Simulating the distribution of cosmological neutral hydrogen over cosmic times

A thesis submitted towards partial fulfilment of

BS-MS Dual Degree Programme

by

KULKARNI MIHIR SANJAY

under the guidance of

DR. TIRTHANKAR ROY CHOUDHURY

NCRA-TIFR, PUNE

Indian Institute of Science Education and Research Pune

Certificate

This is to certify that this thesis entitled **Simulating the distribution of cosmological neutral hydrogen over cosmic times** submitted towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research Pune represents original research carried out by **Kulkarni Mihir Sanjay** at the **National Centre for Radio Astrophysics, Tata Institute of Fundamental Research, Pune**, under the supervision of **Dr. Tirthankar Roy Choudhury** during the academic year 2014-2015.

T. Roy Choudhury

Dr. Tirthankar Roy Choudhury
(Thesis Supervisor)

Place: Pune
Date: 25 March 2015

Declaration

I hereby declare that the matter embodied in the report entitled **Simulating the distribution of cosmological neutral hydrogen over cosmic scales** are the results of the investigations carried out by me at the **National Centre for Radio Astrophysics, Tata Institute of Fundamental Research, Pune**, under the supervision of **Dr. Tirthankar Roy Choudhury** and the same has not been submitted elsewhere for any other degree.

KULKARNI MIHIR SANJAY

(Reg. No. 20101068)

Place: Pune

Date: 25 March 2015

Acknowledgements

I thank my supervisor Dr. Tirthankar Roy Choudhury for allowing me to work him for my 5th year project. It was a truly gratifying experience. A part of the project work was done in collaboration with Dr. Aseem Paranjape, Inter University Centre for Astronomy and Astrophysics (IUCAA). He was very helpful during the work.

Thanks are due to National Centre for Radio Astronomy (NCRA) for providing academic and computing facilities. I would also like to thank teachers and friends at IISER and NCRA for their help and suggestions during the project. I am grateful to my family for their constant support.

Abstract

The reionization of the intergalactic medium is a paramount event in the evolution of the universe. In this thesis, we have studied the distribution of the neutral hydrogen in the universe at two different epochs i.e. post- reionization era and during the reionization. We have used cosmological N-body simulations to make numerical and semi-numerical calculations.

In introduction chapter, we introduce the concepts needed in the thesis. We describe basic cosmology, density perturbations, power spectrum and correlation functions. We also describe about the various methods used in cosmological N-body simulations and about different codes like GADGET-2, Friends-of-Friends, N-GenIC that we have used.

Chapter 2 deals with the study of the distribution of the neutral hydrogen in post-reionization era. Redshifted 21-cm signal and Lyman-α forest are two signals which we get mainly from galactic and intergalactic neutral hydrogen respectively. Though these signals arise from different astrophysical phenomena, they are expected to trace the underlying dark matter distribution at large scales. We calculate cross power spectrum using two methods i.e. Multi-frequency Angular Power Spectrum and using colour plots in ($k_{\parallel}$, $k_{\perp}$) plane.

In chapter 3, we describe the study about HII bubble distribution during the reionization. We first describe the analytical models that use 'excursion set formalism' to calculate it. We calculate the distribution using semi-numerical formalism and compare it with analytical calculations. We observed the presence of a large ionized bubble for particular parameters, which needs further investigation. We also visualized the ionized regions to check if they match with the halos and bubbles calculated.

In chapter 4, we summarise and describe the future work to be done with these projects.

Contents

Chapter 1

Introduction

1.1 Motivation of the thesis

According to current accepted model of cosmology, the universe began with a hot big bang. For nearly first 10^{-34} seconds universe expanded very rapidly, through a process known as inflation. Quantum fluctuations during this time were imprinted as density fluctuations and acted as seeds for structure formation. As the universe cooled down to nearly 10^9 K because of expansion, light atomic nuclei like Deuterium, Helium were formed though a process named as baryogenesis. The universe was ionized at this epoch and protons, electrons, photons were continuously interacting with each other. Nearly 100,000 years after ($z \sim 1100$), the universe cooled down to form neutral hydrogen. Photons decoupled from protons and electrons at this time. This epoch is called as 'epoch of recombination'. We observe these photons as cosmic microwave background today.

As the universe evolved, dark matter clustered and first stars and galaxies started to form at the centres of dark matter halos. High energy ultraviolet photons from these sources started ionizing the regions around them. As more stars and galaxies formed, the hydrogen in the universe was reionized gradually. This epoch is known as the 'epoch of reionization'. The observations today suggest that the reionization of the intergalactic medium was completed by redshift $z > 6$.

Reionization is an important transition in the universe. In the post-reionization era, the neutral hydrogen is found mainly in galaxies. There is small amount of neutral hydrogen present in the intergalactic medium. The neutral hydrogen in these two

regions can be probed using different methods, but is expected to trace the underlying dark matter distribution at large scales. We can get 21-cm signal from galactic neutral hydrogen and we get the Lyman-α forests arising from of intergalactic neutral hydrogen. Cross-correlation of these two signals has been suggested to be a probe of the dark matter distribution. Guha Sarkar et al. (2011); Guha Sarkar and Datta (2015) have presented analytical calculations of the same. We decided to calculate the cross power spectrum using cosmological N-body simulations.

The details of how reionization occurred are not very clear yet. Various models have been suggested for the mechanism of the reionization. Paranjape and Choudhury (2014) proposed an analytical model based on 'excursion set formalism' to calculate the size distribution of the ionized bubbles during the reionization. We have checked the consistency of this analytical model with the semi-numerical calculations involving N-body simulations.

In this chapter, we explain the basic theory about cosmology, density perturbations, power spectrum, N-body simulations that we use in next chapters. This information is widely available in textbooks.

1.2 Basic cosmology

Our universe is assumed to be homogeneous and isotropic at large scales. This is known as the Cosmological principle. Homogeneity means that the properties at each point are nearly same as the properties at some other point and isotropy means that the universe looks same in each direction we look. Friedmann-Robertson-Walker (FRW) metric can be used to describe such universe as

$$ds^2 = dt^2 - a(t)^2(\frac{dr^2}{1-kr^2} + r^2 d\Omega^2) \tag{1.1}$$

$d\Omega^2 = d\theta^2 + sin^2\theta d\phi^2$, which is the metric on a unit sphere. $a(t)$ is called 'scale factor' and is normalised such that $a(t_0) = 1$. (t_0 denotes current epoch). k is negative, zero and positive for hyperbolic geometry, flat geometry and spherical geometry respectively. The magnitude of k is of the order of $1/R_C^2$, where R_C is the radius of curvature. Redshift is defined using the scale factor as follows

$$1 + z = \frac{a(t_0)}{a(t)} = \frac{\lambda_o}{\lambda_e} = \frac{\nu_e}{\nu_o} \tag{1.2}$$

The evolution of $a(t)$ is given by Friedmann equations. Hubble parameter $H(t)$ is defined as follows, where $\dot{a}$ is time derivative of the scale factor.

$$H(t) = \frac{\dot{a}}{a} \tag{1.3}$$

The Hubble parameter at current epoch is generally denoted as $H_0 = 100h$ km/s/Mpc, where h lies between $0.5 - 1.0$. Recent observations have calculated h to be 0.673 (Planck Collaboration et al., 2014). If we assume that the universe is filled with an ideal fluid following pressure-density relation $p = w\rho$, where w is called the equation of state, we get an expression for $\rho(t)$.

$$\rho = \rho_0 a^{-3(1+w)} \tag{1.4}$$

The critical density of the universe is defined as follows:

$$\rho_{0c} = \frac{3H_0^2}{8\pi G} \tag{1.5}$$

The density parameter is defined as the ratio of the density of the fluid with the critical density.

$$\Omega_a = \frac{\rho_a}{\rho_c} \tag{1.6}$$

The curvature density parameter is defined as

$$\Omega_k = 1 - (\Omega_r + \Omega_m + \Omega_\Lambda) \tag{1.7}$$

where Ω_r, Ω_m, Ω_Λ are the density parameters for radiation, matter and cosmological constant. Ω_Λ is defined as

$$\Omega_\Lambda = \frac{\Lambda c^2}{3H_0^2} \tag{1.8}$$

The Hubble parameter varies with time as

$$H(t) = H_0 \left[\sum_i \Omega_{0i} a^{-3(1+w_i)} \right]^{0.5} \tag{1.9}$$

Comoving distance is the distance between the objects which remains unchanged because of Hubble flow. The measured distance at a particular epoch is called proper distance and is related to the comoving distance by scale factor. The Hubble distance is defined as

$$d_H(z) = \frac{c}{H(z)} \tag{1.10}$$

The line-of-sight comoving distance is calculated as follows:

$$d_C(z) = \int_0^z dz d_H(z) = \int_0^z \frac{cdz}{H_0[\sum_i \Omega_{0i} a^{-3(1+w_i)}]^{0.5}} \tag{1.11}$$

1.3 Density perturbations

The structure (galaxies and clusters) forms on scales much smaller than the Hubble distance $d_H(z)$. Hence, the relativistic effects can be ignored and Newtonian physics can be used. Assuming the dark matter and baryons to be perfect fluids, we can write following equations:

$$\dot{\rho}(t,\boldsymbol{r}) = \nabla_{\boldsymbol{r}}[\rho(t,\boldsymbol{r})\boldsymbol{V}(t,\boldsymbol{r})] \tag{1.12}$$

$$\dot{\boldsymbol{V}}(t,\boldsymbol{r}) + [\boldsymbol{V}(t,\boldsymbol{r}\cdot\nabla_{\boldsymbol{r}}]\boldsymbol{V}(t,\boldsymbol{r}) = -\nabla_{\boldsymbol{r}}\Phi(t,\boldsymbol{r}) - \frac{\nabla_{\boldsymbol{r}}P(t,\boldsymbol{r})}{\rho(t,\boldsymbol{r})} \tag{1.13}$$

$$\nabla^2_{\boldsymbol{r}}\Phi(t,\boldsymbol{r}) = 4\pi G\rho(t,\boldsymbol{r}) \tag{1.14}$$

The equations (1.12), (1.13) and (1.14) are called continuity equation, Euler equation and Poission equation respectively. Here, the overdot represents partial derivative with respect to time t and $\nabla_{\boldsymbol{r}}$ denotes spatial divergence with proper coordinates $\boldsymbol{r}$. $\rho(t,\boldsymbol{r})$, $P(t,\boldsymbol{r})$ and $\Phi(t,\boldsymbol{r})$ denote the density, pressure and gravitational potential respectively. $V(t,\boldsymbol{r}) \equiv d\boldsymbol{r}/dt$ is the proper velocity of the fluid.
Using comoving coordinates $\boldsymbol{x}$, defined as

$$\boldsymbol{r} = a(t)\boldsymbol{x} \tag{1.15}$$

and perturbed quantities as

$$\text{Density contrast}: \delta(t,\boldsymbol{x}) = \frac{\rho(t,\boldsymbol{x})}{\bar{\rho}(t)} - 1 \tag{1.16}$$

$$\text{Peculiar velocity field}: \boldsymbol{v}(t,\boldsymbol{x}) \equiv a(t)\frac{d\boldsymbol{x}}{dt} = \boldsymbol{V}(t,\boldsymbol{x}) - \frac{\dot{a}}{a}\boldsymbol{r} \tag{1.17}$$

$$\text{Perturbed gravitational field}: \phi(t,\boldsymbol{x}) = \Phi(t,\boldsymbol{x}) - \bar{\Phi}(t) \tag{1.18}$$

the fluid equations (1.12), (1.13) and (1.14) reduce to the following equations.

