Chapter 3

Cross-correlations of the post-reionization redshifted H I 21 cm signal*

In the last chapter we saw that tomographic intensity mapping of the H I using the redshifted 21 cm observations opens up a new window towards our understanding of cosmological background evolution and structure formation [75, 79, 104, 107]. The astrophysical processes dominating the epoch of reionization is now believed to have completed by redshift $z \sim 6$ [123]. In the post-reionization era most of the neutral HI gas are housed in the Damped Lyman-α (DLA) systems. These DLA clouds are the predominant source of the HI 21-cm signal. Intensity mapping involves a low resolution imaging of the diffuse HI 21-cm radiation background without attempting to resolve the individual DLAs. Such a tomographic imaging shall naturally provide astrophysical and cosmological data regarding the large scale matter distribution, structure formation and background cosmic history in the post-reionization epoch [102, 107, 110, 112, 124]. This is a key science goal of Several functioning and upcoming radio interferometric arrays like Giant Metrewave Radio Telescope (GMRT) †, the Ooty Wide Field Array (OWFA) ‡, the Canadian Hydrogen Intensity Mapping Experiment (CHIME) §, the Meer-Karoo Array Telescope (MeerKAT) ¶, the Square Kilometer Array (SKA) ‖ are aimed

*This chapter is adapted from the paper titled "Redshifted HI 21-cm Signal from the Post-Reionization Epoch: Cross-Correlations with Other Cosmological Probes" [122]

†http://gmrt.ncra.tifr.res.in/

‡http://www.ncra.tifr.res.in/ncra/ort

§http://chime.phas.ubc.ca/

¶http://www.ska.ac.za/meerkat/

‖https://www.skatelescope.org/

towards detecting the cosmological 21-cm background radiation. Detecting the 21 cm signal, is however extremely challenging. This is primarily because of the large astrophysical foregrounds [13, 14, 121] from galactic and extra-galactic sources which are several order of magnitude greater than the signal .

In this chapter we explore the possibility of cross correlating the 21 cm signal with other tracers of the large scale structure. The cross-correlation of the post-reionization H I signal has been studied in the context of correlation with Lyman-break galaxies [21], and late time anisotropies in the CMBR maps like weak lensing [125] and the Integrated Sachs Wolfe effect [126].

We shall first develop the general formalism for cross correlation and then elaborate upon the special case of cross-correlation of the post-reionization 21 cm signal with the Lyman-α forest. We investigate the feasibility of detecting the signal and explore the possibility of obtaining constraints on cosmological models using it. In the next chapters we shall use the cross-correlation of the post-reionization 21 cm signal with the Lyman-α forest as the cosmological probe.

3.1 Why cross-correlation ?

In the last chapter we saw that that the 21-cm brightness temperature maps can be treated as a biased tracer of the underlying dark matter distribution. This has motivated the use of the cross correlation of the H I -maps with other cosmological fields. The key advantage is that the detection of the cross-correlation signal would unambiguously ascertain the cosmological nature of the H I distribution. The cross-correlation has been often proposed as a probe of the epoch of reionization(EoR). The cross-correlation of 21-cm emission EoR signal with galaxy distribution would prove/disprove the theoretical predictions that over dense regions (sampled by galaxies) were the first to be ionized ([101, 127–129]). This shall allow us to make a distinction between an "inside-out" or "outside-in" reionization model using even the first generation 21-cm experiments and galaxy surveys of just a few square degrees. Cross-correlation of Lyman-α emitters with 21-cm emission has been studied in large simulations to model the role of patchy reionization on the clustering properties of α emitters ([130, 131]). Several studies ([57, 132, 133]) have also investigated the sensitivities at which the EoR signal may be detected through a cross correlation with the CMBR (Sunyaev Zel'dovich effect), the IR background or galaxies. These studies show that it is possible to detect the cross-correlation signal at a higher signal-to-noise ratio than the auto-correlation 21-cm power spectrum. Further, cross-correlation of the EoR H I would imprint the effects of inhomogeneous reionization, and may be discerned at higher sensitivity.

Our work is focussed on the post-reionization epoch, when the H I brightness temperature is believed to trace the large scale dark matter field with a linear bias. The bias is now known to be scale independent on large scales (see Chapter-2). Observational evidence of correlation between the H I 21-cm signal and the optical galaxies (6dFGS) ([134]) also strengthens this belief. This strengthens our motivation to use the post-reionization 21-cm maps as a biased tracer and thereby study its cross-correlations. .

Detecting the 21 cm signal, is extremely challenging. This is primarily because of the large astrophysical foregrounds [13, 14, 121] from galactic and extra-galactic sources which are several order of magnitude greater than the signal .

Cross-correlating the 21 cm signal with other probes may prove to be useful towards mitigating the severe effect of foreground contaminants and other systematic effects which plague the signal. Foregrounds which contaminate the H I signal poses considerably less problem when dealing with cross-correlations. We should however note that a certain degree of foreground cleaning is still required for a detection of the cross-correlation signal.

The main advantage of cross-correlation is that the cosmological origin of the signal can only be ascertained only if it is detected with high a statistical significance in the cross- correlation. Further, a cross-correlation signal may be detected at a greater SNR as compared to the H I auto power spectrum.

