Chapter 2

H I Power Spectrum and Foregrounds

In this chapter, we show the relation of the H I power spectrum with the dark matter density power spectrum. We discuss how visibility measurements from radio interferometers can help to estimate the H I power spectrum and present the visibility correlation-based power spectrum estimator we use in this thesis. In the end, we discuss various challenges in measuring the H I power spectrum, particularly the foregrounds, their components, and their removal methods.

2.1 The H1 Power Spectrum and Dark Matter Power Spectrum

The redshifted 21-cm radiation can be used as a probe to the H I distribution in the universe, which almost follows the underlying matter density field during CD. However, during EoR, the H I distribution is determined mainly by the ionized regions. Details of the H I distribution and related processes during CD, EoR, and post-EoR era can be found in Bharadwaj and Ali (2004, 2005). We give a simplified description here for clarity and completeness.

Along any direction in the sky **n**, the CMBR interacts with the neutral hydrogen through the spin-flip hyperfine transition. Due to this interaction, the brightness temperature of the CMBR changes from its original brightness temperature T_γ to $T(\tau_e)$, τ_e being the optical depth. The quantity of interest which quantifies the redshifted 21-cm observations is the excess brightness temperature, observed against CMB, from a redshifted z is

$$\delta T_b(\mathbf{n},z) = \frac{T(\tau_e) - T_\gamma}{1+z} \approx \frac{(T_s - T_\gamma)\tau_e}{1+z}. \tag{2.1}$$

Here T_s is the spin temperature of the H I gas in the IGM which is defined by the relative population of H I atoms between two hyperfine states *i.e.*

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} \exp(-T_*/T_s). \tag{2.2}$$

Here $g_1 = 3$ and $g_0 = 1$ are the degeneracy factors of the excited states and ground states with n_1 and n_0 being numbers of H I atoms in the respective states and $T_* = 0.068\,K = h_p\,\nu_e/k_B$, where h_P is the Planck constant and k_B the Boltzmann constant.

Depending upon the spin temperature T_s and the CMB temperature T_γ, the redshifted 21-cm signal will either be observed in absorption or emission. This absorption/emission signal from the intervening H I is redshifted to a frequency $\nu = \nu_e/(1+z)$ for an observer at a redshift z. The redshift is introduced due to the universe's expansion and the H I peculiar velocity.

The excess brightness temperature of the redshifted 21- cm signal can be expressed as $\delta T_b(\mathbf{n},z) = \bar{T}(z) \times \eta_{H\ i}\ (\mathbf{n},z)$ where

$$\bar{T}(z) = 4.0\ mK(1+z)^2 \left(\frac{\Omega_b h^2}{0.02}\right)\left(\frac{0.7}{h}\right)\frac{H_0}{H(z)} \tag{2.3}$$

depends only on z and the background cosmological parameters, and

$$\eta_{H\,i}\,(\mathbf{n},z) = \frac{\rho_{H\,i}}{\bar{\rho}_H}\left(1-\frac{T_\gamma}{T_s}\right)\left[1-\frac{(1+z)}{H(z)}\frac{\partial v}{\partial r}\right] \tag{2.4}$$

is the "21- cm radiation efficiency" introduced by Madau et al. (1997); the term in square bracket is the effect due to H I peculiar velocity (Bharadwaj and Ali, 2004, 2005). The 21-cm radiation efficiency ($\eta_{H\,i}$) incorporates the details of the H I evolution and the effects of the growth of large-scale structures and varies with position ($\mathbf{x} = r_\nu \mathbf{n}$) and redshift, r_ν is the comoving distance. The H I power spectrum can be defined as

$$\langle \eta_{H\,i}\,(\mathbf{k},z)\,\eta^*_{H\,i}\,(\mathbf{k}^{'},z)\rangle = (2\pi)^3\,\delta^3_D(\mathbf{k}-\mathbf{k}^{'})\,P_{H\,i}\,(\mathbf{k},z) \tag{2.5}$$

where δ^3_D is the three-dimensional Dirac delta function and $\eta_{H\,i}\,(\mathbf{k},z)$ is the Fourier transform of the 21- cm radiation efficiency $\eta_{H\,i}\,(\mathbf{n},z)$.

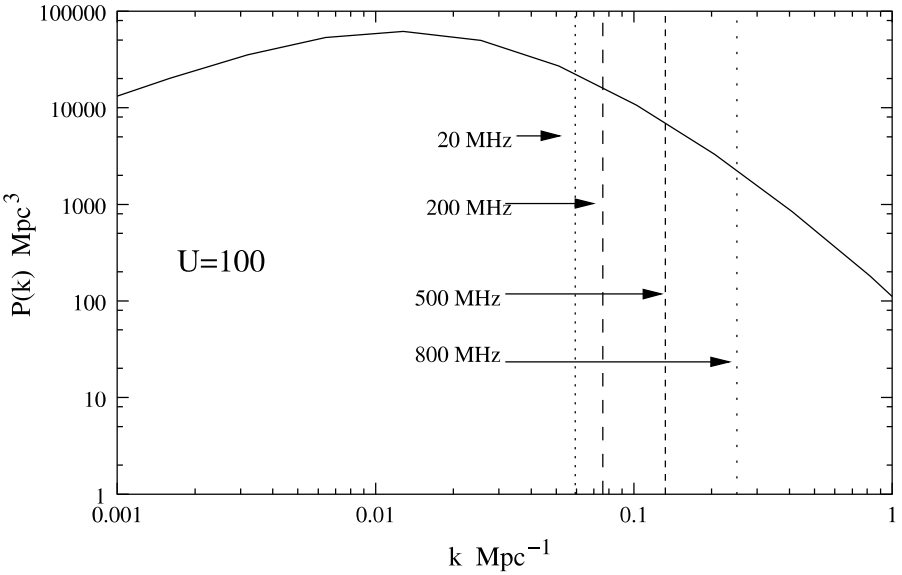

Fig. 2.1 This shows the linear power spectrum of dark matter density fluctuations at the present epoch. A baseline u at a frequency ν will probe the power spectrum at all Fourier modes $k \geq k_{min} = (2\pi/r_\nu)U$. The values of k_{min} are shown at different frequencies for $U = 100$. This figure can be used to determine the length scale probed by any baselines U. The figure is adapted from Bharadwaj and Ali (2005).