$$\dot{\delta} + \frac{1}{a}\nabla\cdot[(1+\delta)\boldsymbol{v}] = 0 \tag{1.19}$$

$$\dot{\boldsymbol{v}} + \frac{\dot{a}}{a}\boldsymbol{v} + \frac{1}{a}(\boldsymbol{v}\cdot\nabla)\boldsymbol{v} = -\frac{1}{a}\nabla\phi - \frac{\nabla p}{a\bar{\rho}(1+\delta)} \tag{1.20}$$

$$\nabla^2\phi = 4\pi G\bar{\rho}a^2\delta \tag{1.21}$$

Here ∇ denotes the spatial gradient with comoving coordinates $\boldsymbol{x}$. Since the dark matter is collisionless, we assume $p = 0$ for it. In the linear perturbation theory, we retain only first order terms in $\boldsymbol{x}$ and $\boldsymbol{v}$ and ignore higher order terms. Simplifying the equations gives

us a second order differential equation in time for the dark matter density contrast δ_{DM}. Since this involves just time derivatives of δ_{DM}, a solution of the form $\delta_{DM} = D(t)f(\boldsymbol{x})$ exist, where $f(\boldsymbol{x})$ is a function of spatial coordinates depending on the initial configuration and $D(t)$ is a time dependent function. Hence we get

$$\ddot{D} + 2\frac{\dot{a}}{a}\dot{D} = \frac{3}{2}H_0^2\Omega_m\frac{D}{a^3} \tag{1.22}$$

This equation has two solutions. One of them decays with time and turns out to be the Hubble parameter $H(a)$. The other solution grows with time and is given by

$$D(a) = H(a)\int^a \frac{da'}{H^3(a')a'^3} \tag{1.23}$$

This is known as growth factor. Normally it is normalised such that $D(t_0) = 1$. This can be used as a new time variable. A new velocity is defined as

$$\boldsymbol{w} \equiv \frac{d\boldsymbol{x}}{dD} = \frac{\boldsymbol{v}}{a\dot{D}} \tag{1.24}$$

It is very difficult to solve the dark matter density perturbations using complete non-linear theory. In the quasi-linear approximation, we assume that the dark matter fluid flows in a force-free field. The comoving coordinates and the velocities are given as

$$\boldsymbol{w}_{DM}[\boldsymbol{x}(\boldsymbol{q}, D), D] = -\nabla_{\boldsymbol{q}}\psi_0(\boldsymbol{q}), \qquad \boldsymbol{x}_{DM}(\boldsymbol{q}, D) = \boldsymbol{q} - D\nabla_{\boldsymbol{q}}\psi_0(\boldsymbol{q}) \tag{1.25}$$

where $\psi \equiv \frac{2}{3H_0^2\Omega_m}\frac{a}{D}\phi$ and $\boldsymbol{q}$ is the Lagrangian coordinate. This approximation is also called as *Zel'dovich approximation*. This approximation is used while generating initial conditions for N-body simulations.

1.4 Power spectrum

The matter density contrast $\delta(\boldsymbol{x}, t)$ is defined in equation (1.16). Its Fourier component is defined as

$$\delta(\boldsymbol{k}, t) = \int_{-\infty}^{\infty} d^3\boldsymbol{x}\delta(\boldsymbol{x}, t)e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \tag{1.26}$$

where $\boldsymbol{k}$ denotes the wave vector.

The density contrast is generated using a stochastic random process. In general, we need to define all the moments of a stochastic process to describe it completely. We assume that that the density contrast is a gaussian random field. This assumption makes

calculations very easy, does not contradict current observations and is predicted from inflationary theory. A gaussian random field is defined using two parameters viz. its mean and variance. The mean of the density contrast is zero.

$$\bar{\delta}(\boldsymbol{x},t) = \overline{\frac{\rho(\boldsymbol{x},t)}{\bar{\rho}} - 1} = \frac{\bar{\rho}}{\bar{\rho}} - 1 = 1 - 1 = 0 \tag{1.27}$$

The power spectrum of the density contrast field is defined as

$$\langle \delta(\boldsymbol{k})\delta^*(\boldsymbol{k}')\rangle = (2\pi)^3\delta_D(\boldsymbol{k}-\boldsymbol{k}')P(k) \tag{1.28}$$

where $\langle\rangle$ denotes the average over the ensemble of the random field. This is conceptually different from volume average, which is calculated for a single realization. The volume average and ensemble average are the same if the process is ergodic. Note that the power spectrum depends only on the magnitude k as we have assumed universe to be isotropic. The power spectrum is the only quantity required to define the gaussian random field $\delta(\boldsymbol{x},t)$. Higher $P(k)$ means having more structure on the scales corresponding to that particular k-mode. We define two point auto-correlation function of $\delta(\boldsymbol{x})$ as

$$\xi(r) = \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{r})\rangle_{\boldsymbol{x}} \tag{1.29}$$

where $\langle\rangle_{\boldsymbol{x}}$ denotes the average over $\boldsymbol{x}$, which is same as the average over ensembles because of ergodicity. Similar to power spectrum, the auto-correlation function depends on r and not on $\boldsymbol{r}$ because of isotropy. The auto-correlation will be zero if $\delta(\boldsymbol{x})$ and $\delta(\boldsymbol{x}+\boldsymbol{r})$ are perfectly random, because of averaging over ensemble. If $\delta(\boldsymbol{x})$ and $\delta(\boldsymbol{x}+\boldsymbol{r})$ have excess similarity over randomness, then the auto-correlation function will be positive. We call the regions separated by r to be correlated if $\xi(r) > 0$ and anti-correlated if $\xi(r) < 0$. Wiener-Khinchin theorem states that the power spectrum and the auto-correlation function are Fourier transform pairs

$$\xi(r) = \frac{1}{(2\pi)^3}\int d^3\boldsymbol{k}P(k)e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \tag{1.30}$$

As $\xi(r)$ is dimensionless and $d^3\boldsymbol{k}$ has dimensions of Vol^{-1}, $P(k)$ has dimensions of of Vol.

$$\frac{d^3\boldsymbol{k}P(k)}{(2\pi)^3} = \frac{4\pi k^2 dkP(k)}{(2\pi)^3} = \frac{dk}{k}\frac{k^3P(k)}{2\pi^2} = dlnk\Delta^2(k) \tag{1.31}$$

Hence, we define the dimensionless power spectrum $\Delta^2(k) = k^3P(k)/2\pi^2$ for 3 dimensions. Similar to 3D power spectrum, we can calculate 2D and 1D power spectra if the real space

input is defined in 2 or 1 dimensions. e.g. We define 1D power spectrum for Lyman-α forests as they are sampled only along an axis. Dimensionless power spectra for 2D and 1D are defined as $k^2P(k)/2\pi$ and $kP(k)/\pi$ respectively.

RMS density fluctuation smoothed over scale R is defined as

$$\sigma^2(R) = \int_0^\infty \frac{dk}{k}\frac{k^3P(k)}{2\pi^2}W^2(kR) = \int_0^\infty dlnk\Delta^2(k)W^2(kR) \tag{1.32}$$

where $W(kR)$ denotes spherical top-hat window function in Fourier space defined as

$$W(kR) = \frac{3(\sin kR - kR\cos kR)}{(kR)^3} \tag{1.33}$$

RMS fluctuations at scale of $8h^{-1}$ Mpc are defined as $\sigma_8 \equiv \sigma(R = 8h^{-1}\text{Mpc})$. Its magnitude is of order of unity and it is used for the normalization of the power spectrum.

1.5 Cosmological N-body simulations

Cosmological N-body simulations are very useful to understand the structure formation. We start with initial conditions with tiny perturbations and allow particles to evolve under influence of the gravitational field. We use dark matter only numerical simulations to understand how structure like dark matter halos, filaments form in the universe. The equations and theory in this section is based on Bagla (1996, 2001). The dynamical evolution of the gravitationally interacting particles is given by

$$\begin{aligned} \ddot{\boldsymbol{r}}_i &= -\nabla_{\boldsymbol{r}_i}\Phi \\ \nabla^2_{\boldsymbol{r}}\Phi &= 4\pi G\rho \\ \rho(\boldsymbol{r}) &= \sum_i m_i\delta^3_D(\boldsymbol{r}-\boldsymbol{r}_i) \end{aligned} \tag{1.34}$$

where $\boldsymbol{r}_i$ denotes the position vector in proper coordinates of the i-th particle. Using the comoving coordinates $\boldsymbol{x} = \boldsymbol{r}/a$ and doing some algebraic manipulations, we get

$$\ddot{\boldsymbol{x}} + 2\frac{\dot{a}}{a}\dot{\boldsymbol{x}} = -\frac{1}{a^2}\nabla_{\boldsymbol{x}}\phi \tag{1.35}$$

where ϕ is perturbed potential and satisfies equation (1.21). There are various methods to study evolution of collision-less dark matter particles. The important methods are mentioned below.

1.5.1 Simulation methods

1.5.1.1 Direct sum method (PP)

Direct sum method is also called particle-particle method (PP). In this method we calculate the forces on each particle from all other particles and add them. This method is very straight-forward and does not use any approximation, but is computationally very expensive. This method can be used only for small number of particles (10^4), as its complexity is $\mathcal{O}(N_p^2)$ where N_p is the number of particles.

1.5.1.2 Particle Mesh (PM)

Equation (1.21) which needs to be solved to calculate the perturbed potential takes the following form in the Fourier space.

$$\phi_{\boldsymbol{k}} = \frac{4\pi G\bar{\rho}(t)a^2\delta_{\boldsymbol{k}}}{k^2} \tag{1.36}$$

where $\phi_{\boldsymbol{k}}$ is the Fourier mode of the perturbed component of the gravitational potential, δ_k is the Fourier mode of the density contrast and k is the corresponding spatial frequency. We can see that the second order differential equation in real space is a simple algebraic equation in the Fourier space. From density contrast values in real space, we can calculate their Fourier components which can be used to calculate the Fourier components of the perturbed potential. Inverse Fourier transform of those gives us the perturbed potential in the real space. The complexity of the process is $\mathcal{O}(N_p log N_p)$ if Fast Fourier Transform (FFT) is used for calculating Fourier components. FFT algorithm can be used on uniformly spaced grids. Hence, we assign particles on the uniform grid. The main disadvantage of this method is that the force calculated because of close particles is incorrect, as it assumes minimum particle distance to be the grid separation.

1.5.1.3 P^{3}M Method

Particle mesh method can be used very effectively to calculate long range forces at very small computational cost. However, it calculates short range forces incorrectly. P^{3}M method (particle-particle particle-mesh) is a method, where long range forces are calculated using PM method and short range corrections are made using direct summation method. P^{3}M method was the first method which allowed N-body simulations with large number of particles.

1.5.1.4 Tree method

In this method, force on a particle is calculated by arranging other particles in tree structure. Force by sufficiently distant particle groups is calculated assuming them as a single entity, instead of calculating separately. This method is relatively computationally expensive and can give large errors for nearly uniform particle distributions (early universe).

1.5.1.5 TreePM method

This method is very similar method to P^3M method. The long range forces are calculated in the Fourier space (PM method) and short range forces are calculated using tree method. It uses best features of the tree and PM method.

1.5.2 GADGET-2

GADGET-2 is publicly available cosmological N-body simulation code written by Springel (2005). GADGET-2 was used in this project using TreePM method and periodic boundary conditions. Periodic boundary conditions ensure that the simulation box is the representative sample of the universe and also avoids edge effects.

GADGET-2 starts with an initial condition of the particles and evolves them gravitationally with time. It returns shapshot files (binary file format) at specified redshifts which contains header and position and velocity of each particle.

1.5.3 Initial conditions

We use N-GenIC code written by Volker Springel (2003) for calculating the initial conditions for GADGET-2 simulations. The initial conditions are calculated using Zel'dovich approximation (section 1.25). Zel'dovich approximation is valid in the quasi-linear regime. We generate the initial conditions at high redshift ($z \sim 100$). N-GenIC returns initial particle positions and velocities in a format readable by GADGET-2.

1.5.4 Cloud-in-cell

Cloud-in-cell (CIC) is an algorithm used to assign particles on the uniform grid. Each particle is associated with some weights to neighbouring 8 grid points. The algorithm

used was as follows. Along each axis, calculate the distance along that axis with the nearest two grid points. Give the weights to the grid points proportional to the distance of the other grid point. e.g. if the distances of a particle are 0.2 and 0.8 from two grid points in units of grid separation, then the weights given to those grid points will be proportional to 0.8 and 0.2 respectively. The weights are later normalised such that the sum of the weights on the 8 grid points equals the particle mass.

