawson, K. S., Font-Ribera, A., Le Goff, J.-M., Ho, S., Honscheid, K., Lee, K.-G., Margala, D., McDonald, P., Medolin, B., Miralda-Escudé, J., Myers, A. D., Nichol, R. C., Noterdaeme, P., Palanque-Delabrouille, N., Pâris, I., Petitjean, P., Pieri, M. M., Piškur, Y., Roe, N. A., Ross, N. P., Rossi, G., Schlegel, D. J., Schneider, D. P., Suzuki, N., Sheldon, E. S., Seljak, U., Viel, M., Weinberg, D. H., and Yèche, C. (2013). Measurement of baryon acoustic oscillations in the Lyman-α forest fluctuations in BOSS data release 9. *JCAP*, 4:026. 200

Smith, B. D., Hallman, E. J., Shull, J. M., and O'Shea, B. W. (2011). The Nature of the Warm/Hot Intergalactic Medium. I. Numerical Methods, Convergence, and O VI Absorption. ApJ, 731:6. xxi, 55, 62, 109, 110, 111, 116, 117, 139, 165, 193

Smith, B. M., Windhorst, R. A., Jansen, R. A., Cohen, S. H., Jiang, L., Dijkstra, M., Koekemoer, A. M., Bielby, R., Inoue, A. K., MacKenty, J. W., O'Connell, R. W., and Silk, J. I. (2016). Hubble Space Telescope Wide Field Camera 3 Observations of Escaping Lyman Continuum Radiation from Galaxies and Weak AGN at Redshifts z˜2.3–5. *ArXiv e-prints*. 138

Songaila, A. (2001). The Minimum Universal Metal Density between Redshifts of 1.5 and 5.5. ApJ, 561:L153–L156. xxiii, 42, 195

Songaila, A., Bryant, W., and Cowie, L. L. (1989). Limits on the Galactic and cosmic ionizing fluxes from measurements of H-alpha emission from the high-velocity neutral hydrogen clouds. ApJ, 345:L71–L73. 138

Springel, V. (2005). The cosmological simulation code GADGET-2. MNRAS, 364:1105–1134. xvi, 30, 31, 37, 55, 56, 65, 73, 74, 190

Springel, V. and Hernquist, L. (2002). Cosmological smoothed particle hydrodynamics simulations: the entropy equation. MNRAS, 333:649–664. 31, 156

Springel, V., Yoshida, N., and White, S. D. M. (2001). GADGET: a code for collisionless and gasdynamical cosmological simulations. *New Astronomy*, 6:79–117. 30, 31

Srianand, R. and Khare, P. (1996). Analysis of Lyalpha absorption lines in the vicinity of QSOs. MNRAS, 280:767–780. 138

Stevans, M. L., Shull, J. M., Danforth, C. W., and Tilton, E. M. (2014). HST-COS Observations of AGNs. II. Extended Survey of Ultraviolet Composite Spectra from 159 Active Galactic Nuclei. ApJ, 794:75. 177

Storrie-Lombardi, L. J., McMahon, R. G., and Irwin, M. J. (1996). Evolution of neutral gas at high redshift: implications for the epoch of galaxy formation. MNRAS, 283:L79–L83. 4, 88

Sutherland, R. S. and Dopita, M. A. (1993). Cooling functions for low-density astrophysical plasmas. ApJS, 88:253–327. 18, 19, 60

Tegmark, M., Blanton, M. R., Strauss, M. A., Hoyle, F., Schlegel, D., Scoccimarro, R., Vogeley, M. S., Weinberg, D. H., Zehavi, I., Berlind, A., Budavari, T., Connolly, A., Eisenstein, D. J., Finkbeiner, D., Frieman, J. A., Gunn, J. E., Hamilton, A. J. S., Hui, L., Jain, B., Johnston, D., Kent, S., Lin, H., Nakajima, R., Nichol, R. C., Ostriker, J. P., Pope, A., Scranton, R., Seljak, U., Sheth, R. K., Stebbins, A., Szalay, A. S., Szapudi, I., Verde, L., Xu, Y., Annis, J., Bahcall, N. A., Brinkmann, J., Burles, S., Castander, F. J., Csabai, I., Loveday, J., Doi, M., Fukugita, M., Gott, III, J. R., Hennessy, G., Hogg, D. W., Ivezić, Ž., Knapp, G. R., Lamb, D. Q., Lee, B. C., Lupton, R. H., McKay, T. A., Kunszt, P., Munn, J. A., O'Connell, L., Peoples, J., Pier, J. R., Richmond, M., Rockosi, C., Schneider, D. P., Stoughton, C., Tucker, D. L., Vanden Berk, D. E., Yanny, B., York, D. G., and SDSS Collaboration (2004). The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey. ApJ, 606:702–740. 4, 87