Cosmological parameter estimation often involves a joint analysis of two or more data sets and this would require not only the auto-correlation but also cross-correlation information. Further, the two different probes may focus on specific $k-$ modes with high signal to noise ratio and in such cases the cross-correlation signal takes advantage of the different cosmological probes simultaneously. This has been studied extensively in the case of the BAO [135] signal. It is to be noted that if the observations of the distinct probes are perfect, there shall be no new advantage of using the cross correlation. However, we expect the first generation observations of the redshifted HI 21 cm signal to have large systematic errors and foreground residuals (even after subtraction). For a detection of the 21 cm signal and subsequent cosmological investigations these measurements can be cross-correlated with other large scale structure tracers to yield information from the 21 cm signal which may not be possible to obtain using the low SNR auto correlation signal. In this chapter we shall study the cross-correlation of the

21 cm signal with the Lyman-α flux distribution in details. On large scales both the Lyman-α forest absorbed flux and the redshifted 21-cm signal are, believed to be biased tracers of the underlying dark matter (DM) distribution [22–25]. The clustering of these signals, is then, directly related to the underlying dark matter power spectrum. In the subsequent chapters we investigate the possibility of using the cross-correlation of the 21-cm signal and the Lyman-α forest for cosmological parameter estimation, neutrino mass measurement, constraining warm dark matter masses and dark energy modelling using the BAO features in the cross-power spectrum.

3.1.1 Cross-correlation between cosmological signals

Consider two cosmological fields $A(\mathbf{r})$ and $B(\mathbf{r})$. We shall be generally interested in biased tracers of the underlying dark matter distribution whereby $A(\mathbf{r})$ and $B(\mathbf{r})$ are expected to be correlated. The 2-point correlation function may be defined as

$$\xi_{AB}(s) = \langle A(\mathbf{r}) B(\mathbf{r} + \mathbf{s}) \rangle \tag{3.1}$$

where the ensemble average may be replaced by a volume average and we have assumed statistical isotropy. It is often convenient to work in the Fourier space for observational reasons. In case of the 21-cm signal, for example, the quantity measured directly in radio interferometric observation is the visibility which is the Fourier transform of the actual 21-cm brightness temperature field. Thus we introduce the Fourier transforms $\tilde{A}(\mathbf{k})$ and $\tilde{B}(\mathbf{k})$ of the two tracers of large scale structure defined as

$$A(\mathbf{r}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3}\, e^{i\mathbf{k}.\mathbf{r}} \tilde{A}(\mathbf{k})\,. \qquad B(\mathbf{r}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3}\, e^{i\mathbf{k}.\mathbf{r}} \tilde{B}(\mathbf{k})\,. \tag{3.2}$$

The statistical quantity of interest is the cross-correlation power spectrum P_{AB} defined as

$$\langle \tilde{A}(\mathbf{k}) \tilde{B}^*(\mathbf{k}') \rangle = P_{AB}(k) \delta_D(\mathbf{k} - \mathbf{k}') \tag{3.3}$$

Observationally one would measure $\tilde{A}_O(\mathbf{k})$ and $\tilde{B}_O(\mathbf{k})$, which would contain the signals $\tilde{A}(\mathbf{k})$ and $\tilde{B}(\mathbf{k})$ along with some additive noise and foreground contaminants.

We define the cross correlation estimator $\hat{E}$ as follows

$$\hat{E} = \frac{1}{2} \left[\, \tilde{A}_O \tilde{B}^*_O \;+\; \tilde{B}_O \tilde{A}^*_O \,\right] \tag{3.4}$$

We note that $\tilde{A}_O$ and $\tilde{B}_O$ can be complex fields. It is important to note that the noise, foregrounds and other systematics shall be uncorrelated between the two fields. Thus we see that the expectation value of the estimator faithfully yields the cosmological cross-correlation power spectrum.

$$\langle \hat{E} \rangle = P_{AB}(k) \tag{3.5}$$

The noise and systematics do not appear in the cross-correlation signal. They however appear, as we shall see in the variance. For a diagonal covariance matrix the noise is uncorrelated between neighbouring modes and we are interested in the variance of the estimator defined as

$$\sigma^2_{\hat{E}} = \langle \ \hat{E}^2 \ \rangle - \ \langle \hat{E} \ \rangle^2 \tag{3.6}$$

Noting that $\langle \ \tilde{A}_O(\mathbf{k}) \ \tilde{A}_O(\mathbf{k}) \ \rangle = \langle \ \tilde{A}_O(\mathbf{k}) \ \tilde{A}^*_O(-\mathbf{k}) \ \rangle = 0$, we have

$$\langle \ \hat{E}^2 \ \rangle = \frac{1}{2} \ \left[\ \langle \tilde{A}_O \tilde{A}^*_O \rangle \langle \tilde{B}_O \tilde{B}^*_O \rangle \ + \ |\langle \tilde{A}_O \tilde{B}_O \rangle|^2 \ + \ 3 \ |\langle \tilde{A}_O \tilde{B}^*_O \rangle|^2 \ \right] \tag{3.7}$$

Further, the term $\langle \tilde{A}_o \tilde{B}_O \rangle$ can be dropped since

$$\langle \ \tilde{A}_O(\mathbf{k}) \ \tilde{B}_O(\mathbf{k}) \ \rangle = \langle \ \tilde{A}_O(\mathbf{k}) \ \tilde{B}^*_O(-\mathbf{k}) \ \rangle = C\delta_{\mathbf{k},-\mathbf{k}} = 0 \tag{3.8}$$