1.5.5 Friends of Friends algorithm

'Friends of Friends' (FoF) (Davis et al., 1985) is an algorithm used mainly for halo finding. For this project, we used FoF code written by Volker Springel. A friend is defined as a particle which is closer to the particle than a particular linking length (typically 0.2 times the inter-particle distance). This algorithm identifies friends and friends of friends and identifies them as a single group called a halo. We use FoF to identify dark matter halos in the GADGET-2 simulation.

Chapter 2

Cross-correlation of redshifted 21-cm signal and Lyman-α forest

2.1 Motivation

Neutral hydrogen (HI) emits line corresponding to hyperfine transition in the ground state which corresponds to wavelength 21-cm. HI lines have been observed from Milky Way and nearby galaxies and are used to calculate rotation curves which gives an indirect evidence of the dark matter. Sources of cosmological HI emission in the post-reionization era ($z < 6$) are mainly Damped Lyman-α Absorbers (DLAs), which have very high amount of neutral hydrogen present in them. This HI 21-cm signal is expected to trace underlying dark matter distribution at the large scales. However, the intensity of the 21-cm signal is very low because of the large distance of sources. Another major hurdle for its direct measurement is the abundant foregrounds resulting from synchrotron emission, free-free emission and extragalactic point sources which fall in the same frequency range and are very strong compared to the background signal.

The diffused neutral hydrogen in the highly ionized intergalactic medium produces absorption lines in the spectra of background quasars. This can be used to calculate the neutral hydrogen distribution along the line of sight towards the background quasar. The quasar spectra look like forests because of multiple Lyman-α absorption lines, hence are named as Lyaman-α forest. The Lyman-α forest flux underlies dark matter density on the large scale according the fluctuating Gunn-Peterson approximation. Though, redshifted 21-cm signal and the Lyman-α forest arise from two completely different astrophysical

phenomena, both are expected to trace the underlying dark matter distribution at the large scale. Lyman-α forest spectra have relatively much less foreground contaminations. Cross-correlation of these two signals is a novel probe which can be used to calculate the matter power spectrum, to estimate cosmological parameters and constraints on the dark energy, and does not have foreground contamination as severe as only HI 21-cm signal.

Previous studies have analytically calculated the estimated cross power spectrum of redshifted 21-cm signal and Lyman-α forest (Guha Sarkar et al., 2011; Guha Sarkar and Datta, 2015). In this project, we calculate the cross-power spectrum using cosmological N-body simulations.

2.2 HI 21-cm signal

2.2.1 Introduction

The ground state of the neutral hydrogen (1S) is split between two states because of spin-spin coupling. The lower energy state corresponds to spins of proton and electron being anti-parallel to each other and the higher energy state corresponds to spins of proton and electron being parallel to each other. The transition of the electron from higher energy state to lower energy state is called hyperfine transition and corresponds to wavelength of 21 cm. The Einstein coefficient for this transition is $A_{ij} = 2.85 \times 10^{-15}\text{sec}^{-1}$. The transition is very weak and has lifetime of 1.1×10^7 years. This signal very important astronomy because of large quantities of neutral hydrogen present in the universe. It has been observed from galactic as well as extragalactic sources. HI emission from galaxies can be used to calculate rotation curves, which provides an evidence of the dark matter in the galaxy.

In the post-reionization era ($z < 6$), the main source of HI 21-cm signal is Damped Lyman-α Absorbers (DLAs), which are gaseous objects having high column density of neutral hydrogen and damp the spectra of background quasars. We use following method to identify the regions of neutral hydrogen from the GADGET-2 simulation.

2.2.2 Identifying 21-cm sources from simulations

We used two low resolution and a high resolution GADGET-2 simulations. The low resolution simulations had 256^3 particles with box sizes 25 h^{-1} Mpc and 150 h^{-1} Mpc. The high resolution simulation had 1024^3 particles with box size of 75 h^{-1} Mpc. We used WMAP-5 parameters for these simulations. ($\Omega_m = 0.26$, $\Omega_\Lambda = 0.74$, $\sigma_8 = 0.79$, $h = 0.72$) (Komatsu et al., 2009). We use snapshot at redshift $z = 3$ for these calculations, as quasar population peaks near redshift 2-3. We generate halo catalogues using friends-of-friends (FoF) with linking length $b = 0.2$.

We used the method given in Bagla et al. (2010) to fill the dark matter halos with the neutral hydrogen (HI). Most of the bulk neutral hydrogen is present in the dark matter halos, which remains unionized because of self-shielding. Halos with very small mass will have very less neutral hydrogen as self-shielding will not be sufficient to prevent the ionization. It is observed that very high mass galaxy clusters have very less amount of neutral hydrogen present in them. To fill halos with neutral hydrogen, we use the first model described in Bagla et al. (2010). Here we put strict upper and lower bounds on the masses of halos that can have neutral hydrogen.

$$f(M) = \begin{cases} f_1, & M_{min} < M < M_{max} \\ 0, & \text{For all other masses} \end{cases} \tag{2.1}$$

Where $f(M)$ is the ratio of HI mass in halo and the halo mass, where M denotes the mass of the halo. f_1 is a constant which is calculated using the normalization $\Omega_{\rm HI} = 0.001$. We use minimum mass and maximum mass limit for the halo to be $1.25 \times 10^9 M_\odot$ and $3.7 \times 10^{11} M_\odot$ at redshift 3. These masses correspond to rotational velocities of 30 m/s and 200 m/s respectively. We then identify the dark matter particles lying in the halos of the specified mass range. We calculate the density at each grid point using only these dark matter particles using cloud-in-cell. We calculate density contrast for this and call it neutral hydrogen density contrast. We calculate the brightness temperature of the HI gas using the following formula from Bagla et al. 2010.

$$\delta T_b(z) = 4.6\ mK \times x_{HI}\,(1+\delta)\left(1 - \frac{T_{cmb}}{T_s}\right)(1+z)^2 \frac{H_0}{H(z)}\left[\frac{H(z)}{(1+z)(dv_\parallel/dr_\parallel)}\right] \tag{2.2}$$

where, x_{HI} is the neutral hydrogen fraction of baryons, δ is the HI density contrast, T_{cmb} is the cosmic microwave background temperature and T_s is the spin temperature calculated

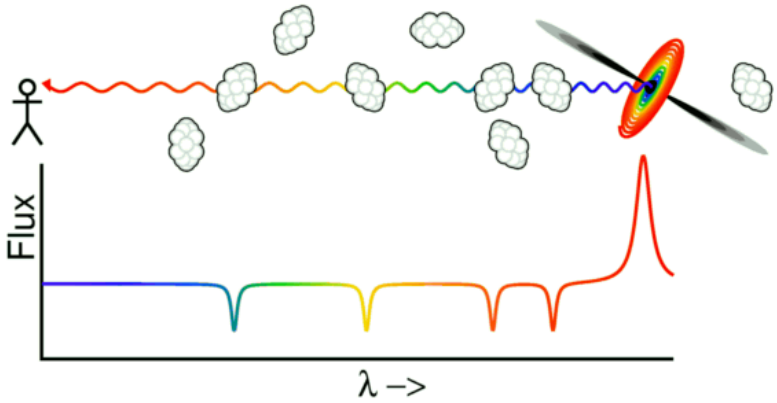

Figure 2.1: *This figure shows a schematic representation of the quasar spectrum and Lyman-α absorption lines by the intervening neutral hydrogen. The absorption line is seen at higher wavelength for neutral hydrogen at higher redshift. Image source: Wright's webpage.*

from relative abundances of HI in the ground state and the excited state. Typically, $T_s \gg T_{cmb}$ at the epoch of our interest, hence we ignore the term $(1 - T_{cmb}/T_s)$. The last term $H(z)/[(1+z)(dv_{\parallel}/dr_{\parallel})]$ accounts for the redshift-space distortions occurring because of peculiar velocities along the line of sight. However, we have not included these distortions in our current work and we ignore that term.

2.3 Lyman-α forest

2.3.1 Introduction

Lyman-α forests arise because of absorption lines by the intervening neutral hydrogen in the spectra of background quasars. Quasars are extremely luminous objects at high redshifts, which are compact objects unlike galaxies. Quasars are subclass of objects called as the active galactic nuclei (AGN) that are powered by accretion of matter around supermassive blackhole at the centres of distant galaxies. Quasars contain strong emission lines like Lyman-α emission line.

If some neutral hydrogen is present along the line of sight of the quasar, it will absorb its photons at $\lambda_{12} = 1215$Å in its rest frame undergoing Lyman-α transition. However, because of cosmological redshift, we will observe the absorption line at $\lambda = \lambda_{12}(1+z)$. Hence, we will get absorption lines by neutral hydrogen at different redshifts to be at different wavelengths/frequencies, even though the transition is same i.e. Lyman-α ($n = 1$ to $n = 2$). Figure 2.1 shows a schematic diagram of Lyman-α absorption lines in the

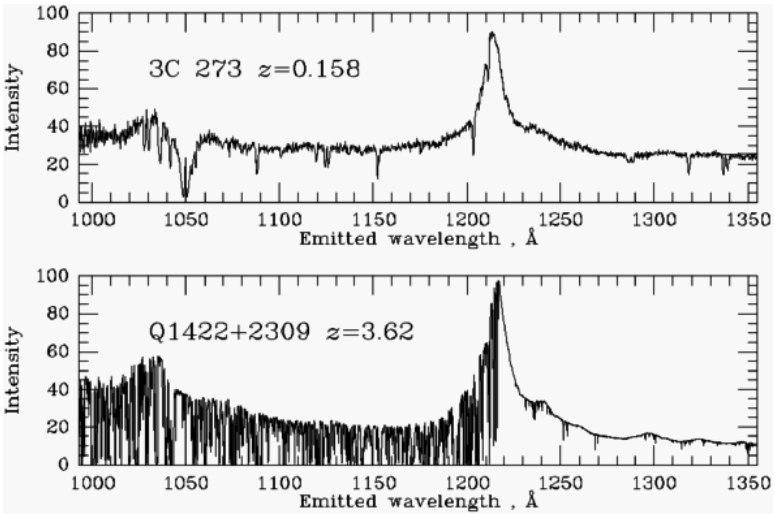

Figure 2.2: *This figure shows observed Lyman-α forests for two quasars at redshifts 0.158 and 3.62 respectively. Image source: Wright's webpage.*

quasar spectrum. Figure 2.2 shows two actual Lyman-α forests from quasars at two different redshifts.

It was predicted by Gunn and Peterson in 1965 that the neutral hydrogen in the intergalactic medium will mostly absorb the photons from quasars causing a trough in the quasar spectra, today called as Gunn-Peterson trough. This was first observed in quasar spectra at $z > 6$ in Sloan Digital Sky Survey (SDSS) nearly 3 decades after its inception. We know that neutral hydrogen fraction as small as 10^{-4} is sufficient to absorb the photons completely. The presence of Gunn-Peterson trough in quasar spectra at $z > 6$ tells us that the reionization was complete before redshift 6.