After the recombination happened, the universe was filled with neutral hydrogen with small fluctuations in the neutral hydrogen density. The fluctuations in the H I density $\rho_{H\ i}$ follows the dark matter density fluctuations Δ, i.e. $\Delta\rho_{H\ i}\ (\mathbf{k},z) = \bar{\rho}_{H\ i}\ \Delta(\mathbf{k},z)$. The fluctuation in the neutral hydrogen density $\Delta\rho_{H\ i}$ will cause the spin temperature fluctuation ΔT_s. The dark matter also determines the H I peculiar velocities. Hence, the hydrogen density and peculiar velocities can be assumed to follow the dark matter with some bias. As the EoR 21- cm signal from the spin-flip transition of neutral hydrogen depends on the H I density fluctuations and peculiar velocities which follow the dark matter, studying the H I power spectrum can give information about the dark matter distribution. The detailed analysis is presented in Bharadwaj and Ali (2005). The H I power spectrum $P_{H\ i}$ is related to the dark matter power spectrum P(k) as (Bharadwaj and Ali, 2005)

$$P_{H\ i}\ (\mathbf{k},z) = \left[(1-\frac{T_\gamma}{T_s})(1+f\mu^2)+\frac{T_\gamma}{T_s}s\right]^2 P(\mathbf{k},z) \tag{2.6}$$

The term $f\mu^2$ comes from the Fourier transform of $[(1+z)H(z)](\partial v/\partial r)$, the H I peculiar velocities term. Here μ is the cosine of the angle between $\mathbf{k}$ and the line of sight $\mathbf{n}$, and in a spatially flat universe $f(\Omega_m) \approx \Omega_m^{0.6} + \frac{1}{70}\left[1-\frac{1}{2}\Omega_m(1+\Omega_m)\right]$. The fluctuations $\Delta\rho_{H\ i}$ which causes the fluctuations in the spin temperature ΔT_s is quantified using a dimensionless function $s(z)$ defined such that $\Delta T_s = s\ T_s\ \Delta\rho_{H\ i}\ /\rho_{H\ i}$. If the P_{HI} can be determined by observing the angular fluctuations in the H I signal by any means, we can directly probe the dark matter power spectrum with the help of the H I power spectrum given the evolution of radiation temperature T_γ and T_s are known. The figure 2.1 is taken from Bharadwaj and Ali (2005). It shows the part of the dark matter power spectrum which will be probed at different frequencies by the visibility correlations at a baseline $U = 100$ wavelengths.

2.2 Interferometers and visibility measurements

In the Raleigh-Jeans limit, the brightness temperature and specific intensities of emission are related as

$$I_\nu = [2k_B/\lambda_e^2(1+z)^2] \times T_b. \tag{2.7}$$

A radio interferometer is a collection of antennae, with each antenna recording the electric field incident on it. The interferometers evaluate the spatial coherence function of the signal by cross-correlating the electric fields from each pair of antennae. This spatial coherence function is known as visibility, which, with certain approximations, gives the Fourier transform of the specific intensity I_ν (Zernike, 1938).

For a pair of antennae at a given time, the positional separation of antennae **d**, projected perpendicular to the direction of observation in units of the observing wavelength λ is denoted with a dimensionless quantity **U**, which is referred as 'baseline vector'. Usually, in radio-interferometric convention, the baseline **U** is expressed in units of kilo wavelength ($k\lambda$). We also follow the same convention and, throughout the thesis, express **U** in units of kilo wavelengths ($k\lambda$). Visibility $V(\mathbf{U},\nu)$ is a function of baseline, and it is recorded for every pair of antennae at different frequency channels in the frequency range of observation. Note that due to the apparent motion of the source in the sky, the baseline **U** as measured by the same pair of antenna changes with time, as well.

A typical radio interferometric array simultaneously measures visibilities at many baselines and frequency channels. Ideally, each visibility records a single mode of the Fourier transform of the specific intensity distribution $I_\nu(\theta)$ on the sky, where θ is a two-dimensional vector in the plane of the sky. In reality, the specific intensity distribution of the sky measured is multiplied by the antenna beam response $A(\theta,\nu)$ which quantifies how the individual antenna, pointing upwards, responds to signals from different directions

in the sky. Hence, the visibility measurements can be given as

$$V(u,v,w) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_\nu(l,m) I_\nu(l,m) e^{-2\pi i(ul+vm+wn)} \frac{dldm}{\sqrt{1-l^2-m^2}} \tag{2.8}$$

where (u, v, w) are the components of the baseline vector **U** and (l, m, n) specifies the direction cosines with respect to the phase center, associated with θ. $A_\nu(l,m)$ is the normalized antenna beam response pattern, and $I_\nu(l,m)$ is the specific intensity of the source at frequency ν. These visibilities are later Fourier inverted to determine the specific intensity distribution $I(\theta,\nu)$, which is known as image. Usually, the visibility at zero spacing $\mathbf{U} = 0$ is not measured, and a uniform specific intensity distribution does not contribute to the visibilities. Radio interferometers pick only the fluctuations in the specific intensity $I(\theta,\nu)$.

In practice, the visibility is not a continuous function but is sampled only at particular places in the uv-plane. This sampling can be described by a sampling function $S(u,v) = \sum_i \delta_{2D}(u-u_i, v-v_i)$, u_i, v_i being the points in the uv-plane where visibilities are sampled, and δ_{2D} is the 2D delta function. At a particular instant of time, each antenna pair of an array measures the visibility at a single uv-point corresponding to the distance between the antenna pair. As time progresses, with the Earth's rotation, antenna pairs trace elliptical curves called uv-tracks and sample more and more points in the uv-plane. This is what is referred to as the "Earth Rotation Synthesis". Figure 2.2 is the uv-coverage of uGMRT at 130 MHz, pointing towards the declination 50°. The black curves are the uv-tracks traced by antenna pairs, consisting of closely placed discrete points, showing the points in uv-space where visibilities are sampled. This is for 8 hours of observation time. The uv-points are denser in the central regions of the uv-plane compared to the outer parts due to the availability of more antenna pairs at shorter spacing. As we go into the outer region of the uv-plane, the number of baselines decreases.