This gives

$$\sigma^2_{\hat{E}} = \langle \ \hat{E}^2 \ \rangle - \ \langle \hat{E} \ \rangle^2 = \ \frac{1}{2}[\ \langle \tilde{A}_O \tilde{A}^*_O \rangle \langle \tilde{B}_O \tilde{B}^*_O \rangle \ + \ |\langle \tilde{A}_O \tilde{B}^*_O \rangle|^2 \ ] \tag{3.9}$$

The variance is suppressed by a factor of N_c for that many number of independent estimates. Thus, finally we have

$$\sigma^2_{\hat{E}} = = \ \frac{1}{2N_c}[\ \langle \tilde{A}_O \tilde{A}^*_O \rangle \langle \tilde{B}_O \tilde{B}^*_O \rangle \ + \ |\langle \tilde{A}_O \tilde{B}_O \rangle|^2 \ ] \tag{3.10}$$

For detection, one is often interested in the signal to noise ratio defined as

$$SNR = \sqrt{N_c} \frac{P_{AB}(k)}{\sigma_{\hat{E}}} \tag{3.11}$$

We note that the quantities $\langle \tilde{A}_O \tilde{A}^*_O \rangle$ and $\langle \tilde{B}_O \tilde{B}^*_O \rangle$ denote the observed power spectrum of A and B respectively. We may hence write

$$\langle \tilde{A}_O \tilde{A}_O^* \rangle = P_A(k) + N_A(k) + F_A(k) \quad (3.12)$$

$$\langle \tilde{B}_O \tilde{B}_O^* \rangle = P_B(k) + N_B(k) + F_B(k) \quad (3.13)$$

where P_A and P_B denote the auto-correlation power spectra for A and B respectively and N and F stands for the noise and foreground power spectra. It is important to recognize that P_A and P_B may actually be the convolution of the pristine cosmological power spectra with the respective observational aperture (or selection) functions respectively.

3.2 Cross-correlation of Post-reionization 21 cm signal with Lyman-α forest

The observations of the post reionization redshifted 21-cm radiation (in emission) may be used as a probe of cosmological matter distribution over a wide redshift range ($z \leq 6$) through its cross-correlations with other probes.

Neutral gas in the same redshift range $z \leq 6$ (post reionization epoch) also produces distinct absorption features, in the spectra of background quasars [136]. The Lyman-α forest, traces the HI density fluctuations along one dimensional quasar lines of sight. The Lyman-α forest observations finds several cosmological applications [137–142]. On large cosmological scales the Lyman-α forest and the redshifted 21-cm signal are, both expected to be biased tracers of the underlying dark matter (DM) distribution [22–25]. This allows to study their cross clustering properties in n-point functions. Also the Baryon Oscillation Spectroscopic Survey (BOSS) * is aimed towards probing the dark energy through measurements of the BAO signature in Lyman-α forest [143]. The availability of Lyman-α forest spectra with high signal to noise ratio for a large number of quasars from the BOSS survey allows 3D statistics to be done with Lyman-α forest data [144, 145].

A neutral hydrogen cloud in the line of sight (LOS) to a distant QSO will cause absorption of the QSO continuum for a frequency redshifted to the Lyman-α (1215.67A°) UV resonance line in the rest frame of the gas. The inhomogeneous distribution of H I in an expanding Universe will thereby lead to the formation of

*https://www.sdss3.org/surveys/boss.php

a series of absorption lines blue ward of 1215.67A°, known as the Lyman-α forest. The optical depth of the Lyman-α forest is believed to be a tracer of the underlying dark matter distribution, and is of observational interest in cosmology. It is to be noted that a tiny fraction of the absorption lines in the Lyman-α forest are formed due to absorption by heavier elements. We shall be largely ignoring these metal line contaminations.

The Lyman-α forest is an well established cosmological probe [138, 146] with diverse cosmological applications which include determining the matter power spectrum [137, 139, 147], parameter estimation [141, 148, 149], constraining small scales clustering properties of dark matter [150], constraining neutrino mass [140, 149] and extensive use in probing the reionization history of the Universe [142, 151, 152].

Most analytical studies of the Lyman-α forest assumes that these features arise from small scale density fluctuations in the H I . The transmitted-flux fluctuations through the forest is believed to be a tracer of the underlying dark matter distribution with a possible bias. Most of the study of the IGM which incorporates most of the ongoing physical processes are through hydrodynamical simulations [153–156] or through semi analytical modelling ([157–163]). These techniques are however not complete and incomplete/inadequate modelling of the IGM often reveals as systematic errors. There are also a large number of observational issues which come in the way of constraining theoretical models. Some of the key sources of observational uncertainty includes: inadequate understanding of the of the background ionizing field, fitting the QSO continuum fluctuations, uncertainties in the slope of the temperature-density relationship in the diffuse IGM and its inevitable fluctuations about the mean relation, the contamination of the Lyman-α forest from metal lines in the quasar spectra and the poorly quantified effects of galactic super winds on the Lyman-α forest.