2.3.2 Generation of Lyman-α forests

The theory and equations in this subsection are mainly based on Choudhury et al. (2001); Choudhury (2003). We have used these equations for the generation of Lyman-α forests and hence have included here for completeness.
The quasar flux after an absorption is given as

$$\mathcal{F} = \mathcal{F}_0 e^{-\tau} \tag{2.3}$$

where, $\mathcal{F}_0$ is the flux of continuum and τ is the optical depth. The optical depth of Lyman-α absorption by neutral hydrogen at redshift z at observed frequency ν_0 is given

by

$$\tau(\nu_0) = c \int dt \sigma_\alpha[\nu_0(1+z)] n_{\rm HI}(z) = \int_{z_{min}}^{z_{max}} dz \frac{d_H(z)}{1+z} \sigma_\alpha[\nu_0(1+z)] n_{\rm HI}(z) \tag{2.4}$$

where $d_H(z)$ is the Hubble distance at redshift z, $\sigma_\alpha(\nu)$ is the Lyman-α absorption cross-section and $n_{\rm HI}(z)$ is the neutral hydrogen number density at redshift z. Here the effect of scattering is neglected as it is small compared to the absorption in the intergalactic medium. The absorption cross-section is given by

$$\sigma(\nu) = \frac{I_\alpha c}{b\sqrt{\pi}} V\left(\alpha, \frac{c(\nu - \nu_\alpha)}{b\nu_\alpha}\right) \tag{2.5}$$

Here, $I_\alpha = 4.45 \times 10^{-18}$ cm^2 is a constant, b is the velocity dispersion of the intergalactic medium and $V(\alpha, \Delta v/b)$ is called Voigt profile. The shape of the line is determined by Voigt profile, which is convolution of Lorentzian profile and Gaussian profile. The Lorentzian profile arises because of natural broadening of line and is given as

$$\frac{\Gamma/4\pi^2}{(\nu - \nu_\alpha)^2 + (\Gamma/4\pi)^2} \tag{2.6}$$

where Γ is the natural line-width. The Gaussian profile arises because of thermal broadening and is given by

$$\frac{e^{-v^2/b^2}}{b\sqrt{\pi}} \tag{2.7}$$

The convolved Voigt profile is given as

$$V(\alpha, \frac{\Delta v}{b}) = \frac{\alpha b}{\pi} \int_{-\infty}^{\infty} du \frac{e^{-v^2/b^2}}{(\Delta v - u)^2 + (\alpha b)^2}; \alpha := \frac{\Gamma c}{4\pi\nu_\alpha b} \tag{2.8}$$

Here we have used $(\nu - \nu_\alpha)/\nu_\alpha = \Delta v/c$.

We assume the baryonic density field to follow the dark matter density. Once we know the baryonic density, we can calculate the neutral hydrogen fraction $x_{\rm HI}$ using the ionization equilibrium in the intergalactic medium. We assume that the rate of formation of neutral hydrogen because of recombination of protons and electrons is same as rate of ionization of neutral hydrogen because of collisions and photo-ionization. The equilibrium equation for hydrogen is given as

$$\alpha(T) n_p n_e = \Gamma_{ci}(T) n_e n_{\rm HI} + J n_{\rm HI} \tag{2.9}$$

where $\alpha(T)$ is the recombination rate of proton and electron, $\Gamma_{ci}(T)$ is the collision rate of the electron and the neutral hydrogen and J is the photoionization rate of the neutral

hydrogen. The neutral hydrogen fraction is given by

$$x_{\rm HI} = \frac{n_{\rm HI}}{n_B} = \frac{n_{\rm HI}}{n_{\rm HI} + n_p} \tag{2.10}$$

Using equations (2.9) and (2.10) and ignoring the helium number density contribution, we get

$$x_{\rm HI}(x,z) = \frac{\alpha(T(x,z))}{\alpha(T(x,z)) + \Gamma_{ci}(T(x,z)) + J(z)/n_e(x,z)}. \tag{2.11}$$

Expressing n_e in terms of n_B as $n_e/n_b \equiv \mu_e$, we get,

$$n_{\rm HI}(x,z) = \frac{\alpha(T(x,z))n_B(x,z)}{\alpha(T(x,z)) + \Gamma_{\rm ci}(T(x,z)) + J(z)/(\mu_e n_B(x,z))} \tag{2.12}$$

The temperature of the intergalactic medium is calculated as

$$T(x,z) = T_0(z)\left[\frac{n_B(x,z)}{n_0(z)}\right]^{\gamma-1} \tag{2.13}$$

where $T_0(z)$ is the mean IGM temperature at redshift z, γ is polytropic index and $n_0(z)$ is the mean baryonic density at redshift z given as

$$n_0(z) = \frac{\Omega_B \rho_c}{\mu_B m_p}(1+z)^3 \tag{2.14}$$

where Ω_B is baryon density parameter, ρ_c is the critical density of the universe and $\mu_B m_p$ is the mass per baryonic particle. The velocity dispersion used in equation (2.8) is calculated as

$$b[x, z(x)] = \sqrt{\frac{2k_B T(x, z(x))}{m_p}} \tag{2.15}$$

Using equations (2.12) and (2.15), we can calculate $n_{\rm HI}(x,z)$ and $b[x,z(x)]$, which can be used to calculate the optical depth in equation (2.4). For generating Lyman-α forests, we used $\gamma = 1.01$, $T_0 = 2.66 \times 10^4$ K. Figure 2.3 shows a Lyman-α forest generated by shooting a line of sight in a random direction.

2.4 Cross-correlation

Similar to power spectrum, we define the cross power spectrum as the Fourier transform of the cross-correlation function. Here we calculate the cross power spectrum of 21-cm brightness temperature and Lyman-α flux contrast. We define the Lyman-α flux contrast as

$$\tilde{\delta}_{\mathcal{F}} = \frac{\mathcal{F}}{\bar{\mathcal{F}}} - 1 \tag{2.16}$$

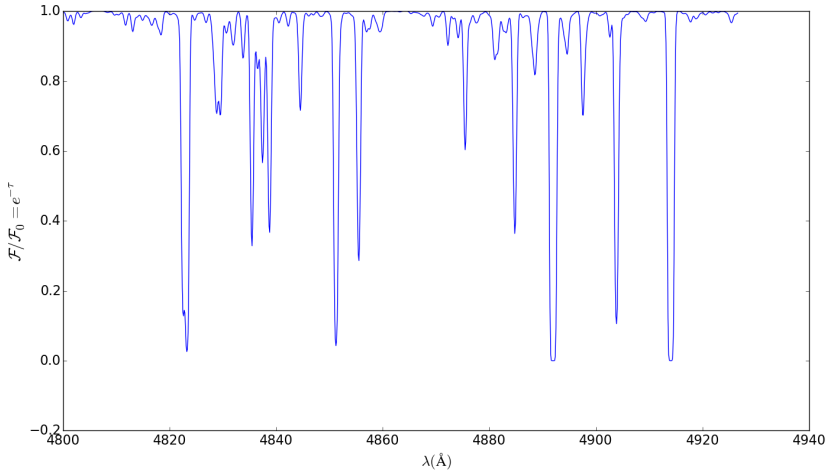

Figure 2.3: *The figure shows a Lyman-α forest generated by shooting a random line of sight in the GASGET-2 simulation of* 1024^3 *particles with box size 75* h^{-1} *Mpc.*

where $\mathcal{F}$ is the Lyman-α flux and $\bar{\mathcal{F}}$ is the mean flux. We can observe Lyman-α forests only along few lines of sight where quasars are present. We define the quasar sampling function as $\rho(\vec{\theta}) = N^{-1} \sum_n \delta_D^2(\vec{\theta} - \vec{\theta_n})$. We define sampled Lyman-α flux contrast as

$$\delta_{\mathcal{F}0}(\vec{\theta}) = \rho(\vec{\theta}) \delta_{\mathcal{F}}(\vec{\theta}) \tag{2.17}$$

We calculate the cross power spectrum in two methods. In one method we calculate the Multi-frequency Angular Power Spectrum (MAPS) and in the other one, we calculate 3D power spectrum and plot it against $k_\parallel$ and $k_\perp$.

2.4.1 Multi-frequency Angular Power Spectrum

We calculated the Multi-frequency Angular Power Spectrum (MAPS) following Datta et al. (2007); Guha Sarkar et al. (2011) We consider a sufficiently small field of view such that it can be assumed flat. The unit vector along a line of sight in the field of view can be written as $\hat{\boldsymbol{n}} = \hat{\boldsymbol{m}} + \vec{\theta}$ where $\hat{\boldsymbol{m}}$ is the unit vector along the line of sight towards the centre of the field of view and $\vec{\theta}$ is a 2D vector in the plane perpendicular to $\hat{\boldsymbol{m}}$. We decompose $\delta(\vec{\theta}, z)$ into Fourier modes where $\boldsymbol{U}$ is conjugate variable of $\vec{\theta}$. MAPS estimator for the cross correlation of HI brightness temperature and Lyman-α flux contrast is defined as follows (Guha Sarkar et al., 2011)

$$\hat{E}(\boldsymbol{U}, \Delta z) = \frac{1}{2} \left[\tilde{\delta}_{\mathcal{F}o}(\boldsymbol{U}, z) \tilde{\delta}_T^*(\boldsymbol{U}, z + \Delta z) \right] + \frac{1}{2} \left[\tilde{\delta}_{\mathcal{F}o}^*(\boldsymbol{U}, z) \tilde{\delta}_T(\boldsymbol{U}, z + \Delta z) \right] \tag{2.18}$$

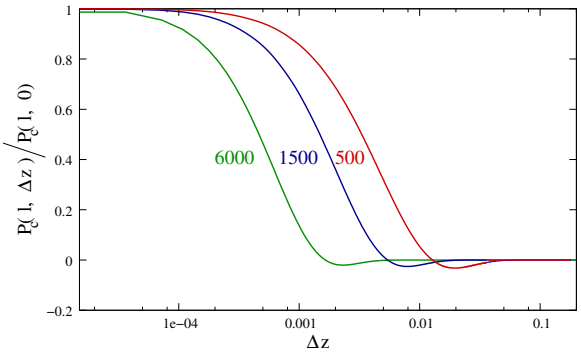

Figure 2.4: *This figure from Guha Sarkar et al. (2011) shows $P_c(l, \Delta z)/P_c(l, 0)$ with Δz for cross-correlation of 21-cm signal and Lyman-α forest. We can see that the normalized power spectrum falls of rapidly for higher l values.*

Here $\tilde{\delta}_{\mathcal{F}o}(\boldsymbol{U}, z)$ denotes the 2D Fourier transform of the the flux contrast of Lyman-α at redshift z and $\tilde{\delta}^*_T(\boldsymbol{U}, z + \Delta z)$ denotes the complex conjugate of 2D Fourier transform of 21-cm brightness temperature at redshift $z + \Delta z$. We calculate this estimator by averaging over all possible z. MAPS can also be defined using $\boldsymbol{k}_\perp$ which is related to $\boldsymbol{U}$ as $\boldsymbol{k}_\perp = 2\pi\boldsymbol{U}/r(z)$ where $r(z)$ denotes the line of sight comoving distance at redshift z. Assuming isotropy in the plane, we bin $\boldsymbol{k}_\perp$ in k by averaging over the circles in $\boldsymbol{k}_\perp$-plane (k means $k_\perp$ for this subsection). This reduces our Poission error as the number points in each bin increase.

$P(k, \Delta z)$ has the information of 3D box stored in it. Fixing Δz gives us the information about the power in Fourier modes in the plane perpendicular to the line of sight. Fixing k gives us information about the axis along the line of sight.
As Δz increases, the two slices at redshift z and $z + \Delta z$ start to be uncorrelated. It means $P(k, \Delta z)$ decreases as Δz increases. It will be nearly 0 for $\Delta z > z_{\text{correlation}}$ where $z_{\text{correlation}}$ is the typical redshift range over which the signal remains correlated. We also expect $P(k, \Delta z)$ to fall rapidly with Δz for higher values of k or $l = 2\pi U$. Lower value of k denotes the structure on large scale in the perpendicular plane. So structure over large scale will have higher $z_{\text{correlation}}$. Figure 2.4 shows a plot from Guha Sarkar et al. (2011) of cross-correlation using analytical calculations. We plot normalized power spectrum $P(k, \Delta z)/P(k, 0)$. It's value should be unity for $\Delta z = 0$ and fall down to 0 as Δz increases.