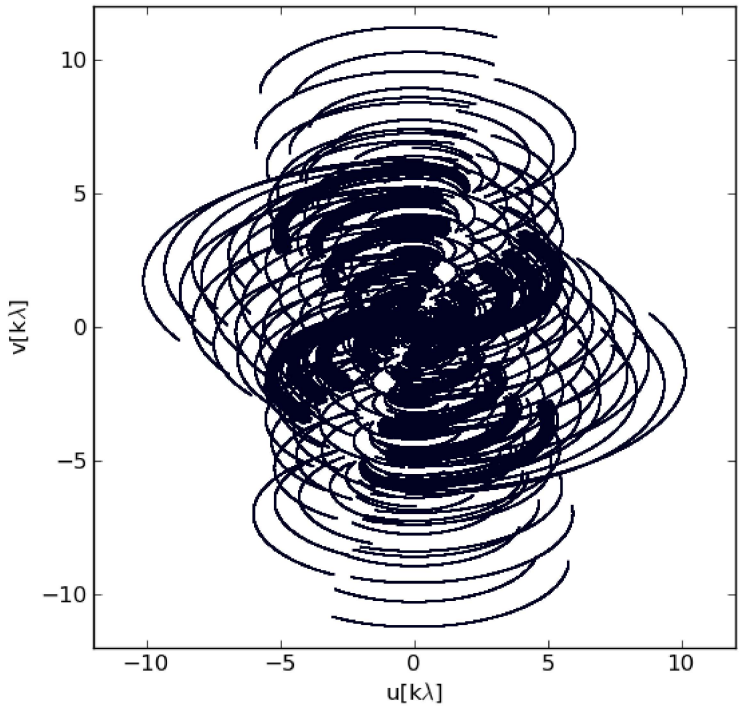

Fig. 2.2 The sample uv-coverage of the uGMRT at 130 MHz, for 8 hours of observation. This shows the points in uv-plane where visibilities are sampled. Here, (u, v) are the antenna separations in wavelength units at the observing frequency of 150 MHz.

In interferometric terms, the spatial coherence function of the electric field of the radiation at the position of the antenna is denoted as the 'sky' visibility $\tilde{V}^S(\vec{U})$. Here we use the symbol '~' to denote the complex nature of the visibility and other quantities. The visibility function observed by a pair of the antenna includes the astronomical signals from the objects as well as an unavoidable component called noise. The main sources of noise are receivers, Earth's atmosphere, and the random radio signals from the background sky. The noise is usually quantified in terms of temperature, and the total noise of the system is denoted by the system temperature, which is a measure of the total random noise in the system while looking at a blank sky. The other contributions of the noise come from the spillover noise, the radiation picked by the antenna feeds from the ground, and any lossy

element in the feed path. The noise contributed from the various sources discussed above is random in nature. Hence, the system noise $\tilde{N}(\vec{U})$ present in the visibility measurements can be considered Gaussian random with zero mean. Given antenna characteristics and its source equivalent flux density (SEFD), frequency width of the channel $\Delta\nu$ and integration time for each visibility $\Delta\tau$, the standard deviation of the noise in the real or imaginary part of each visibility can be written as (Thompson et al., 1986)

$$\sigma_N(\vec{U}) = \frac{\text{SEFD}}{\sqrt{2\Delta\nu\,\Delta\tau}}. \tag{2.9}$$

It is safe to assume that the interferometric noise is uncorrelated in time. Hence, its auto-correlation functions are zero except at zero delay.

2.3 Power Spectrum Estimator

A widely used statistical property of the sky brightness distribution is its power spectrum (Bharadwaj and Sethi (2001); Lazarian (1995) etc.). As the redshifted 21-cm signal is expected to be faint and hard to detect with imaging, estimating its power spectrum or equivalently intensity mapping gives a possible probe of the evolution of the baryonic matter distribution over cosmic time.

The power spectrum of the sky brightness distribution can be estimated using images of the sky and the visibilities measured from the radio interferometric observations. Dillon et al. (2015) have proposed an image-based estimator to measure the three-dimensional (3-D) power spectrum P(k) of the redshifted 21- cm signal intensity fluctuations. An image-based power spectrum estimator for the angular power spectrum C_ℓ is proposed by Seljak (1997) and used by Bernardi et al. (2009) and Iacobelli et al. (2013) to measure C_ℓ of the Galactic diffuse synchrotron emission using 150MHz observations with WSRT and LOFAR respectively.

It is convenient to use the measured visibilities from radio interferometers to estimate the power spectrum (Bharadwaj and Sethi, 2001; Bharadwaj and Ali, 2005; Bharadwaj and Pandey, 2003). The visibility-based power spectrum estimators have been extensively used in literature to measure the power spectrum. Visibility-based estimators for the three-dimensional 21-cm signal power spectrum P(k) are proposed by Liu and Tegmark (2012) and Trott et al. (2016). Paul et al. (2014) have proposed that visibility correlations can be used to detect the EoR signal using drift scan observations with the MWA, which is further used by Patwa and Sethi (2019) to detect the 21- cm power spectrum for drift scan. Visibility-based power spectrum estimators have also been used to analyze GMRT data in the context of H I observations by Ali et al. (2008); Ghosh et al. (2011, 2012); Paciga et al. (2011), etc.

The visibility-based power spectrum estimator and its various variants (Bharadwaj et al. (2019); Choudhuri et al. (2014, 2016, 2019); Datta et al. (2007) etc.) has been used to estimate the power spectrum of H I distribution in the ISM of nearby galaxies by Begum et al. (2006); Dutta and Bharadwaj (2013); Dutta et al. (2009); Nandakumar and Dutta (2020), and the angular power spectrum of the Galactic diffuse synchrotron emission (Chakraborty et al., 2019a; Choudhuri et al., 2017a, 2020; Ghosh et al., 2012).
The main benefit visibility-based estimators have over image-based estimators is that they are free from imaging artifacts. To remove the noise bias, most of the image-based estimators depend on modeling the noise in the data and subtracting the expected noise bias (Seljak, 1997). On the other hand, the visibility-based estimators, which individually correlate visibilities, avoid the noise bias by excluding the visibility self-correlations (Begum et al., 2006; Dutta and Bharadwaj, 2013).