Tepper-García, T., Richter, P., Schaye, J., Booth, C. M., Dalla Vecchia, C., and Theuns, T. (2012). Absorption signatures of warm-hot gas at low redshift: broad H I Lyα absorbers. MNRAS, 425:1640–1663. 55, 62, 116, 139, 140

Teyssier, R. (2002). Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES. A&A, 385:337–364. 34

Theuns, T., Bernardi, M., Frieman, J., Hewett, P., Schaye, J., Sheth, R. K., and Subbarao, M. (2002a). Detection of He II Reionization in the Sloan Digital Sky Survey Quasar Sample. ApJ, 574:L111–L114. 59

Theuns, T., Leonard, A., and Efstathiou, G. (1998a). The low-redshift evolution of the Lyman-alpha forest. MNRAS, 297:L49–L52. 55, 139

Theuns, T., Leonard, A., Efstathiou, G., Pearce, F. R., and Thomas, P. A. (1998b). P^3M-SPH simulations of the Lyalpha forest. MNRAS, 301:478–502. 15, 18, 21, 59, 60, 110, 112

Theuns, T. and Zaroubi, S. (2000). A wavelet analysis of the spectra of quasi-stellar objects. MNRAS, 317:989–995. 124

Theuns, T., Zaroubi, S., Kim, T.-S., Tzanavaris, P., and Carswell, R. F. (2002b). Temperature fluctuations in the intergalactic medium. MNRAS, 332:367–382. 124

Tripp, T. M., Sembach, K. R., Bowen, D. V., Savage, B. D., Jenkins, E. B., Lehner, N., and Richter, P. (2008). A High-Resolution Survey of Low-Redshift QSO Absorption Lines: Statistics and Physical Conditions of O VI Absorbers. ApJS, 177:39–102. xxv, 156, 196

Upton Sanderbeck, P. R., D'Aloisio, A., and McQuinn, M. J. (2016). Models of the thermal evolution of the intergalactic medium after reionization. MNRAS, 460:1885–1897. 128

Vanzella, E., de Barros, S., Vasei, K., Alavi, A., Giavalisco, M., Siana, B., Grazian, A., Hasinger, G., Suh, H., Cappelluti, N., Vito, F., Amorin, R., Balestra, I., Brusa, M., Calura, F., Castellano, M., Comastri, A., Fontana, A., Gilli, R., Mignoli, M., Pentericci, L., Vignali, C., and Zamorani, G. (2016). Hubble Imaging of the Ionizing Radiation from a Star-forming Galaxy at Z=3.2 with fesc>50%. ApJ, 825:41. xxiii

Vasei, K., Siana, B., Shapley, A. E., Quider, A. M., Alavi, A., Rafelski, M., Steidel, C. C., Pettini, M., and Lewis, G. F. (2016). The Lyman Continuum Escape Fraction of the Cosmic Horseshoe: A Test of Indirect Estimates. ApJ, 831:38. 138

Verner, D. A., Ferland, G., and Korista, K. (1994). New Atomic Data for Astrophysics. In *American Astronomical Society Meeting Abstracts*, volume 26 of *Bulletin of the American Astronomical Society*, page 1371. 6

Viel, M., Bolton, J. S., and Haehnelt, M. G. (2009). Cosmological and astrophysical constraints from the Lyman α forest flux probability distribution function. MNRAS, 399:L39–L43. xiv, xix, 4, 87

Viel, M. and Haehnelt, M. G. (2006). Cosmological and astrophysical parameters from the Sloan Digital Sky Survey flux power spectrum and hydrodynamical simulations of the Lyman α forest. MNRAS, 365:231–244. xiv, 4, 32, 36, 87

Viel, M., Haehnelt, M. G., Bolton, J. S., Kim, T.-S., Puchwein, E., Nasir, F., and Wakker, B. P. (2017). Diagnosing galactic feedback with line broadening in the low-redshift Lyα forest. MNRAS, 467:L86–L90. xxv, 25, 36, 53, 116, 129, 131, 155, 172, 173, 175, 178, 196

Viel, M., Haehnelt, M. G., and Lewis, A. (2006). The Lyman α forest and WMAP year three. MNRAS, 370:L51–L55. xix, 4, 87

Viel, M., Haehnelt, M. G., and Springel, V. (2004a). Inferring the dark matter power spectrum from the Lyman α forest in high-resolution QSO absorption spectra. MNRAS, 354:684–694. xix, 4, 53, 65, 87, 123