Similar behaviour is expected for MAPS of the dark matter density contrast. However, we find that the power spectrum is nearly flat as a function of Δz. Figure 2.5 shows

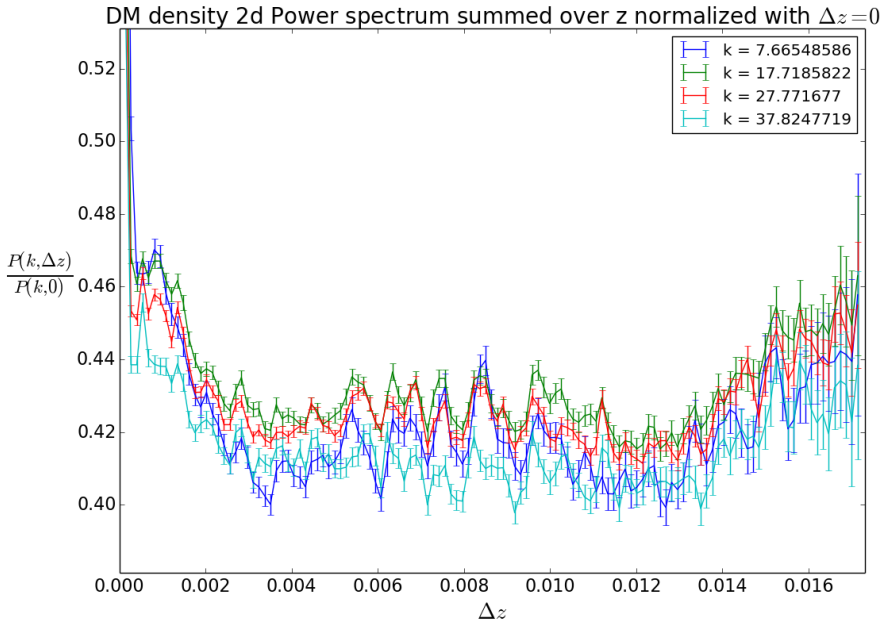

Figure 2.5: *This figure shows the normalized Multi-frequency angular power spectrum (MAPS) for dark matter density contrast. We see that that the power spectrum is nearly flat as a function of redshift and the shape is similar for different k values.*

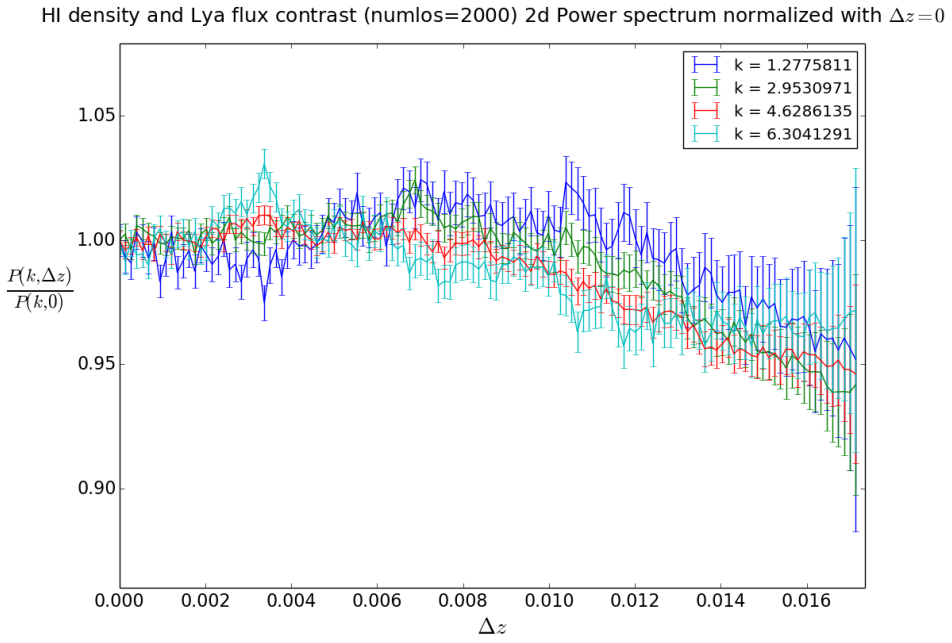

Figure 2.6: *This figure shows the normalized MAPS for cross-correlation of* HI *density and Lyman-α flux contrast for 200 lines of sight.*

one such plot. We don't see variation in the plots for different k values. This is probably because of the poor resolution of the box. We similarly calculate the cross-power spec-

trum of HI density and Lyman-α forest with 2000 lines of sight shot randomly along a particular axis. Figure 2.6 shows a plot of power spectrum. We see that the normalized multi-frequency angular power spectrum is nearly unity for all Δz values and does not fall at all.

To check the effect of number of lines of sight, we calculated the cross power spectrum of dark matter density contrast and dark matter density contrast sampled only along few lines of sight. We generated different sets of lines of sight randomly using different seeds. We found that the intensity of MAPS is much lower than the case where we sample the density contrast along all lines of sight. We can see that the shape is also different i.e. we don't see the peak that is present while sampling all lines of sight. We also see that there is a lot of variation in the power spectrum for different realizations of the lines of sight. See figure 2.7.

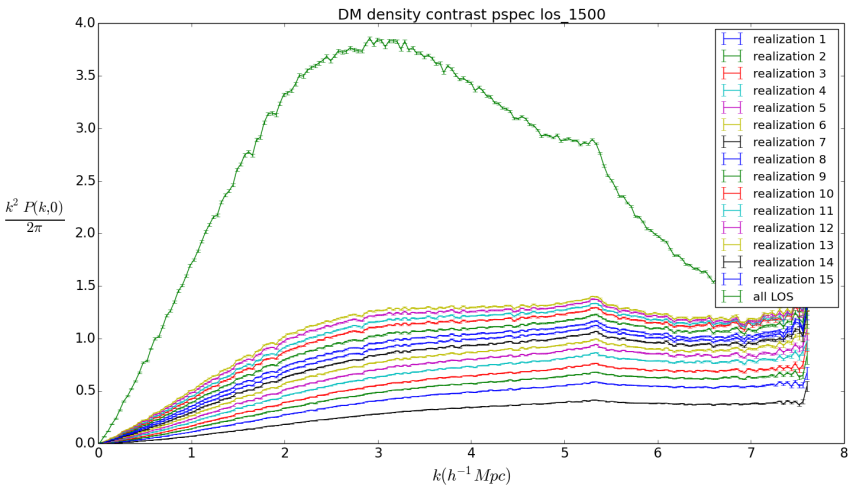

Figure 2.7: *This figure shows the 2D cross power spectrum of the two dark matter density contrast samples, one of which is sampled at all places and other one is sampled only along 1500 lines of sight. We have also plotted the power spectrum when dark matter density contrast is sampled over all points for comparison. Grid size is* 256^3*. The plot corresponds to* $\Delta z = 0$*.*

Here we are essentially missing $\boldsymbol{k}_\perp$ modes, as we sample Lyman-α forests only along few lines of sights. This was supported by the fact that 1D power spectrum calculated along the line of sight converged. Hence, we decided to use another method to calculate the power spectrum which is mentioned in the next section.

2.4.2 3D power spectrum contour plots

Here we plot power spectrum as a function of $k_\parallel$ and $k_\perp$, as these two are not on equal footing. We calculate 3D power spectrum, unlike MAPS where we had calculated the Fourier modes only along the plane perpendicular to the line of sight. Guha Sarkar and Datta (2015) present the similar calculations using analytical methods. We use the cross-power spectrum estimator defined in Guha Sarkar and Datta (2015) as

$$\hat{\mathcal{E}} = \frac{1}{2}\left[\Delta_{\mathcal{F}_0}(\boldsymbol{k})\Delta^*_{T_0}(\boldsymbol{k}) + \Delta^*_{\mathcal{F}_0}(\boldsymbol{k})\Delta_{T_0}\boldsymbol{k})\right] \tag{2.19}$$

We divide $\boldsymbol{k}$ modes into $k_\parallel$ and $k_\perp$. We calculate $k_\parallel$ by binning component of $\boldsymbol{k}$ parallel to the line of sight. We calculate $k_\perp$ by binning $\boldsymbol{k}$ modes in the plane perpendicular to the line of sight into circular bins, assuming 2D isotropy. We calculate the power spectrum $P(k)$ for each ($k_\parallel$, $k_\perp$) pair. We also calculate the 3D dimensionless power spectrum $k^3P(k)/2\pi^2$ where $k = \sqrt{k_\parallel^2 + k_\perp^2}$.

We use the high resolution GADGET-2 simulation with 1024^3 particles for this part. We first calculate the 21-cm brightness temperature δT_b using Eq. (2.2) on each grid point and then calculate its value for each ($k_\parallel$, $k_\perp$) pair after binning.

Figure 2.8 shows the $k_\parallel$, $k_\perp$ colour plot of $k^3P(k)/2\pi^2$ for 21-cm brightness temperature in units of mK^2, where axes are logscaled. Figure 2.9 shows similar plot using linear scale for axes. We see circular arcs having nearly same $k^3P(k)/2\pi^2$ values. This matches well with the isotropy, as we expect $P(k)$ to be a function of k where $k = \sqrt{k_\parallel^2 + k_\perp^2}$.

Figure 2.10 shows similar plot of $k^3P(k)/2\pi^2$ in units of mK, where $P(k)$ is cross power spectrum of 21-cm brightness temperature and Lyman-α forest sampled along all lines of sight. Here we don't have 3D isotropy as there is a preferred direction parallel to the line of sight.

We plan to continue this work further. The next step would be to calculate the cross power spectrum plots for less number of randomly shot lines of sight for Lyman-α forest. We also plan to calculate the noise and the Signal-to-Noise (SNR) ratio contours assuming observations from SKA-1.

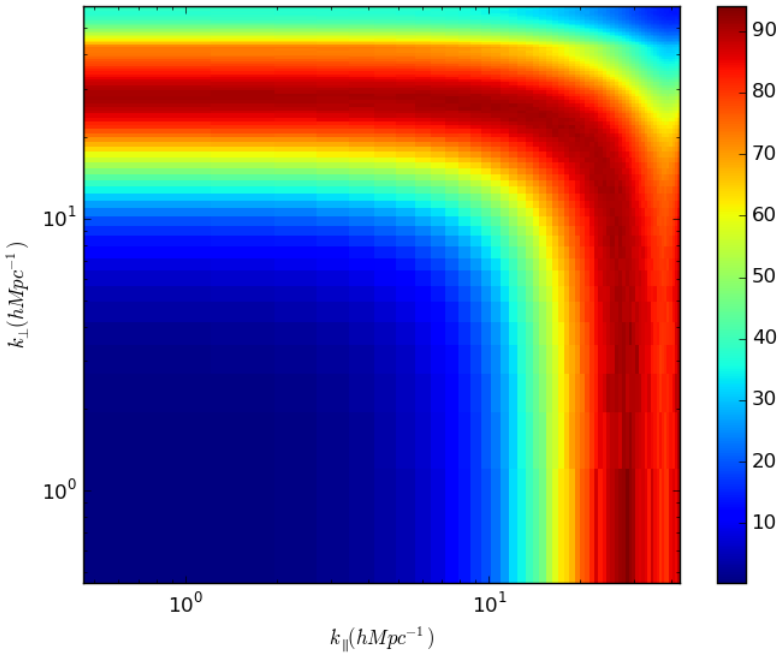

Figure 2.8: *This figure shows colour plot of* $k^3P(k)/2\pi^2$ *of* HI *brightness temperature in units of* mK^2*. The* x *and* y *axes denote* $k_\parallel$ *and* $k_\perp$ *respectively.*

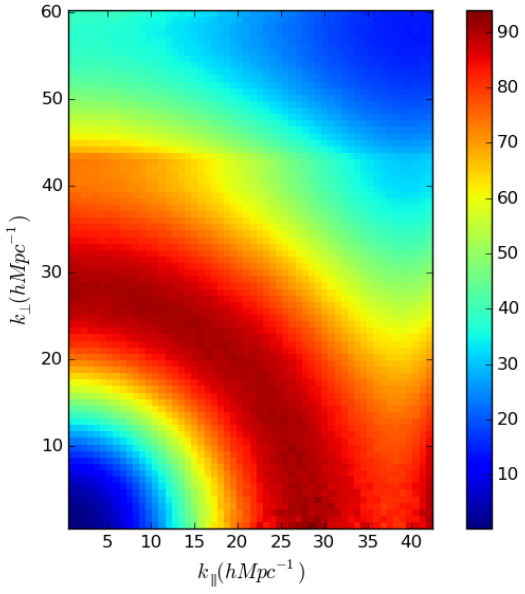

Figure 2.9: *This figure shows colour plot of* $k^3P(k)/2\pi^2$ *of* HI *brightness temperature in units of* mK^2*. The* x *and* y *axes denote* $k_\parallel$ *and* $k_\perp$ *respectively on linear scale. We can see circular arcs having same value, as expected by isotropy.*