In this thesis, we use the visibility-based power spectrum estimator discussed in Choudhuri et al. (2014). As the visibilities in the uv-plane are sparsely sampled, first, we grid the visibility in the uv-plane and then estimate the power spectrum by correlating the

visibilities in each grid. The power spectrum at baseline $\vec{U}$ can be calculated by correlating the visibilities within a grid around those baselines. Given an angular field of view of radius θ_0 to which the telescope is sensitive, it has been shown (Bharadwaj and Sethi, 2001; Bharadwaj and Ali, 2005; Choudhuri et al., 2014) that the visibilities in the nearby baselines remains correlated to a baseline separation of $\Delta U < \frac{1}{\pi\theta_0}$. The size of the uv-grids is chosen such that they are large enough to include a sufficient number of baselines in a given uv-grid and small enough to have all visibilities in the uv-grid correlated. In the figure 2.3 we schematically show the gridding of the visibilities in the uv-space. As mentioned earlier, we first grid the visibilities in the uv-space and then correlate the visibilities in the grids to estimate the power spectrum. We consider a grid in an annular bin of radius U; the black dots inside that grid represents the visibility measurements present in that grid. Similarly, there will be visibility measurements in other grids as well (which we haven't shown in the cartoon). In each uv-grid, we estimate the power spectrum by correlating visibilities only in nearby baselines, omitting the visibility auto-correlations. This is referred to as a nearby baseline correlation, which drastically reduces the noise bias in estimates of the power spectrum in uv-grids. The contribution from each uv-grid within a given annulus in $U = \mid \vec{U} \mid$ is then combined, and the real part of it is used to quote the value of the isotropic power spectrum for the baseline separation U. We may schematically write it as

$$\mathscr{E}\left\{P(U)\right\} = \mathscr{R}\left[\langle \tilde{V}(\vec{U})^* \tilde{V}(\vec{U}+\Delta\vec{U})\rangle\right]. \tag{2.10}$$

Here, $\mathscr{R}$ is denoting the real part, and the angular brackets denote the ensemble average and have to be taken over many realizations of the sky. It is practically not feasible to have many independent observations of the sky. We assume that the fluctuations are isotropic and the power spectrum depends on $|U|$. Hence, the ensemble average above can be replaced by the average over the independent estimates of the visibility correlation in an annular region in the uv-plane at U. To have an estimate of the power spectrum $P(U)$ at

a particular baseline **U**, we consider an annular bin at baseline **U**. We measure $P(U)$ by correlating the visibilities present in the grid. Within these grids, the individual estimates of P are not independent; however, estimates of P from different such grids are independent. Hence, we average all the individual estimates of P possible within the grids, and then we average over all such grids within the annulus to get $P(U)$.

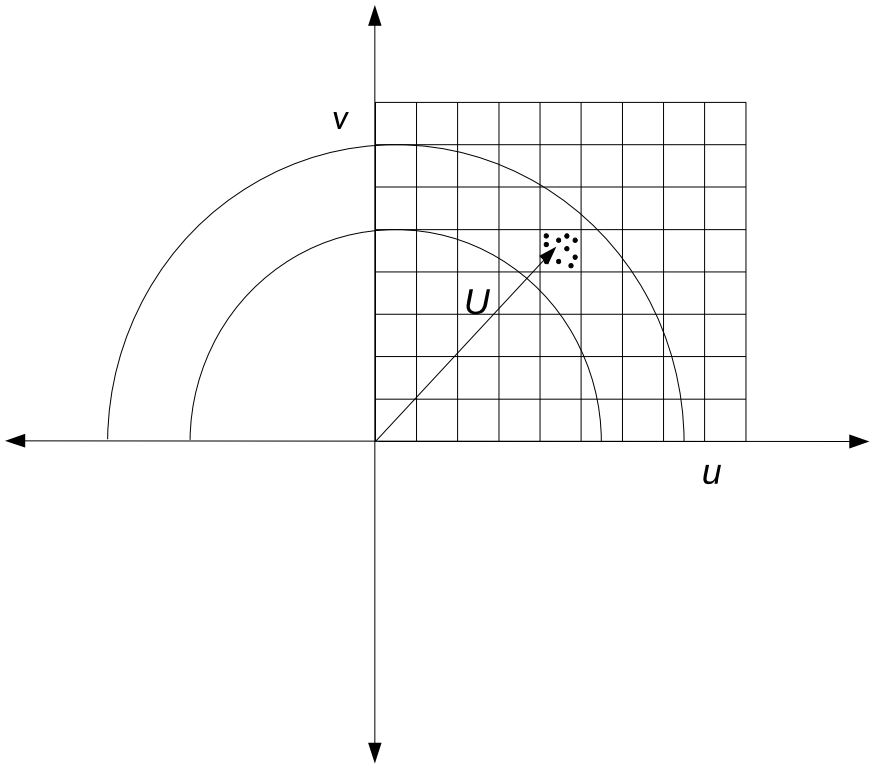

Fig. 2.3 This cartoon schematically shows the gridding in uv-space. The black dots in the grid under consideration represents the measured visibilities which will be correlated to estimate the power spectrum at a baseline **U**.

Note that the power spectrum estimator here assumes that a perfect calibration is done and the gains are all unity. In such a case, the power spectrum estimate has no bias arising from instrumental noise, and its uncertainties can be written as (Ali et al., 2008; Dutta, 2011)

$$\sigma_P^2 = \frac{P^2(U)}{N_G} + 2\frac{P(U)\sigma_N^2}{N_B} + 2\frac{\sigma_N^4}{N_B}, \tag{2.11}$$

where N_G is the number of independent estimates of the power spectrum in a given U-bin, N_B is the total number of visibility pairs possible in the bin, and N_2 is the noise variance. The first term is the contribution from the sample variance, and the other two terms come from the noise correlation. The sample variance decreases with the baseline. For expected H I 21- cm power spectrum estimates of cosmological origin, the contribution from the sample variance term is negligible in the baseline ranges of our interest, and the last term dominates over the two.

A variation of the visibility-based estimator more relevant to the EoR observation is the Tapered Gridded Estimator (TGE) (Choudhuri et al., 2016). The TGE uses a Gaussian kernel to taper the response of compact sources far away from the phase center of observation, hence reducing the effects of foregrounds clipping in through beam sidelobes. In this thesis, we consider that the power spectrum estimator we used here uses the same tapering. This also helps us in the creation of the foreground sky model in a restricted region of the sky.