Semi-analytical modelling of the Lyman-α forest reveals that these absorption lines can be explained from linear or quasi linear density fluctuations. Highly non-linear processes may, hence be ignored in a first approximation. We shall assume that on large cosmological scales, the fluctuations in the Lyman-α flux field shall be a faithful tracer of the dark matter density fluctuation field. However, the forest can be seen as a continuous field being sampled at discrete line of sights (corresponding to the positions of individual background quasars). The cosmological evolution of the diffuse IGM is known to be primarily governed by the cosmological Hubble expansion and gravitational instabilities. At the linear level the underlying Physics being quite simple and remains so even when weak nonlinearities are included.

There is a considerable transmitted flux in the Lyman-α forest indicating that the IGM is highly ionized (the neutral hydrogen fraction being $\sim 10^{-5}$). Along a given line of sight to a quasar there exists a very large large number of absorption lines which makes a statistically analysis sensible even for a single spectra. The quasar surveys like SDSS III Baryon Oscillation Spectroscopic Survey (BOSS) [164] aims to measure the absorption spectra of $160,000$ QSOs and the data may improve significantly in future surveys. This makes it possible to use Lyman-α forest for precision cosmology.

Since we shall be looking at the cross-correlation of the 21-cm signal with Lyman-α forest, it is important to point out here that though both the 21-cm emission and the Lyman-α forest from H I in the post-reionization epoch at the same z, the two tracers actually belong to two different kinds of astrophysical systems. The Lyman-α forest has its origin in the small H I fluctuations in a primarily ionized IGM; and the 21-cm emission from these diffuse H I regions is almost negligible. The bulk of the 21-cm signal, on the contrary originate from the Damped Lyman-α Absorber (DLAs) systems which actually house most of the H I at these redshifts ([82–84]). Numerical simulations however reveal that it is reasonable to assume that on large scales both these trace the same underlying large scale structure, whereby we expect them to be correlated. Statistical study of the cross-correlation of the Lyman-α forest and the 21-cm signal has the potential to probe the dynamics, thermal state of the IGM, and also the clustering properties of the dark matter.

In this work, we construct an unbiased estimator of the cross-correlation signal as well as its variance, which can be used for a Fisher matrix analysis for constraining cosmology.

Detection the individual auto-correlation signals are observationally challenging. For the HI 21-cm a detection of the signal requires careful modelling of the foregrounds [17, 18]. Some of the difficulties faced by Lyman-α observations include proper modelling of the continuum, fluctuations of the ionizing sources, poor modelling of the temperature-density relation [165] and metal lines contamination in the spectra [166]. The two signals are tracers of the underlying dark matter distribution. Thus they are correlated on large scales. However foregrounds and other systematics are uncorrelated between the two independent observations. Hence, the cosmological nature of a detected signal can be only ascertained in a cross-correlation. The 2D and 3D cross correlation of the redshifted HI 21-cm signal with other tracers such as the Lyman-α forest, and the Lyman break galaxies have been proposed as a way to avoid some of the observational issues [19, 21]. The foregrounds in HI 21-cm observations appear as noise in the cross correlation and hence, a significant degree foreground cleaning is still required for a detection. The cross-correlation signal shall independently probe the the same astrophysical and cosmological information as the individual auto-correlations. However the cross-correlation has the added advantage that the problems of foregrounds and systematics are expected to be much less severe for its detection. There is also some literature that considered the possibility of cross-correlating the Lyman-α forest with the CMBR [167], large scale structure [168] and weak lensing [169, 170].

The fluctuations in the transmitted flux $\mathcal{F}(\hat{\mathbf{n}}, z)$ along a specified line of sight $\hat{\mathbf{n}}$ through the Lyman-α forest may be quantified using

$$\delta_{\mathcal{F}}(\hat{\mathbf{n}}, z) = \mathcal{F}(\hat{\mathbf{n}}, z)/\bar{\mathcal{F}} - 1. \tag{3.14}$$

where $\bar{\mathcal{F}}$ denotes the mean transmitted flux through the forest.

On large scales one may use the the fluctuating Gunn-Peterson approximation [49, 137, 147, 171] which relates the transmitted flux and the matter density contrast through the relation

$$\mathcal{F} = \exp[-A(1+\delta)^{\mathcal{K}}] \tag{3.15}$$

where A and $\mathcal{K}$ are both functions of the redshift. The function A is related to the the mean flux level, IGM temperature, photo-ionization rate and other cosmological parameters [137]. We now know that A is of order unity [166]. The quantity $\mathcal{K}$ depends on the power law exponent in the IGM temperature density relation [162, 165]. For an analytic estimate of the cross-correlation signal, we assume that $\delta_{\mathcal{F}}$ to be smoothed whereby it is we may use the linear relationship $\delta_{\mathcal{F}} \propto \delta$ [49, 137, 147, 171–174] Numerical simulations of the Lyman-α forest however justifies the linear approximation and the higher order terms has subdomimant contribution to the cross-correlation.

We have already discussed the post-reionization 21-cm signal in the last chapter. In the redshift range ($z < 3.5$) the fluctuation in the redshifted 21-cm brightness temperature $\delta_T(\hat{\mathbf{n}}, z)$ is a tracer of the the underlying dark matter distribution with a possible scale dependent bias function $b_T(k, z)$.