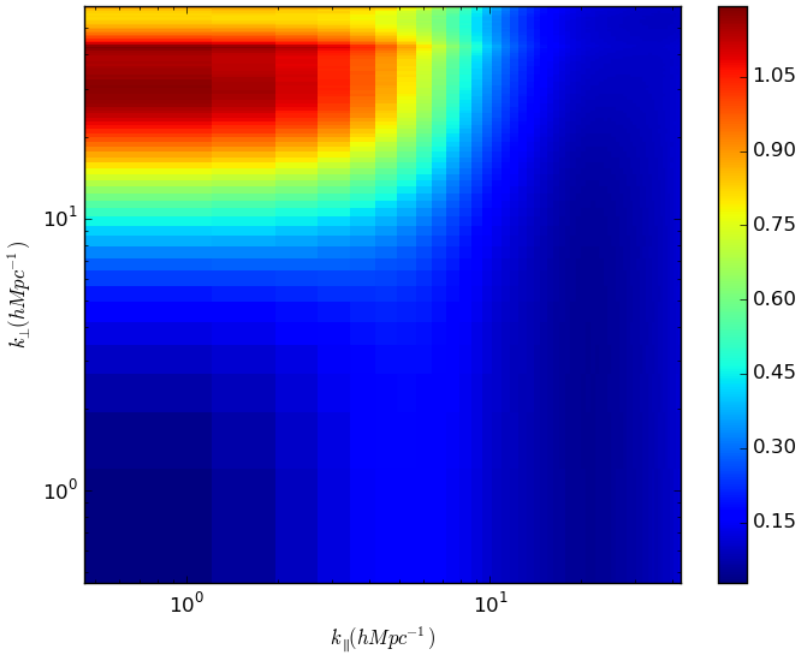

Figure 2.10: *This figure shows colour plot of* $k^3P(k)/2\pi^2$ *of* HI *brightness temperature and Lyman-*α *forests sampled over all lines of sight in units of* mK*. The* x *and* y *axes denote* $k_\parallel$ *and* $k_\perp$ *respectively.*

Chapter 3

HII bubble distribution during the reionization

3.1 History

Observations suggest that the intergalactic medium is highly ionized as of today. The reionization of the neutral hydrogen occurred during $20 \leqslant z \leqslant 6$. As the density perturbations in the early universe evolved, dark matter collapsed into halos. Baryons fall in the gravitational potential wells of the dark matter, forming galaxies at the centres of the dark matter halos. These first stars and galaxies started emitting high energy ultraviolet photons and ionizing the regions around them. These ionized regions grow with time and overlap with each other and eventually occupy the whole universe.

The study of reionization is imperative, as it is the important transition in the cosmic evolution that can be probed directly using observations. One important probe for observing the neutral hydrogen is Lyman-α forest (see section 2.3). However, quasar population peaks between redshift $z = 2$ and $z = 3$, and number of quasars observed at $z > 6$ is very less. Also, transmitted flux of the becomes zero if neutral hydrogen fraction is greater than 10^{-4}. We get nearly zero flux from quasars at $z > 6$ and hence, studying neutral hydrogen distribution gets highly model dependent at this epoch.

Observing the emission line of 21-cm from neutral hydrogen is another probe for its study. The redshifted 21-cm line will give us the neutral hydrogen distribution and ionized HII bubble distribution as well. The upcoming projects like the upgraded Giant Metrewave Radio Telescope (GMRT), Square Kilometre Array (SKA) are expected to

detect the redshifted 21-cm signal. Analytical calculations play an important role here and can be used to predict to what kind of signal can be expected from the observations. Semi-numerical calculations using the simulations can also serve the same purpose. These models can predict the size distribution of the ionized regions (ionized bubbles)

One important set of such models use 'excursion set formalism'. Initially this formalism was used to calculate halo mass function. (Press and Schechter, 1974; Bond et al., 1991) In this formalism, the probability of forming halo of a particular mass is related with the crossing of the barrier density of the random walks in the smoothing scale.

Furlanetto et al. (2004) showed that the excursion set formalism can be used for calculating the ionized bubble distribution as well. Based on their treatment, Paranjape and Choudhury (2014) developed an improved analytical model to calculate the bubble distribution, using the recent advances in the field. In this project, we calculated the size distribution of ionized bubbles using semi-numerical approach and checked its consistency with the analytical model by Paranjape and Choudhury (2014).

3.2 Analytical model

We summarise the basics of analytical models which use excursion set formalism from Furlanetto et al. (2004); Paranjape and Choudhury (2014). The following ansatz has been used to define ionized mass.

$$m_{ion} = \zeta m_{gal} \tag{3.1}$$

where m_{gal} denotes the mass of galaxy that is the source of ionizing photons and m_{ion} is the mass of ionized region around it. The parameter $\zeta > 1$ can depend on various physical parameters. One such dependence could be

$$\zeta = \frac{f_{esc} f_* N_{\gamma/b}}{1 + n_{rec}} \tag{3.2}$$

where f_{esc} is the escape fraction of the ionizing photons, f_* is the star formation efficiency, $N_{\gamma/b}$ is the number of ionizing photons produced per baryon and n_{rec} is the typical number over which hydrogen atom is recombined. Here we use ζ as a single parameter and do not focus on its dependences on other parameters. For an isolated region to be completely ionized, following condition should be satisfied

$$f_{coll} \geqslant f_x \equiv \zeta^{-1} \tag{3.3}$$

where f_{coll} denotes the collapsed mass fraction i.e. ratio of collapsed dark matter halo mass and the total mass in particular region. In the extended Press-Schechter model (Bond et al., 1991) sharp-k filter is used for smoothing, which makes calculations simple. In this model, the collapsed fraction of a region of mean overdensity δ_m is given by

$$f_{coll} = \text{erfc}\left[\frac{\delta_c(z) - \delta_m}{\sqrt{2[\sigma^2_{min} - \sigma^2(m)]}}\right] \tag{3.4}$$

where $\delta_c(z)$ is the critical density for collapse and m_{min} is the minimum mass of an ionizing source. $\sigma^2(m)$ is the variance of the density fluctuations on the scale of mass m and is given as.

$$\sigma^2(m) \equiv \left\langle \delta^2_{R_L} \right\rangle = \int dlnk\Delta^2(k)W^2(kR_L) \tag{3.5}$$

where $\Delta^2(k) \equiv k^3P(k)/2\pi^2$ is the dimensionless matter power spectrum and $W(kR_L)$ is the sharp-k smoothing filter in Fourier space. The Lagrangian radius is related with mass through

$$m = (4\pi/3)\bar{\rho}R^3_L \tag{3.6}$$

$\sigma^2(m)$ increases as scale we go on smaller scales. Using equation (3.4), we can rewrite condition 3.3 as a constrain on the density

$$\delta_m \geqslant \delta_x(m, z) \equiv \delta_c(z) - \sqrt{2}K(\zeta)[\sigma^2_{min} - \sigma^2(m)]^{1/2} \tag{3.7}$$

where $K(\zeta) = \text{erf}^{-1}(1 - \zeta^{-1})$. To use this formula, we need to fix a smoothing scale corresponding to m.

We get a naturally preferred smoothing scale in this problem. Consider an overdense region of space having $\zeta f_{coll} > 1$. It means that this region will have extra photons remaining after ionizing the whole region and these photons will leak outside the region and ionize neighbouring regions, that don't have enough f_{coll} to ionize themselves completely on their own. This region will ionize itself as well as its neighbouring underdense region. This problem can be solved if we use larger smoothing scale and consider the neighbouring overdense and underdense regions to be a single region, such that it is just able to ionize itself i.e. $\zeta f_{coll} = 1$. Hence, the smoothing scale should be the largest scale over which the region is able to ionize itself.

In case of halo mass function, the barrier to be crossed is a fixed number for a particular redshift. Here, as we have seen in equation (3.7), the barrier depends on the

smoothing scale as well. Furlanetto et al. (2004) used the linear best fit to the barrier to calculate the bubble distribution.

Ionized fraction is defined as

$$Q = \int_{V,min}^{\infty} dlnV \left(V \frac{dn}{dlnV} \right) \tag{3.8}$$

Furlanetto et al. (2004) used sharp-k filter for smoothing, which essentially corresponds to uncorrelated random steps. It also assumes no special location for halo formation, whereas we know that halos form preferably near peaks in the density field. Though sharp-k filters make calculations easy, filters like top-hat filter are more physical. Use of top-hat filter makes steps of random walk strongly correlated and first-crossing of random walks of the barrier can be replaced by a simpler up-crossing condition. This allows peak constraint to be included in a straightforward way. Paranjape and Choudhury (2014) calculated the HII bubble distribution using the correlated steps and peak constraint. The density barrier similar to equation (3.7) needs to be solved numerically, since there is no closed form conditional mass fraction expression available here.

3.3 Methodology

We use GADGET-2 simulations run by Dr. Aseem Paranjape. We used 5 different realizations of 1024^3 particles in the box of comoving size 50 h^{-1} Mpc. We used Planck-13 parameters ($\Omega_m = 0.315$, $\Omega_\Lambda = 0.685$, $h = 0.673$, $\sigma_8 = 0.829$). (Planck Collaboration et al., 2014) The snapshots were saved at redshifts $z = 18, 15, 12, 9, 6$.

We used Rockstar halo-finder to identify halos in the simulation. (Behroozi et al., 2013) Rockstar (Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement) is a halo finder which uses adaptive hierarchical refinement of FoF groups in 6 dimensional phase-space. Rockstar returns the halo catalogue with various parameters like number of dark matter particles, position, velocity, virial mass, m_{200b}, m_{200c}, m_{500c} etc. of the halo. m_{200b} is the mass of the halo such that the density of the region enclosing this mass is 200 times the mean background density. Similarly, m_{200c}, m_{500c} denotes the masses such that the corresponding density is 200 and 500 times the critical density respectively. We assume that the halos having more than 100 dark matter particles are the sources of ionizing photons. We also ignore sub-halos for this calculations.

We calculated the density contrast (δ) at all grid points using grids of sizes 250^3, 512^3 and 750^3 using cloud-in-cell (see 1.5.4). Then we calculated the density at each grid point using $\rho(i, j, k) = \rho_m(1 + \delta(i, j, k))$.

We similarly calculated the photo-ionization field at each grid point. As mentioned in equation (3.1), we assume that each dark matter halo of mass m_{halo} ionizes $\zeta \cdot m_{halo}$. Using cloud-in-cell on ionized mass and dividing it by the volume of the grid cell, we get the photo-ionization field.

We smooth the density and the photo-ionization field at various scales using spherical top-hat filter (1.33). We use the largest scale for smoothing to be the box length or the radius of sphere having volume corresponding to the sum of photo-ionization field, choosing the smaller one of the two. We use the smallest scale to be the grid separation. If the photo-ionization field is greater than the dark matter density at any scale for a grid point, we identify that grid point as ionized. We also consider partially ionized grid points without smoothing as ionized. This ensures that the grid points neighbouring halos are always ionized.

We then assign a specific mass to each grid point using the dark matter density and the volume of the grid cell. We then use Friends-of-Friends (see Sec. 1.5.5) to assign ionized grid points into ionized bubbles. The typical linking length parameter is $b = 0.2$ for FoF, which means that the linking length is 0.2 times the inter-particle distance. Here, we want linking length to be slightly higher than the grid separation. So, for each before each FoF run, we calculate the inter-particle distance for ionized points and calculate the required linking length parameter b and modify and recompile FoF accordingly. Volume of each bubble is calculated using the number of ionized particles in the bubble. We calculate the ionized fraction (Q) as the ratio of the ionized volume and the total volume of the box.