There are other versions of the visibility-based power spectrum which are discussed below. The multifrequency angular power spectrum (MAPS) (Datta et al., 2007) quantifies the statistics of the sky brightness distribution as a function of angular multipole (ℓ) and frequency (ν). This assumes that the sky brightness temperature fluctuations are Gaussian random in nature, statistically homogeneous, and isotropic on the sky. For a particular particular frequency ν, the brightness temperature fluctuations across the sky $\delta T_b(\mathbf{n},\nu)$, can be decomposed in terms of spherical harmonics as $Y_\ell^m(bfn)$ as $\delta T_b(\mathbf{n},\nu) = a_{\ell,m}(\nu)Y_\ell^m(\mathbf{n})$. The MAPS is defined as $C_\ell(\nu_1,\nu_2) = \langle a_{\ell,m}(\nu_1)(a^*_{\ell,m}(\nu_2)\rangle$, where the angular brackets $\langle ...\rangle$ denote an ensemble average over different statistically independent realizations of the brightness temperature fluctuations field $\delta T_b(\mathbf{n},\nu)$. Under the assumption that the signal is ergodic, the visibility correlations at nearby baseline give estimates of angular power spectrum as $C_{2\pi U}(\Delta\nu) \sim \langle \tilde{V}(\mathbf{U},\nu)\tilde{V}(\mathbf{U}',\nu+\Delta\nu)\rangle$. The angular multipole ℓ and baseline

U are related as $\ell = 2\pi U$.

The 3D power spectrum of the redshifted 21-cm brightness temperature fluctuations $P(k_\perp, k_\parallel)$ can be estimated using the MAPS. Assuming the ergodicity of the signal along the line of sight, for the small bandwidth the MAPS can be given as $C_\ell(\nu_1, \nu_2) = C_\ell(\Delta\nu)$ where $\Delta\nu = |\nu_1 - \nu_2|$. Under the flat sky approximation the 3D power spectrum $P(k_\perp, k_\parallel)$ becomes the Fourier transform of $C_\ell(\Delta\nu)$ (Datta et al., 2007).

The primary beam pattern of the telescope is frequency dependent and usually not well quantified at the locations away from the pointing center, especially around nulls and side lobes. Thus the bright foreground sources located near these nulls and side lobes introduce problems, such as ripples along the frequency direction, in estimating the angular power spectrum. Such problems can be avoided by tapering the sky response using some kind of frequency-independent window function. Motivated by this, various estimators have been developed. The visibility-based tapered gridded estimators (TGE) for the redshifted 21-cm power spectrum (Choudhuri et al., 2016) suppresses the contribution from the sources lying in the outer regions of the telescope's field of view by tapering the visibilities using a window function. In this TGE the tapering is achieved by convolving the measured visibilities with the window function. Later these convolved visibilities are used to estimate the power spectrum. The extensions of the tapered gridded estimator are the image-based tapered gridded estimator *ITGE* (Choudhuri et al., 2019) for the angular power spectrum and a Tapered Gridded Estimator $C_\ell TGE$ (Bharadwaj et al., 2019), for the *MAPS* to estimate the power spectrum P(k) of redshifted H I 21-cm brightness temperature fluctuations. The TGE uses the gridded visibilities to estimate the power spectrum. Gridding the visibility makes this estimator computationally much faster than the traditional visibility correlation-based estimators, which individually correlate the visibilities. The positive noise bias of these estimators is removed by subtracting the auto-correlation of the visibilities. For the drift scan observations, where the antennas are fixed to the ground and the sky above the

antennae changes continuously due to the earth's rotation, the Tracking Tapered Grided Estimator TTGE (Chatterjee et al., 2022) has also been developed which is another version of TGE. The TTGE quantifies the power spectrum of the sky signal estimated directly from the visibilities measured in drift scan radio interferometric observations.

The flagging of the channels affected by the presence of RFI results in the non-uniform sampling of the bandpass response of the instrument. This, while transforming the visibilities from the frequency domain to the delay domain (Fourier Transformation), contaminates the otherwise foreground-free delay modes (Chakraborty et al., 2022) and the separation of H I 21- cm signal from the smooth foreground becomes difficult. The delay-transformed visibility can be given as the convolution of Fourier-transformed visibility with the convolving kernel resulting from the Fourier conjugate of the non-uniform bandpass response of the instrument. This convolution will result in leakage of the spectrally smooth foregrounds beyond the horizon limit into the EoR window (Chakraborty et al., 2022). To restrict the spectral leakage, Chakraborty et al. (2021, 2022) discuss a power spectrum estimator which uses 1-dimensional Hogbom clean algorithm across the delay-axis to deconvolve this convolving kernel. This simple deconvolution technique which uses the delay space CLEANing effectively deconvolves the convolving kernel and reduces the leakage of foreground beyond the horizon limit due to flagged channels.

2.4 Foregrounds

At present several experiments are planned to estimate the power spectrum of the brightness temperature fluctuation (Bharadwaj and Sethi, 2001; Bharadwaj and Ali, 2005) from the epoch of reionization, including the experiments with the Giant Metrewave Radio Telescope (Paciga et al., 2013), Low-Frequency Array (van Haarlem et al., 2013), Murchison Widefield Array (Bowman et al., 2013; Dillon et al., 2015; Tingay et al., 2013), the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER; Ali et al.

(2015); Parsons et al. (2010)), the Hydrogen Epoch of Reionization Array (DeBoer et al., 2017), the Square Kilometer Array (Koopmans et al., 2015; Mellema et al., 2013) etc. One of the major challenges in the detection of H I 21- cm signal from the cosmological origin is the presence of four to five orders of magnitude stronger emission from the astrophysical foregrounds at the observation frequency of the redshifted 21-cm emission (Ali et al., 2008; Bernardi et al., 2010; Shaver et al., 1999). The foreground emissions include the radiation from the compact sources such as the radio galaxies as well as the Galactic diffuse synchrotron radiation and free-free emissions from the Galaxy and other galaxies (Ali et al., 2008, 2016; Choudhuri et al., 2020; Cooray and Furlanetto, 2004; Di Matteo et al., 2002; Ghosh et al., 2012; Oh and Mack, 2003; Santos et al., 2005; Shaver et al., 1999). Bright compact sources and the Galactic diffuse synchrotron emission dominate the foreground radiation at redshifted 21- cm signal observation frequency $\sim$ 150 MHz. In contrast, the free-free emissions from our Galaxy and external galaxies make much smaller contributions to the total foreground emission. However, each of these is individually larger than the H I signal.

As astrophysical foregrounds are several orders of magnitude large compared to the redshifted 21- cm signal in the observation frequency range, it becomes crucial to study and accurately characterize the foregrounds to detect the H I signal through various interferometric experiments.