We use δ_T to denote the redshifted 21-cm brightness temperature fluctuations and $\delta_{\mathcal{F}}$ as the fluctuation in the transmitted flux through the Lyman-α forest. We write $\delta_{\mathcal{F}}$ and δ_T in Fourier space as

$$\delta_a(\mathbf{r}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3}\, e^{i\mathbf{k}.\mathbf{r}}\Delta_a(\mathbf{k})\,. \tag{3.16}$$

where $a = \mathcal{F}$ and T refer to the Lyman-α forest transmitted flux and 21-cm brightness temperature respectively. On large scales we may write

$$\Delta_a(\mathbf{k}) = C_a[1 + \beta_a\mu^2]\Delta(\mathbf{k}) \tag{3.17}$$

where $\Delta(\mathbf{k})$ is the dark matter density contrast in Fourier space and μ denotes the cosine of the angle between the line of sight direction $\hat{\mathbf{n}}$ and the wave vector ($\mu = \hat{\mathbf{k}}\cdot\hat{\mathbf{n}}$). β_a is similar to the linear redshift distortion parameter. The corresponding power spectra are

$$P_a(k,\mu) = C_a^2[1 + \beta_a\mu^2]^2 P(k) \tag{3.18}$$

where $P(k)$ is the dark matter power spectrum.

For the 21-cm brightness temperature fluctuations we have

$$C_T = 4.0\,\mathrm{mK}\, b_T\, \bar{x}_{\rm HI}(1+z)^2 \left(\frac{\Omega_{b0}h^2}{0.02}\right)\left(\frac{0.7}{h}\right)\left(\frac{H_0}{H(z)}\right) \tag{3.19}$$

The neutral hydrogen fraction $\bar{x}_{\rm HI}$ is assumed to be a constant with a value $\bar{x}_{\rm HI} = 2.45\times10^{-2}$ [82, 84, 175]. For the HI 21-cm signal the parameter β_T, is the ratio of the growth rate of linear perturbations $f(z)$ and the HI bias b_T (see Eq. 2.11). The 21 cm bias is assumed to be a constant. This assumption of linear bias is supported by several independent numerical simulations [23, 24] which shows that over a wide range of k modes, a constant bias model is adequately describes the 21 cm signal for $z < 3$. We have adopted a constant bias $b_T = 2$ from simulations [23–25]. For the Lyman-α forest, $\beta_{\mathcal{F}}$, can not be interpreted in the usual manner as β_T. This is because Lyman-α transmitted flux and the underlying dark matter distribution [145] do not have a simple linear relationship. The parameters $(C_{\mathcal{F}}, \beta_{\mathcal{F}})$ are independent of each other.

We adopt approximately $(C_{\mathcal{F}}, \beta_{\mathcal{F}}) \approx (-0.15, 1.11)$ from the numerical simulations of Lyman-α forest [22]. We note that for cross-correlation studies the Lyman-α forest has to be smoothed to the observed frequency resolution of the HI 21 cm frequency channels.

We now consider the 3D cross-correlation power spectrum of the HI 21-cm signal and Lyman-α forest flux. We consider an observational survey volume V which on the sky plane consists of a patch $L \times L$ and of line of sight thickness l along the radial direction. We consider the flat sky approximation. The Lyman-α flux fluctuations are now written as a 3-D field

$$\delta_{\mathcal{F}}(\mathbf{r}) = \left[\frac{\mathcal{F}(\mathbf{r}) - \bar{\mathcal{F}}}{\bar{\mathcal{F}}} \right] \tag{3.20}$$

The observed quantity is $\delta_{\mathcal{F}o}(\mathbf{r}) = \delta_{\mathcal{F}o}(\mathbf{r}) \times \rho(\mathbf{r})$, where the sampling function $\rho(\mathbf{r})$ is defined as

$$\rho(\mathbf{r}) = \frac{\sum_a w_a \, \delta_D^2(\mathbf{r}_\perp - \mathbf{r}_{\perp a})}{l \sum_a w_a} \tag{3.21}$$

and is normalized to unity ($\int dV \rho(\mathbf{r}) = 1$). The summation as before extends up to N. The weights w_a shall in general be related to the pixel noise. However, for measurements of transmitted hight SNR flux, the effect of the weight functions can be ignored. With this simplification we have used $w_a = 1$, so that $\sum_a w_a = N$. In Fourier space we have

$$\Delta_{\mathcal{F}}(\mathbf{k}) = \int_{-L/2}^{L/2} \int_{-L/2}^{L/2} \int_{-l/2}^{l/2} d^2\mathbf{r}_\perp dr_\parallel \; e^{i\mathbf{k}\cdot\mathbf{r}} \; \delta_{\mathcal{F}}(\mathbf{r}) \tag{3.22}$$

One may relate $\mathbf{k}_\perp$ to $\mathbf{U}$ as $\mathbf{k}_\perp = \frac{2\pi\mathbf{U}}{r}$. We have, in Fourier space

$$\Delta_{\mathcal{F}o}(\mathbf{k}) = \tilde{\rho}(\mathbf{k}) \otimes \Delta_{\mathcal{F}}(\mathbf{k}) + \Delta_{N\mathcal{F}}(\mathbf{k}) \tag{3.23}$$

where $\tilde{\rho}$ is the Fourier transform of ρ and $\otimes$ denotes a convolution defined as

$$\tilde{\rho}(\mathbf{k}) \otimes \Delta_{\mathcal{F}}(\mathbf{k}) = \frac{1}{V} \sum_{\mathbf{k}'} \; \tilde{\rho}(\mathbf{k} - \mathbf{k}') \Delta_{\mathcal{F}}(\mathbf{k}') \tag{3.24}$$