We bin the bubble volumes in the logarithmic bins. We then plot $Vdn/dlnV$ vs. V, where n is the comoving number density of the bubbles in that particular volume bin. V is the volume of the bin and $dlnV$ is the logarithmic volume bin interval. We plot the value averaged over 5 realizations with errors. We repeat the same exercise for different grid sizes, different ζ paramater values. We also plot values calculate from the analytical solution and compare both.

3.4 Results

We first calculated the density field and ionization field on the grids of 250^3. We calculated the $VdN/dlnV$ for 5 different realizations. Each value shown in the plots denotes the averaged value over realizations. The errorbars denote standard error on mean calculated as

$$\sigma = \left[\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N(N-1)} \right]^{1/2} \tag{3.9}$$

where x_i denotes the value for each realization, $\bar{x}$ is the mean value over realizations and N is the number of realizations, which is 5 in our case.

We have plotted two lines for analytical solutions. One of those lines is labelled as 'unscaled' which uses the halo mass function from purely analytical calculations. The other line is labelled as 'scaled' which uses scaled halo mass function which is linearly scaled so as to match halo mass function from N-body simulations with least square fitting.

Figure 3.1 shows the HII bubble volume distribution for $z = 9$ and $\zeta = 30$. We can see that the simulated values match very well with the rescaled analytical model. The ionized fractions are 0.061 and 0.052 respectively which match well. Figure 3.2 shows a similar plot for $z = 12$ and $\zeta = 30$. We can see that the values calculated from simulations are higher in each volume bin. The ionized fraction from simulations (7.5×10^{-3}) is nearly 3.5 times the value from rescaled analytical value (2.1×10^{-3}). We observe that the predicted ionized fraction from analytical calculation is lesser than the calculated value from simulations at high redshifts.

We plotted $VdN/dlnV$ for $\zeta = 60$ and $z = 9$ as well. Here we found that though the ionized fraction is similar for simulations and the rescaled analytical calculation, the distribution of the bubbles is not. We observed a very large bubble present in the same volume bin for all 5 realizations and no bubble present contiguous volume bins. Figure 3.3 shows the bubble distribution for $\zeta = 30$, 40, 50 and 60. The y-errorbar on the value in the largest volume bin is 0 for $\zeta = 60$, because there is exactly one bubble present in that volume bin for all 5 realizations. We see that for $\zeta = 50$ and 40, the bubble in the largest volume bin is present only in some of the realizations. A plausible reason for this could be that the ionized regions which are physically different are connected to each other because FoF linking length is equal to grid separation. We calculated the density,

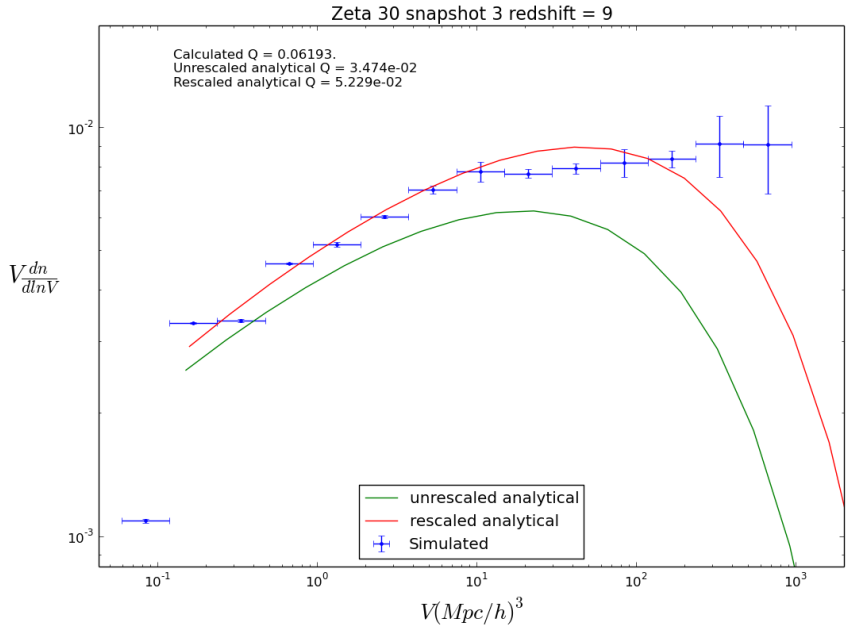

Figure 3.1: *Figure shows the volume distribution of* HII *bubbles at redshift* $z = 9$ *and* $\zeta = 30$ *for simulations and analytical calculations.*

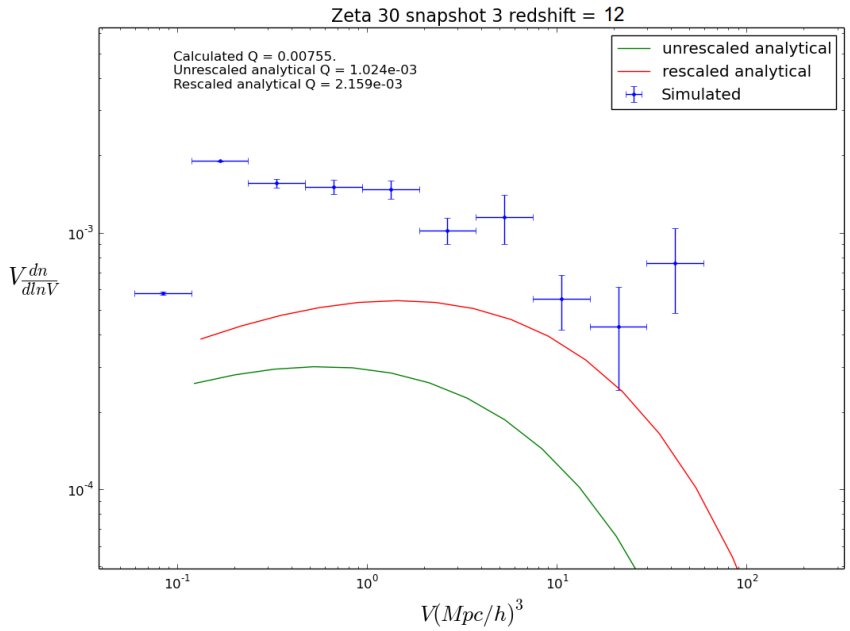

Figure 3.2: *Figure shows the volume distribution of* HII *bubbles at redshift* $z = 12$ *and* $\zeta = 30$ *for simulations and analytical calculations.*

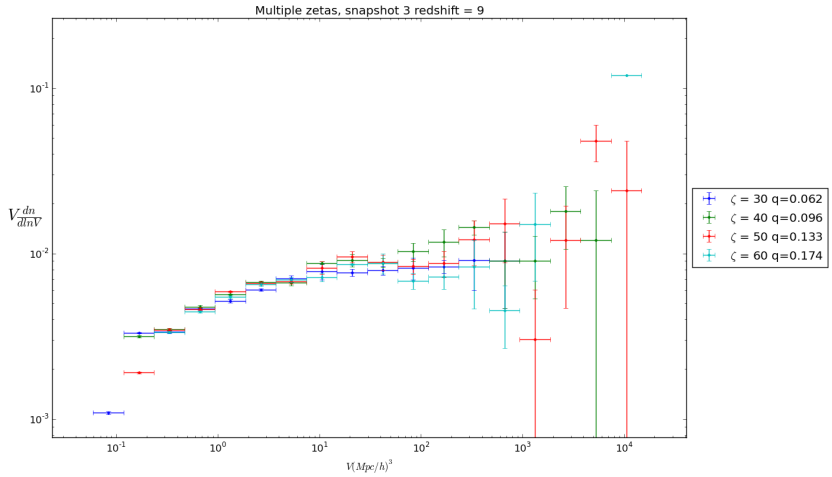

Figure 3.3: *This plot shows the volume distribution of* HII *bubbles at redshift z =9 and at ζ = 30, 40, 50 and 60. We can see that there are some bubbles present in the largest volume bin for some realizations for ζ = 40 and 50.*

ionization field and ionized grid points at better resolution i.e. grid size of 512^3 and 750^3. Figure 3.4 shows the bubble distribution calculated for 3 different grid sizes (250^3, 512^3

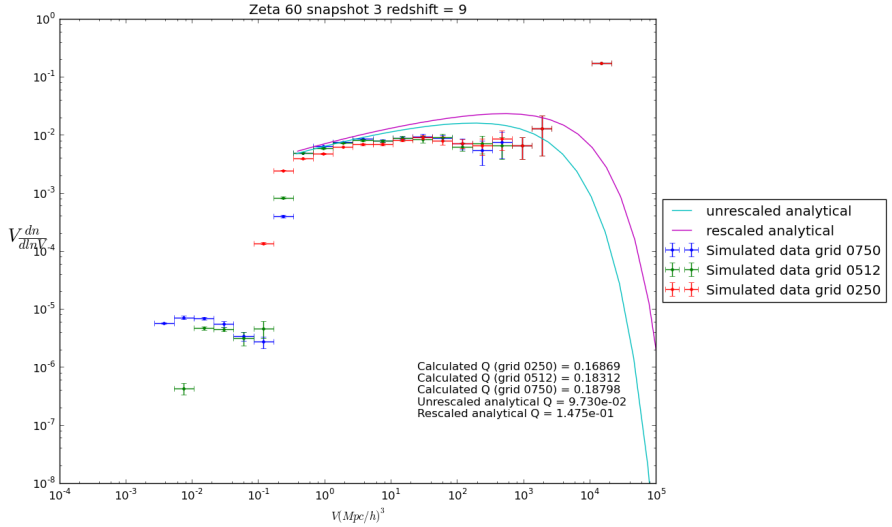

Figure 3.4: *This plot shows the volume distribution of* HII *bubbles at redshift z = 9, ζ = 60 using simulations with different grid sizes and analytical calculations. We can see that the largest bubble is present in the same volume bin for all 3 grid sizes (250, 512, 750)*

and 750^3) for ζ =60 and z =9. There exist a very large bubble present in each realization and for each grid size in the same volume bin. This makes sure that the largest bubble is not present only because of artefacts arising because of poor grid separation at 250 grid.

The calculations from simulations don't match with the analytical calculations for higher ζ and higher z. The excellent agreement for ζ =30 and z =9 seems to be a coincidence. The issue of largest bubble present needs further investigation.

3.5 Visualization

We visualized the ionized regions to check if they are actually surrounding the halos, if the voids are ionized etc. We took a slice through the box having width 1/10 the box size (5 h^{-1} Mpc in this case) and plotted all the ionized grid points in that slice in 2D plane. Since we use translucent gray to show ionized grid point, a point in the image appears darker if there are more ionized grid point in the slice at that location in 2D plane. FoF returns the centre of volume for each bubble, which is actually calculated as the centre of mass of the ionized grid points which are joined by FoF to identify them as one bubble. The volume of each bubble is calculated as the number of ionized grid points times the volume of each grid cell. We calculate the radius for each bubble assuming $V = 4\pi R^3/3$. We calculate the projected radius of the ionized bubble on the slice that we are considering. The plot on the left shows ionized grid points and circles centred at the centres of volume of the bubbles having radii as the projected radii on the slice. We do not expect each ionized region to lie inside the circles plotted because of the following reasons: The shape of bubbles is not necessarily spherical. There can be ionized region where the sphere around the centre of volume does not intersect with the slice and hence the centre of volume is not shown in the plot. It is possible to have two regions that look disconnected in 2D slice, but are part of the same bubble as they could be connected through a region which does not lie in the slice. We plot circles having projected radii larger than a cut-off value.

In the other subplot, we plot the centre of mass of the halos in that slice over the ionized grid points. We have used 3 different colours to denote 3 mass ranges. Halos having mass less than 30th percentile are shown in red. Those having mass between 30th and 60th percentile are shown in green and the remaining halos are shown in blue. This plot can be used to check that ionized regions are near dark matter halos.