Various efforts have been made to study and characterize the foregrounds in detail in the context of the EoR 21- cm signal measurements. Shaver et al. (1999) have studied the diffuse synchrotron and free-free emission from our Galaxy and extragalactic sources; Di Matteo et al. (2002) and Di Matteo et al. (2004) have considered the emission from compact extragalactic sources at low radio frequencies. The effect of free-free emission from extragalactic sources like haloes is also studied by Oh and Mack (2003) and Cooray and Furlanetto (2004). The most dominant component of the foregrounds is the Galactic

diffuse synchrotron emission and the radiation from extragalactic point sources. The free–free emissions from our Galaxy and external galaxies contribute very less and are around 1 percent of the total foreground budget Shaver et al. (1999). Here we discuss the two dominant components of the foreground *viz* extragalactic point sources and the Galactic diffuse synchrotron emission.

2.4.1 Extragalactic point sources

The bright, compact sources are expected to dominate the 150 MHz radio sky, especially at the angular scales important for EoR observations. Hence it is of considerable importance to study the properties of these sources (Dunlop and Peacock, 1990; Garn et al., 2007; Jackson and Wall, 1999; Peacock, 1999; Seymour et al., 2004). The contribution from extragalactic point sources is primarily due to the emission from normal galaxies, radio galaxies, star-forming galaxies, and active galactic nuclei (Condon et al., 2012; Santos et al., 2005; Singal et al., 2010). Different radio surveys have been conducted at various frequencies with a wide range of angular resolutions, and differential source counts at different frequency bands are estimated with the help of these surveys. These full-sky surveys include the NRAO VLA Sky Survey (NVSS) at 1.4 GHz (Condon et al., 1998) Faint Images of the Radio Sky at Twenty-centimeters (The VLA FIRST Survey) (Becker et al., 1995) etc. At rather lower frequencies of 73.8 - 231 MHz there have been several full sky surveys, namely the VLSS (Cohen et al., 2007), VLSSr (Lane et al., 2014), 8C (Rees, 1990), MSSS-LBA and MSSS-HBA (Heald et al., 2015), LoTTS (Shimwell et al., 2017, 2019) etc. In a recent deep and wideband observation of the ELAIS N1 field with the uGMRT band 3 Chakraborty et al. (2019b) has estimated the differential source count at 400 MHz. Differential source counts at $\sim$ 150 MHz have been calculated from a single deep GMRT integration and a larger-area GMRT survey centered on the Boötes field (Intema et al., 2011; Williams et al., 2016), from the 7C survey (Hales et al., 2007;

McGilchrist et al., 1990), from deep, small-area surveys with the LOFAR HBA system Hardcastle et al. (2016); Mahony et al. (2016); Williams et al. (2016), from the MWA GLEAM survey (Wayth et al., 2015) as well as from deep but single-pointing MWA survey Franzen et al. (2016, 2019); Hurley-Walker et al. (2016) etc. Intema et al. (2017) has used the entire TGSS survey to estimate the differential source count at 150 MHz based on the TGSS-ADR1 (Intema et al., 2017). Differential source counts $\frac{dN}{dS}$ from various surveys were found to be consistent with each other.

From the analysis of large samples of nearby radio galaxies, it has been established that the point sources are clustered, and the clustering of the sources can be quantified by the two-point correlation function of the sources, which usually follows a power-law dependence (Blake and Wall, 2002; Blake et al., 2004; Cress et al., 1996; Magliocchetti et al., 1998; Overzier et al., 2003; Peebles, 1993; Rengelink, 1999; Wilman et al., 2003). In the TGSS at 150 MHz Rana and Bagla (2019) has studied the angular clustering of point sources, and they found that the amplitude of clustering varies with the flux limit. These results are in agreement with other previous works (Magliocchetti et al., 2017; Peacock and Nicholson, 1991; Rengelink, 1999), etc.). These clustering properties and the differential source counts can be used to estimate angular power spectrum (Ali and Bharadwaj, 2014; Ali et al., 2008) quantifying the foreground contamination due to the clustering and Poisson contribution of point sources for low-frequency H I intensity mapping surveys to study the epoch of reionization.

2.4.2 Galactic Diffuse Synchrotron Emission

Galactic foregrounds mainly have three components. Galactic diffuse synchrotron emission (GDSE) is the most prominent component, which is the dominant component of foregrounds at the frequency ranges we are interested in. The other components are synchrotron emission from discrete sources, mostly supernova remnants (SNRs), and the

free-free emission from diffuse ionized gas. These are relatively weak and contribute very less to the total foreground emission. The GDSE is believed to originate from the interaction between the cosmic ray electrons in the interstellar medium and the Galactic magnetic field (Rybicki and Lightman, 1979). The intensity of the synchrotron emission or its brightness temperature depends on position and frequency. A power law can describe both position and spectral dependence. In the context of EoR observations, the diffuse component of the foreground is not very accurately modeled, and observations with very high resolution are not available at the relevant frequency and angular scales. An all-sky map of GDSE intensity at 150 MHz with a rather low resolution of 5° is available (Landecker and Wielebinski, 1970). Similarly, there are other all-sky maps at 408 MHz survey with a resolution of 0.5° by Haslam et al. (1982) (The Haslam Map) and by Reich and Reich (1986); Reich (1982) at 1420 MHz with 0.95° resolution. Recently Choudhuri et al. (2020) have used the TGSS survey at 150 MHz and estimated the all-sky angular power spectrum C_ℓ of Galactic diffuse synchrotron emission, before and after subtracting the point sources. They found that C_ℓ decreases significantly after subtracting the point sources. The angular power spectra of the GDSE determined by (La Porta et al., 2008) at angular scales greater than 0.5° using the Haslam map and the one by (Reich and Reich, 1986; Reich, 1982) at 1.4 GHz, and the other observational measurements of the angular power spectra C_ℓ at 150 MHz done by (Bernardi et al., 2009; Choudhuri et al., 2020; Ghosh et al., 2012; Iacobelli et al., 2013; Parsons et al., 2010), etc., establishes the fact that the angular power spectra of GDSE follow a power-law dependence both in frequency and angular position. For the high Galactic altitude region, at around a frequency range of 150 MHz, the power-law index of angular scale varies from -1. to -3.0, whereas the value of temperature spectral index is -2 to -2.8 (Giardino et al., 2002; Jelić et al., 2008; Platania et al., 1998; Santos et al., 2005; Tegmark et al., 2000). In a recent study, in the

frequency range of 150 - 408 MHz, the mean spectral index of the synchrotron emission at high Galactic latitude is found to be -2.5 Rogers and Bowman (2008).