$\Delta_{N\mathcal{F}}(\mathbf{k})$ denotes a possible noise term. Similarly the 21-cm signal in Fourier space is written as

$$\Delta_{To}(\mathbf{k}) = \Delta_T(\mathbf{k}) + \Delta_{NT}(\mathbf{k}) \tag{3.25}$$

where Δ_{NT} is the corresponding noise. The cross-correlation 3-D power spectrum $P_c(\mathbf{k})$ for the two fields is defined as

$$\langle \; \Delta_{\mathcal{F}}(\mathbf{k}) \Delta_T^*(\mathbf{k}') \; \rangle = V P_c(\mathbf{k}) \delta_{\mathbf{k},\mathbf{k}'} \tag{3.26}$$

Similarly, we define the two auto-correlation multi frequency angular power spectra, $P_T(\mathbf{k})$ for 21-cm radiation and $P_{\mathcal{F}}(\mathbf{k})$ for Lyman-α forest flux fluctuations as

$$\langle \ \Delta_T(\mathbf{k})\Delta_T^*(\mathbf{k}') \ \rangle = V P_T(\mathbf{k})\delta_{\mathbf{k},\mathbf{k}'} \tag{3.27}$$

$$\langle \ \Delta_{\mathcal{F}}(\mathbf{k})\Delta_{\mathcal{F}}^*(\mathbf{k}') \ \rangle = V P_{\mathcal{F}}(\mathbf{k})\delta_{\mathbf{k},\mathbf{k}'} \tag{3.28}$$

We define the cross-correlation estimator $\hat{\mathcal{E}}$ as

$$\hat{\mathcal{E}}(\mathbf{k},\mathbf{k}') = \frac{1}{2}\left[\Delta_{\mathcal{F}o}(\mathbf{k})\Delta_{To}^*(\mathbf{k}') + \Delta_{\mathcal{F}o}^*(\mathbf{k})\Delta_{To}(\mathbf{k}')\right] \tag{3.29}$$

We are interested in the various statistical properties of this estimator. Using the definitions of $\Delta_{\mathcal{F}o}(\mathbf{k})$ and $\Delta_{To}(\mathbf{k})$ we have the expectation value of $\hat{\mathcal{E}}$ as

$$\begin{aligned}\langle \ \hat{\mathcal{E}}(\mathbf{k},\mathbf{k}') \ \rangle &= \frac{1}{2}\langle \ [\tilde{\rho}(\mathbf{k}) \otimes \Delta_{\mathcal{F}}(\mathbf{k}) + \Delta_{N\mathcal{F}}(\mathbf{k})] \times [\Delta_T^*(\mathbf{k}') + \Delta_{NT}^*(\mathbf{k}')] \ \rangle \\ &+ \frac{1}{2}\langle \ [\tilde{\rho}^*(\mathbf{k}) \otimes {\Delta_{\mathcal{F}}}^*(\mathbf{k}) + \Delta_{N\mathcal{F}}^*(\mathbf{k})] \times [\Delta_T(\mathbf{k}') + \Delta_{NT}(\mathbf{k}')] \ \rangle \end{aligned} \tag{3.30}$$

We assume that the quasars are distributed in a random fashion, are not clustered and the different noises are uncorrelated. Further, we note that the quasars are assumed to be at a redshift different from rest of the quantities and hence ρ is uncorrelated with both Δ_T and $\Delta_{\mathcal{F}}$. Therefore we have

$$\langle \ \hat{\mathcal{E}}(\mathbf{k},\mathbf{k}') \ \rangle = \frac{1}{V}\sum_{\mathbf{k}''}\langle\tilde{\rho}(\mathbf{k}-\mathbf{k}'')\rangle \times V P_{\mathcal{F}T}(\mathbf{k}'')\delta_{\mathbf{k}'',\mathbf{k}'} \tag{3.31}$$

Noting that

$$\langle\tilde{\rho}(\mathbf{k})\rangle = \delta_{\mathbf{k}_\perp,0}\delta_{\mathbf{k}_\parallel,0} \tag{3.32}$$

we have

$$\langle \ \hat{\mathcal{E}}(\mathbf{k},\mathbf{k}') \ \rangle = P_{\mathcal{F}T}(\mathbf{k})\delta_{\mathbf{k},\mathbf{k}'} \tag{3.33}$$

Thus, the expectation value of the estimator faithfully returns the quantity we are probing, namely the 3-D cross-correlation power spectrum $P_{\mathcal{F}T}(\mathbf{k})$.