Figure 3.5 shows such a plot of yz -plane centred at $x = 2.55$ h^{-1} Mpc. We can see that most of the ionized regions have corresponding bubble circle on the left plot. In the

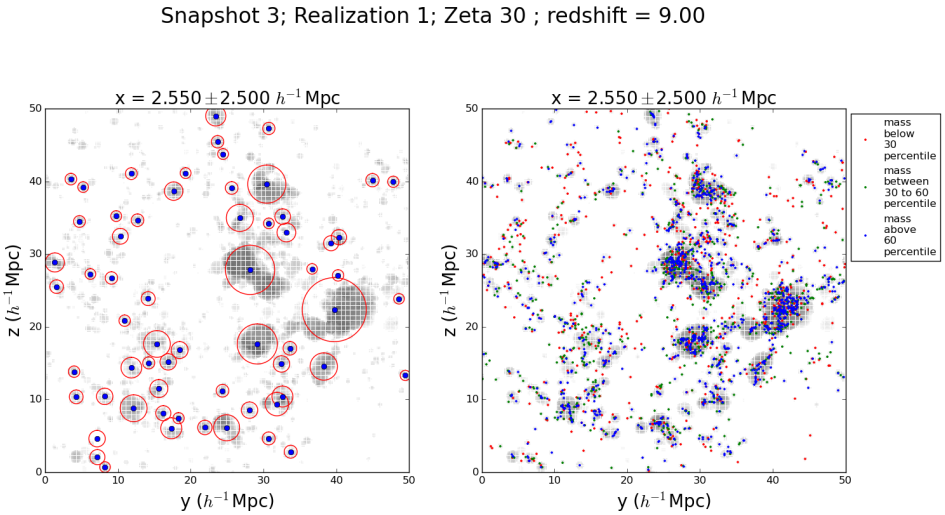

Figure 3.5: *This figure shows a slice centred at $x = 2.55h^{-1}$ Mpc. The left subplot has circles denoting the centres and calculated projected radii of the ionized bubbles. The right subplot shows the dark matter halos over the ionized grid points. We can see that the ionized regions are surrounding the dark matter halos.*

right subplot, we see that the large ionized regions are accompanied by many dark matter halos which is as expected. Figure 3.6 shows the similar plot with same slice for $\zeta = 120$. The central location of slice $x = 2.55$ h^{-1} Mpc is the x-coordinate of the centre of volume of the largest bubble for $\zeta = 120$. Note that FoF considers periodic boundary conditions while identifying the bubbles and calculating the centres of volume. Here we see a large connected ionized region and a large bubble circle having its centre in a corner of the image. This is because of the non-circularity of the largest bubble formed. The plot on the right shows ionized grid points even when no halos are present in that region. Voids are as well getting ionized here. We see that the small bubbles in figure 3.5 have joined to form a large bubble in figure 3.6. The vertical and horizontal lines on the ionized region are because of image aliasing.

FoF indexfile returns indices of particle belonging to different bubbles. Using this we plot only ionized grid points which are part of the largest bubble for $\zeta = 120$ and $z = 9$ in the slice same as figures 3.5 and 3.6 in figure 3.7. We can see that some of the ionized regions from figure 3.6 have disappeared.

We also made a 3D visualization video of the largest bubble present for $\zeta = 60$ and $z = 9$. The video can be accessed at `https://drive.google.com/file/d/0B11aWOD-uhwAR3hKalhXN0R2OFU/view?usp=sharing`

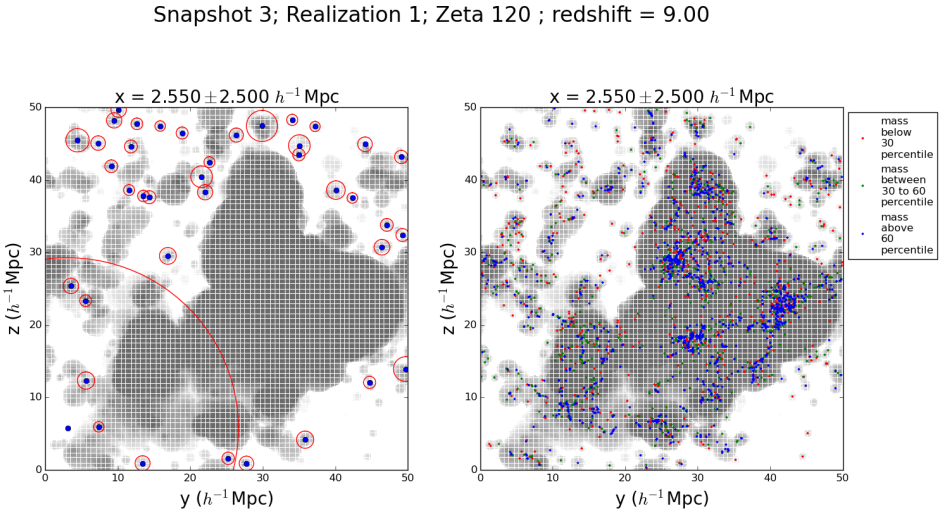

Figure 3.6: *This figure shows similar plot as Figure 3.5 at same slice. Here ζ is 120 instead of 30. Here we see a large bubble present. We can see that the regions not having dark matter halos are also ionized here.*

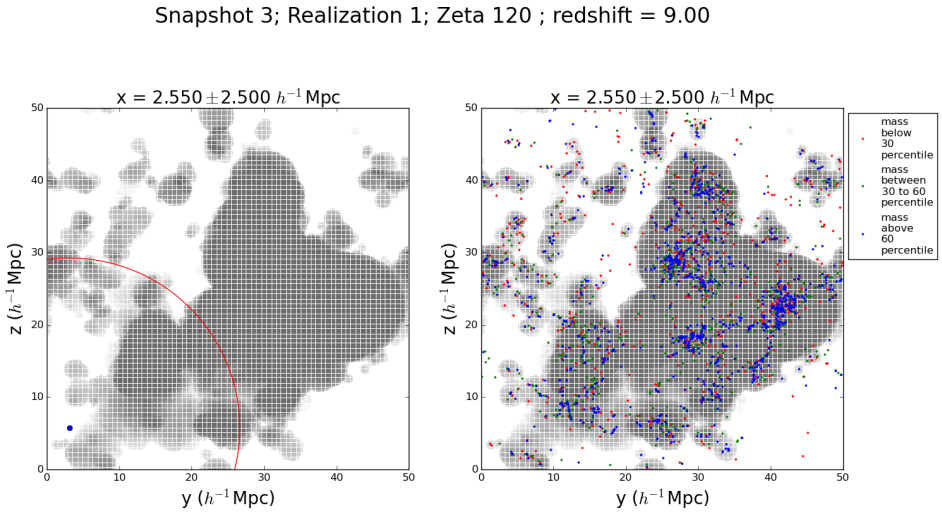

Figure 3.7: *This figure shows only the largest ionized bubble and ionized grid points corresponding to it from figure 3.6. The halo positions on the right plot, however are corresponding to all halos in that slice as we cannot associate halos only with a single bubble.*

Chapter 4

Summary and future work

Reionization of the neutral hydrogen is a vital event in the evolution of the universe. There are various probes to measure the dark matter power spectrum in the post-reionization era like redshifted 21-cm signal, Lyman-α forests etc. The cross-correlation of these two signal has been previously proposed to be a probe of the dark matter distribution on the large scale. The cross-correlation does not have problems of foreground contaminations as severe as just 21-cm signal.

We used Multi-frequency Angular Power Spectrum (MAPS) defined by Datta et al. (2007); Guha Sarkar et al. (2011) to calculate the cross power spectrum. We observed that MAPS did not show the expected behaviour for the variation of Δz. We also observed that the amplitude of MAPS depends greatly on the number of lines of sight shot and is very small comparing data sampled over all lines of sight. This is because of missing $k_\perp$ modes as we have Lyman-α forests only along few lines of sight. Hence, we stopped using MAPS and calculated 3D power spectrum for each ($k_\parallel$, $k_\perp$) pair. We first calculated it for 21-cm brightness temperature. We observed circles having equal values in $k_\parallel$ - $k_\perp$ plane, as expected because of isotropy. We then calculated the cross power spectrum for 21-cm brightness temperature and Lyman-α forest flux contrast sampled along all lines of sight.

We plan to continue this work further by calculating the cross power spectrum where Lyman-α forests are sampled only along randomly shot lines of sight. The next step would be calculation of noise following Guha Sarkar and Datta (2015) using SKA-1 as the observing telescope and plot contours of signal to noise ratio (SNR) so as to find k ranges over which detection can be possible in coming years.

The other project of this thesis was to calculate the distribution of HII bubbles formed during the reionization and compare it with the analytical model developed by Paranjape and Choudhury (2014) that uses 'excursion set formalism'. We used semi-numurical calculations for this purpose. We found that the values from simulations matched well with the analytical model for $\zeta = 30$ and $z = 9$. However, for higher redshifts the expected ionized fraction by analytical model was significantly lesser than the values calculated from simulations. Presence of a very large bubble was observed for $\zeta = 60$ and 120 for $z = 9$ for all realizations. The calculations at higher resolution of the grids confirmed that this was not just an effect of poor grid resolution causing FoF to join the physically distinct ionized regions. The presence of bubble needs further investigation.

We visualized the ionized regions formed in 2D slices. It was confirmed that the ionized regions are formed surrounding the halos. We also created a 3D visualization video of the largest bubble formed.

References

T. Guha Sarkar, S. Bharadwaj, T. R. Choudhury, and K. K. Datta, "Cross-correlation of the HI 21-cm signal and Lyα forest: a probe of cosmology", *Monthly Notices of Royal Astronomical Society* **410** (2011) 1130–1134, arXiv:1002.1368.

T. Guha Sarkar and K. K. Datta, "On using large scale correlation of the Ly-α forest and redshifted 21-cm signal to probe HI distribution during the post reionization era", arXiv:1501.02308.

A. Paranjape and T. R. Choudhury, "An improved model of HII bubbles during the epoch of reionization", *Monthly Notices of Royal Astronomical Society* **442** (2014) 1470–1482, arXiv:1401.7994.

Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al., "Planck 2013 results. XVI. Cosmological parameters", *Astronomy & Astrophysics* **571** (2014) A16, arXiv:1303.5076.

J. S. Bagla, "Gravitational clustering in an expanding universe", PhD thesis, University of Pune, 1996.

J. S. Bagla, "Cosmological N-body simulations", *Khagol* **48** (2001).

V. Springel, "The Cosmological simulation code GADGET-2", *Monthly Notices of Royal Astronomical Society* **364** (2005) 1105–1134, arXiv:astro-ph/0505010.

M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White, "The evolution of large-scale structure in a universe dominated by cold dark matter", *The Astrophysical Journal* **292** (1985) 371–394.

E. Komatsu, J. Dunkley, M. R. Nolta, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Limon, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, S. S. Meyer, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, "Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation", *The Astrophysical Journal* **180** (2009) 330–376, 0803.0547.

J. Bagla, N. Khandai, and K. K. Datta, "HI as a Probe of the Large Scale Structure in the Post-Reionization Universe", *Monthly Notices of Royal Astronomical Society* **407** (2010) 567, arXiv:0908.3796.

E. L. Wright, "Lyman Alpha Forest by Ned Wright". http://www.astro.ucla.edu/~wright/Lyman-alpha-forest.html.

T. R. Choudhury, R. Srianand, and T. Padmanabhan, "Semianalytic Approach to Understanding the Distribution of Neutral Hydrogen in the Universe: Comparison of Simulations with Observations", *The Astrophysical Journal* **559** (2001) 29–40, astro-ph/0012498.

T. R. Choudhury, "Physics of structure formation in the universe", PhD thesis, University of Pune, 2003.

K. K. Datta, T. R. Choudhury, and S. Bharadwaj, "The multifrequency angular power spectrum of the epoch of reionization 21-cm signal", *Monthly Notices of Royal Astronomical Society* **378** (2007) 119–128, astro-ph/0605546.

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

J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser, "Excursion set mass functions for hierarchical Gaussian fluctuations", *The Astrophysical Journal* **379** (1991) 440–460.

S. Furlanetto, M. Zaldarriaga, and L. Hernquist, "The Growth of HII regions during reionization", *The Astrophysical Journal* **613** (2004) 1–15, arXiv:astro-ph/0403697.

P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu, "The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores", *The Astrophysical Journal* **762** (2013) 109, arXiv:1110.4372.