The polarized foreground presents another challenge in the detection of 21- cm signal. Reich (2006) reviews various polarization survey projects and summarizes their results. The extragalactic compact radio sources are usually very weakly polarized or unpolarized in the frequency range of EoR observations, but the diffuse emission from the Galaxy is polarized with the strength $\sim$ 1K (Jelić et al., 2010). Polarized foreground ionospheric and instrumental effects modify the total intensity part of the visibility signal. However, since the polarized intensity in the foreground is much smaller than its total intensity and it is expected that the instrumental effects are smaller, the effect of polarized foreground and instrumental effects from polarization leakage, etc. is rather small compared to the direct effect from gain errors. In this thesis, we neglect the polarization-related effects, these higher-order effects will be considered in following works.

A considerable amount of effort has been invested in simulating the foregrounds (Bowman et al., 2009; Gleser et al., 2008; Harker et al., 2009; Petrovic and Oh, 2011; Sun et al., 2008). Jelić et al. (2008, 2010) also simulate the foreground and presented a detailed astrophysical foreground simulation that includes emission of both Galactic and extragalactic origin as a part of the LOFAR-EoR key science project, the (Jelić et al., 2010) presents simulations for galactic polarized foregrounds. Although these simulations are important in developing algorithms to remove the effect of foregrounds and detect the redshifted 21- cm signal, they require guidance from observations. It is crucial to characterize them accurately using real observations.

Multiple observations of the foregrounds around the frequency of 150 MHz have been done to characterize both the dominant components of foregrounds. Using 14 hours of observation with GMRT at 153 MHz, Ali et al. (2008) have characterized the extragalactic radio sources and the Galactic diffuse synchrotron radiation. They have ignored our

Galaxy and external galaxies' free-free emissions as they contribute much less to the total foreground budget. In the same project, later Ghosh et al. (2012) observed three more fields and estimated the angular power spectrum of Galactic foregrounds after subtracting the bright, compact sources. Authors have also estimated the differential source counts $\frac{dN}{dS}$ from their source catalog, and their results are in agreement with those of the available radio surveys in the literature. Bernardi et al. (2009) have used the 150 MHz WSRT observations and done a similar analysis by subtracting the point sources above 15 mJy and using the resulting image to estimate the angular power spectrum C_ℓ. These studies report that a power law can describe the angular power spectrum C_ℓ at angular scales $\ell \leq 1000$.

The foreground components discussed here are continuum sources and have a good frequency dependence. As discussed above, based on the various studies in literature, it is well accepted that power laws can model the angular scale and frequency dependence of these continuum foregrounds(Santos et al. (2005), etc.). We model the multi-frequency angular power spectrum for different components of the foreground as

$$C_\ell(\nu_1,\nu_2) = A\left(\frac{1000}{\ell}\right)^{\beta}\left(\frac{\nu_f}{\nu_1}\right)^{\alpha}\left(\frac{\nu_f}{\nu_2}\right)^{\alpha} I_l(\nu_1,\nu_2) \tag{2.12}$$

where A is the amplitude and β and α (spectral index) are the power law indices for the ℓ and the ν dependence respectively, $\nu_f = 130$MHz, and $I_l(\nu_1,\nu_2)$ is a parametric function that describes the departure from the complete correlation of foregrounds in frequency. It is expected that the foregrounds are highly coherent in such cases, i.e. $I_l(\nu_1,\nu_2) \simeq 1$. For the frequency range we are interested in, $I_l(\nu_1,\nu_2)$ can be given as

$$I_l(\nu_1,\nu_2) \sim 1 - \log^2(\nu_1/\nu_2)/(2\zeta^2). \tag{2.13}$$

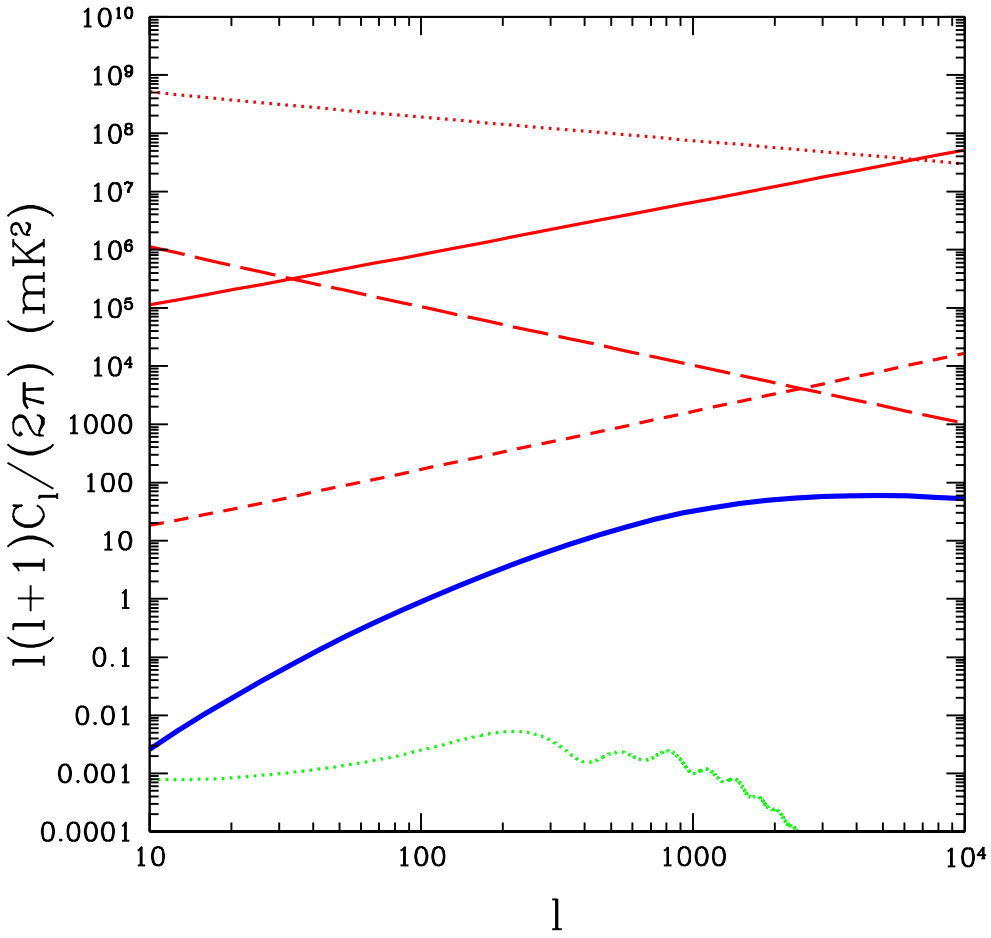

Fig. 2.4 the angular power spectrum plot for various components of foregrounds at z=9.2 ($\nu = 140$ MHz), along with the 21- cm power spectrum (thick blue line). The red thin solid line is for the point sources and the red thin dotted line represents the Galactic Synchrotron emission. The extragalactic free-free and Galactic free-free components are shown with a red thin dashed line and a red thin long-dashed line, respectively. The green dot-dashed line represents the CMB power spectrum. The figure is reproduced from Santos et al. (2005).