We next consider the variance of the estimator $\hat{\mathcal{E}}$ defined as

$$\sigma_{\hat{\mathcal{E}}}^2 = \langle\hat{\mathcal{E}}^2\rangle - \langle\hat{\mathcal{E}}\rangle^2 \tag{3.34}$$

$$\sigma_{\hat{\mathcal{E}}}^2 = \frac{1}{2}\langle\Delta_{\mathcal{F}o}(\mathbf{k})\Delta_{\mathcal{F}o}^*(\mathbf{k})\rangle\langle\Delta_{To}(\mathbf{k}')\Delta_{To}^*(\mathbf{k}')\rangle + \frac{1}{2}|\langle\Delta_{\mathcal{F}o}(\mathbf{k})\Delta_{To}^*(\mathbf{k}')\rangle|^2 \tag{3.35}$$

We saw that

$$\langle\Delta_{\mathcal{F}o}(\mathbf{k})\Delta_{To}^*(\mathbf{k}')\rangle = P_{\mathcal{F}T}(\mathbf{k})\delta_{\mathbf{k},\mathbf{k}'} \tag{3.36}$$

and we note that

$$\langle \Delta_{To}(\mathbf{k}) \Delta^*_{To}(\mathbf{k}) \rangle = V[P_T(\mathbf{k}) + N_T(\mathbf{k})] \tag{3.37}$$

where N_T is the 21-cm noise power spectrum. We also have for the Lyman-α forest

$$\langle \Delta_{\mathcal{F}o}(\mathbf{k}) \Delta^*_{\mathcal{F}o}(\mathbf{k}) \rangle = \langle \; \tilde{\rho}(\mathbf{k}) \otimes \Delta_{\mathcal{F}}(\mathbf{k}) \; \tilde{\rho}^*(\mathbf{k}) \otimes \Delta_{\mathcal{F}}{}^*(\mathbf{k}) \; \rangle + N_{\mathcal{F}} L^2 \tag{3.38}$$

where $N_{\mathcal{F}}$ is the Noise power spectrum corresponding to the Lyman-α flux fluctuations. Using the relation

$$\langle \; \tilde{\rho}(\mathbf{k}) \tilde{\rho}^*(\mathbf{k}') \; \rangle = \frac{1}{N} \delta_{\mathbf{k}_\perp, \; \mathbf{k}'_\perp} \delta_{k_\parallel, 0} \delta_{k'_\parallel, 0} + (1 - \frac{1}{N}) \delta_{\mathbf{k},0} \delta_{\mathbf{k}',0} \tag{3.39}$$

we have

$$\langle \Delta_{\mathcal{F}o}(\mathbf{k}) \Delta^*_{\mathcal{F}o}(\mathbf{k}) \rangle = \frac{1}{V^2} \sum_{\mathbf{k_1}, \mathbf{k_2}} \langle \tilde{\rho}(\mathbf{k} - \mathbf{k}_1) \tilde{\rho}^*(\mathbf{k} - \mathbf{k}_2) \rangle \langle \Delta_{\mathcal{F}}(\mathbf{k}_{1\perp}, k_{1\parallel}) \Delta^*_{\mathcal{F}}(\mathbf{k}_{2\perp}, k_{2\parallel}) \rangle \tag{3.40}$$

or

$$\begin{aligned} \langle \Delta_{\mathcal{F}o}(\mathbf{k}) \Delta^*_{\mathcal{F}o}(\mathbf{k}) \rangle &= \frac{1}{V^2} \sum_{\mathbf{k_1}, \mathbf{k_2}} \delta_{(\mathbf{k} - \mathbf{k_1}),0} \delta_{(\mathbf{k} - \mathbf{k_2}),0} \\ &+ \frac{1}{N} \left(\delta_{(\mathbf{k}_\perp - \mathbf{k_1}_\perp),(\mathbf{k}_\perp - \mathbf{k_2}_\perp)} \delta_{(k_\parallel - k_{1\parallel}),(k_\parallel - k_{2\parallel})} \right) \times \langle \Delta_{\mathcal{F}}(\mathbf{k}_{1\perp}, k_{1\parallel}) \Delta^*_{\mathcal{F}}(\mathbf{k}_{2\perp}, k_{2\parallel}) \rangle \end{aligned}$$

This gives

$$\sigma^2_{\hat{\mathcal{E}}} = \frac{1}{2} \left[\frac{1}{N} \sum_{\mathbf{k}_\perp} P_{\mathcal{F}}(\mathbf{k}) + P_{\mathcal{F}}(\mathbf{k}) + N_{\mathcal{F}} \right] \times [P_T(\mathbf{k}) + N_T] + \frac{1}{2} P^2_{\mathcal{F}T} \tag{3.41}$$

Writing the summation as an integral we get

$$\sigma^2_{\hat{\mathcal{E}}} = \frac{1}{2} \left[\frac{1}{\bar{n}} \left(\int d^2\mathbf{k}_\perp \; P_{\mathcal{F}}(\mathbf{k}) \right) + P_{\mathcal{F}}(\mathbf{k}) + N_{\mathcal{F}} \right] \times [P_T(\mathbf{k}) + N_T] + \frac{1}{2} P^2_{\mathcal{F}T}$$

where $\bar{n}$ is the angular density of quasars and $\bar{n} = N/L^2$. We assume that the variance $\sigma^2_{\mathcal{F}N}$ of the pixel noise contribution to $\delta_{\mathcal{F}}$ is a constant and is same across all the quasar spectra whereby we have $N_{\mathcal{F}} = \sigma^2_{\mathcal{F}N}/\bar{n}$ for its noise power spectrum. An uniform weighing scheme for all quasars is a good approximation when most of the spectra are measured with a sufficiently high SNR [176]. We have not incorporated quasar clustering which is supposed to be sub-dominant as compared to Poisson noise. In reality, the clustering would enhance the term $\left(P^{1\mathrm{D}}_{\mathcal{F}\mathcal{F}}(k_\parallel) P^{2\mathrm{D}}_w + N_{\mathcal{F}} \right)$

by a factor $(1 + \bar{n}C_Q(\mathbf{k}_\perp))$, where $C_Q(\mathbf{k}_\perp)$ is the angular power spectrum of the quasars[177].