Viel, M., Haehnelt, M. G., and Springel, V. (2010). The effect of neutrinos on the matter distribution as probed by the intergalactic medium. *JCAP*, 6:015. 200

Viel, M., Matarrese, S., Mo, H. J., Theuns, T., and Haehnelt, M. G. (2002). Modelling the IGM and the Ly α forest at high redshift from the dark matter distribution. MNRAS, 336:685–698. 28

Viel, M., Schaye, J., and Booth, C. M. (2013). The impact of feedback from galaxy formation on the Lyman α transmitted flux. MNRAS, 429:1734–1746. xiv

Viel, M., Weller, J., and Haehnelt, M. G. (2004b). Constraints on the primordial power spectrum from high-resolution Lyman α forest spectra and WMAP. MNRAS, 355:L23–L28. xix, 4, 35, 53, 87

Vogel, S. N., Weymann, R., Rauch, M., and Hamilton, T. (1995). New upper limits on the local metagalactic ionizing radiation density. ApJ, 441:162–169. 138

Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G., Bird, S., Nelson, D., and Hernquist, L. (2014). Properties of galaxies reproduced by a hydrodynamic simulation. Nature, 509:177–182. 32

Wadsley, J. W., Stadel, J., and Quinn, T. (2004). Gasoline: a flexible, parallel implementation of TreeSPH. *New Astronomy*, 9:137–158. 32

Webb, J. K. and Carswell, R. F. (1991). The Lyman Alpha Forest b-N Controversy. In Shaver, P. A., Wampler, E. J., and Wolfe, A. M., editors, *Quasar Absorption Lines*, page 3. 157

Weinberg, D. H., Hernquist, L., and Katz, N. (1997). Photoionization, Numerical Resolution, and Galaxy Formation. ApJ, 477:8–20. 18, 60

Weinberg, D. H., Katz, N., and Hernquist, L. (1998). Simulating Cosmic Structure Formation. In Woodward, C. E., Shull, J. M., and Thronson, Jr., H. A., editors, *Origins*, volume 148 of *Astronomical Society of the Pacific Conference Series*, page 21. 4, 15, 24, 139

Weymann, R. J., Vogel, S. N., Veilleux, S., and Epps, H. W. (2001). A New Observational Upper Limit to the Low-Redshift Ionizing Background Radiation. ApJ, 561:559–572. 138

Wiersma, R. P. C., Schaye, J., and Smith, B. D. (2009). The effect of photoionization on the cooling rates of enriched, astrophysical plasmas. MNRAS, 393:99–107. liii, 19, 60, 134, 165, 166

Wise, J. H., Demchenko, V. G., Halicek, M. T., Norman, M. L., Turk, M. J., Abel, T., and Smith, B. D. (2014). The birth of a galaxy - III. Propelling reionization with the faintest galaxies. MNRAS, 442:2560–2579. 178

Worseck, G., Prochaska, J. X., Hennawi, J. F., and McQuinn, M. (2014). Early and Extended Helium Reionization Over More Than 600 Million Years of Cosmic Time. *ArXiv e-prints*. 3, 59

Zaldarriaga, M. (2002). Searching for Fluctuations in the Intergalactic Medium Temperature Using the Lyα Forest. ApJ, 564:153–161. 123, 124

Zaldarriaga, M., Hui, L., and Tegmark, M. (2001). Constraints from the Lyα Forest Power Spectrum. ApJ, 557:519–526. 25, 36, 53, 87, 123

Zaroubi, S. (2013). The Epoch of Reionization. In Wiklind, T., Mobasher, B., and Bromm, V., editors, *The First Galaxies*, volume 396 of *Astrophysics and Space Science Library*, page 45. 3

Zel'dovich, Y. B. (1970). Gravitational instability: An approximate theory for large density perturbations. A&A, 5:84–89. 25, 37

Zhan, H., Davé, R., Eisenstein, D., and Katz, N. (2005). Lyα flux power spectrum and its covariance. MNRAS, 363:1145–1154. 35, 36, 119, 123, 150

Zhang, Y., Anninos, P., and Norman, M. L. (1995). A Multispecies Model for Hydrogen and Helium Absorbers in Lyman-Alpha Forest Clouds. ApJ, 453:L57. xiv, 35, 53, 55, 139

Zheng, W., Kriss, G. A., Deharveng, J.-M., Dixon, W. V., Kruk, J. W., Shull, J. M., Giroux, M. L., Morton, D. C., Williger, G. M., Friedman, S. D., and Moos, H. W. (2004). A Study of the Reionization History of Intergalactic Helium with FUSE and the Very Large Telescope. ApJ, 605:631–644. 59