For a radio-interferometric measurement of the 21-cm signal we have [64, 111]

$$N_T(k, z) = \frac{T_{sys}^2}{Bt_0} \left(\frac{\lambda^2}{A_e}\right)^2 \frac{r_\nu^2 L}{n_b(U, \nu)} \tag{3.42}$$

Here $\lambda = 21cm(1+z)$, T_{sys} denotes the system temperature. B is the observation bandwidth, t_0 is the total observation time, r_ν is the comoving distance to the redshift z , $n_b(U, \nu)$ is the density of baseline U, and A_e is the effective collecting area of each antenna.

The expression for $\sigma_{\hat{\mathcal{E}}}^2$ has two parts - The cosmic variance term $P_{\mathcal{F}T}^2$ which owes it origin to the fact that estimates are made based on finite sampling and the term involving observational noises in the auto correlation power spectra.

The signal to noise ratio for the detection of the cross-correlation power spectrum is given by

$$SNR = \sqrt{N_c}\frac{P_{\mathcal{F}T}}{\sigma_{\hat{\mathcal{E}}}} \tag{3.43}$$

where N_c denotes the number of independent measurements of the cross-correlation.

While the cosmic variance term puts a fundamental bound to the sensitivity of a detection, it scales as $\sim \frac{1}{\sqrt{N_c}}$ and may only be improved by considering larger survey volumes.

In a typical observation N_c denotes the number of $k-$ modes that would fit the survey volume.

We shall use this analysis for predicting sensitivities of detection of the cross-power spectrum. Further, the noise $\sigma_{\hat{\mathcal{E}}}$ is crucial in the Fisher matrix analysis for model parameter estimation that we shall take up in the subsequent chapters.

3.2.1 Observational aspects

The detection of the cross-correlation signal is faced with numerous observational difficulties some of which which pose considerable challenge. Let us consider first the Lyman-α forest observations.

The main sources of errors in these observations occur from continuum fitting and subtraction. This would often result in an additive error in the estimated $\delta_{\mathcal{F}}$. This shall make the recovery of the underlying power spectrum at large scales difficult. Continuum modelling and subtraction has been studies extensively [178], [179]. Several techniques are proposed to mitigate or minimise the contribution from such additive errors. While such errors in the observation of $\delta_{\mathcal{F}}$ could have severe issues in estimating the large scale power spectrum from auto-correlation of $\delta_{\mathcal{F}}$ [180], these errors are not expected to be correlated with the 21-cm observed data. An additive error in the measurement of Lyman-α transmitted flux will however manifest as an extra contribution to the noise for the cross-correlation signal. This will invariably degrade the SNR and hence reduce the sensitivity for the detection of the cross-correlation signal. Metal line contamination is also expected to pose challenge in the detection of the Lyman-α power spectrum. We shall see how we can eliminate some parts of the Lyman-α forest spectrum to avoid the confusion from these other lines.

The major issue in the detection of the H I - 21 cm signal is the issue of foregrounds. The cosmological signal is weak as is buried under both galactic and extra-galactic foregrounds which are known to be several orders of magnitude larger. Though several foreground removal techniques are explored, it still poses a severe challenge for detecting the signal (see references in the last chapter). The foregrounds and systematics of the H I 21 cm observations and those in the Lyman-α forest are expected to be uncorrelated, thus the 21-cm signal can only be genuinely detected when detected in a cross-correlation. Residual foregrounds remaining from improper foreground subtraction will contribute to an extra source of noise for the cross-correlation signal. The fast decorrelation of the cross-correlation between the 21-cm δ_T and the Lyman-α $\delta_{\mathcal{F}}$ at two different redshifts separated by Δz with an increase of Δz, has been proposed as an important method for identifying any residual foreground contamination.

Another possible source of uncertainty in the 21-cm signal are the errors in calibrating the radio observations. This will lead to errors in the overall amplitude of the cross-correlation signal. The incomplete modelling of the neutral fraction also multiples the overall amplitude making it uncertain. In parameter estimation one has to typically marginalize over the overall amplitude of the cross-correlation power spectrum.

The cross-correlation is only possible in the region of overlap between the 21-cm observation field and the survey region of the Lyman-α forest survey. The quasar surveys generally cover much greater volume than the 21-cm observational volume. Multiple pointings of the radio array may be used by dividing the total observation time for the 21-cm observations. This helps in the reduction of the cosmic variance error by a factor of $\sqrt{N_p}$ where N_p denotes the number of radio pointings.

The data that is used to estimate the respective auto-correlations also provides an independent estimate of the cross-correlation. The sensitivity of detecting the 21-cm signal is comparable, if not higher, in the cross-correlation than its auto-correlation with the added advantage that effects of residual foregrounds and systematics are expected to be considerably less severe.

In the next chapters we shall use the cross-correlation signal as a tool to measure the large-scale matter distribution and thereby probe a variety of issues like the Neutrino mass, warm dark matter and constraining nature of dark energy.