with ζ being the correlation length. The model discussed above is adapted from the one given in Santos et al. (2005). In Table 2.1 the parameters of the fiducial models for the four different components of foregrounds are given.

	$A(\mathrm{mK}^2)$	β	α	ζ
extragalactic point sources	57.0	1.1	2.07	1.0
extragalactic free-free	0.014	1.0	2.10	35
Galactic synchrotron	700	2.4	2.80	4.0
Galactic free-free	0.088	3.0	2.15	35

Table 2.1 The values of the fiducial parameters used in equations 6.14 and 2.13, at $l = 1000$ and 130MHz. This table is reproduced from Santos et al. (2005).

In general, the angular power spectrum can be assumed to be isotropic in frequency and we can write $C_\ell(\Delta\nu) \equiv C_\ell(\nu_1, \nu_2)$, with $\Delta\nu = \nu_2 - \nu_1$ being the frequency separation between the two frequencies ν_1 and ν_2. As foregrounds are highly correlated the $C_\ell(\Delta\nu)$ varies slowly with $\Delta\nu$, and for a typical frequency separation of value $\Delta\nu = 2$ MHz $C_\ell(\Delta\nu)$ fall by only few percent. While the redshifted 21- cm signal is expected to be highly uncorrelated in frequency (Bharadwaj and Ali, 2005) and decline by $\sim 80\%$ for the similar value of $\Delta\nu$ (Ali and Bharadwaj, 2014; Santos et al., 2005).

Figure 2.4 (adopted from Santos et al. (2005)) shows the variation of the angular power spectrum of various foreground components at 140MHz with angular multipole ℓ, for the parameters given in tabel 2.1. The quantity plotted on y-axis is C_ℓ multiplied with $\ell(\ell+1)/(2\pi)$. The redshifted 21- cm signal power spectrum is also shown for comparison. This figure shows that the foreground power spectrum is 7-8 orders of magnitude fainter than the foregrounds. The point source component dominates over others at these angular scales in the foreground.

2.5 Foregrounds Mitigation

Several orders of magnitude larger foregrounds pose huge difficulty in detecting H I 21-cm signal from EoR. Hence, it becomes essential to remove the foregrounds from the observed data. As discussed above, the foregrounds to the redshifted 21- cm signal have a good frequency range dependence, and the foreground signal at two different but nearby frequency channels are correlated. On the other hand, the H1 21- cm signal is expected to be highly uncorrelated at these larger frequency separations (Bharadwaj and Ali, 2005). This property holds the possibility of separating the redshifted 21- cm signal from the foregrounds. Using this property, either we can subtract best-fitting continuum spectra to get the residual signal (Wang et al., 2006), which should in principle be used to determine the H1 power spectrum, or we can estimate the statistical property, say the power spectrum of the observed signal and subtract the spectrally smooth foregrounds to extract the H I signal (Zaldarriaga et al., 2004). The issue of foreground removal has been studied in the literature (Liu et al. (2009); McQuinn et al. (2006); Morales et al. (2006); Petrovic and Oh (2011), etc.) and different techniques are discussed to mitigate the effect of foregrounds. Here we discuss them briefly.

2.5.1 Foreground Avoidance

One of the methods is the foreground avoidance (Datta et al., 2010; Morales et al., 2012; Vedantham et al., 2012), which is regularly exercised in literature. It uses the property that the foregrounds to the redshifted 21-cm emissions remain correlated across a relatively larger bandwidth (Ali et al., 2008; Platania et al., 1998; Santos et al., 2005), whereas the H I signal decorrelates faster (Bharadwaj and Ali, 2005). As a result, when the power spectrum is observed as a function of $(k_{\parallel}, k_{\perp})$ the foreground emission remains concentrated in a wedge shape near the low $k_{\parallel}$, known as the "foreground-wedge", leaving the upper $k_{\parallel}$ space with low $k_{\perp}$ values free from foreground contamination. This foreground-free

space is known as the "EoR-window". The fig 2.5 (adopted from the Barry et al. (2016)) schematically shows the foreground wedge and the EoR window. The modern EoR experiments using the foreground avoidance method aim to detect the EoR power spectrum in this EoR-window. Note that the smaller frequency separation in the multi-frequency angular power spectrum contributes to larger $k_{||}$ modes of the power spectrum.

2.5.2 Foreground Subtraction

The other method which is often used in literature is the foreground subtraction method (Bowman et al., 2009; Choudhuri et al., 2017b; Morales et al., 2006). This method requires accurate knowledge of the foreground emissions. We assume that a good estimation of the foreground emission already exists, and we can directly subtract the foregrounds from the observed data to extract the 21-cm signal. It is possible to observe the point sources with good accuracy using largely repeated observations, but high dynamic range imaging is yet to be achieved for diffuse emission. This limits the usability of the foreground subtraction method to the point source foregrounds. Cho et al. (2012) propose another foreground subtraction technique where, under some assumptions, the 21- cm power spectrum can be obtained by subtracting the average foreground power spectrum.

2.5.3 Foreground supression

To enhance the accuracy of foreground removal, further foreground suppression methods are reported in literature (Arora and Dutta, 2021; Bharadwaj et al., 2019; Choudhuri et al., 2016, 2019). In this method, we taper the response of the primary beam using some window function which in turn suppresses the contribution from the bright sources lying in the outer regions of the telescope's field of view. This significantly reduces the effect of foregrounds from the residual data. Arora and Dutta (2021) discuss a lens power spectrum

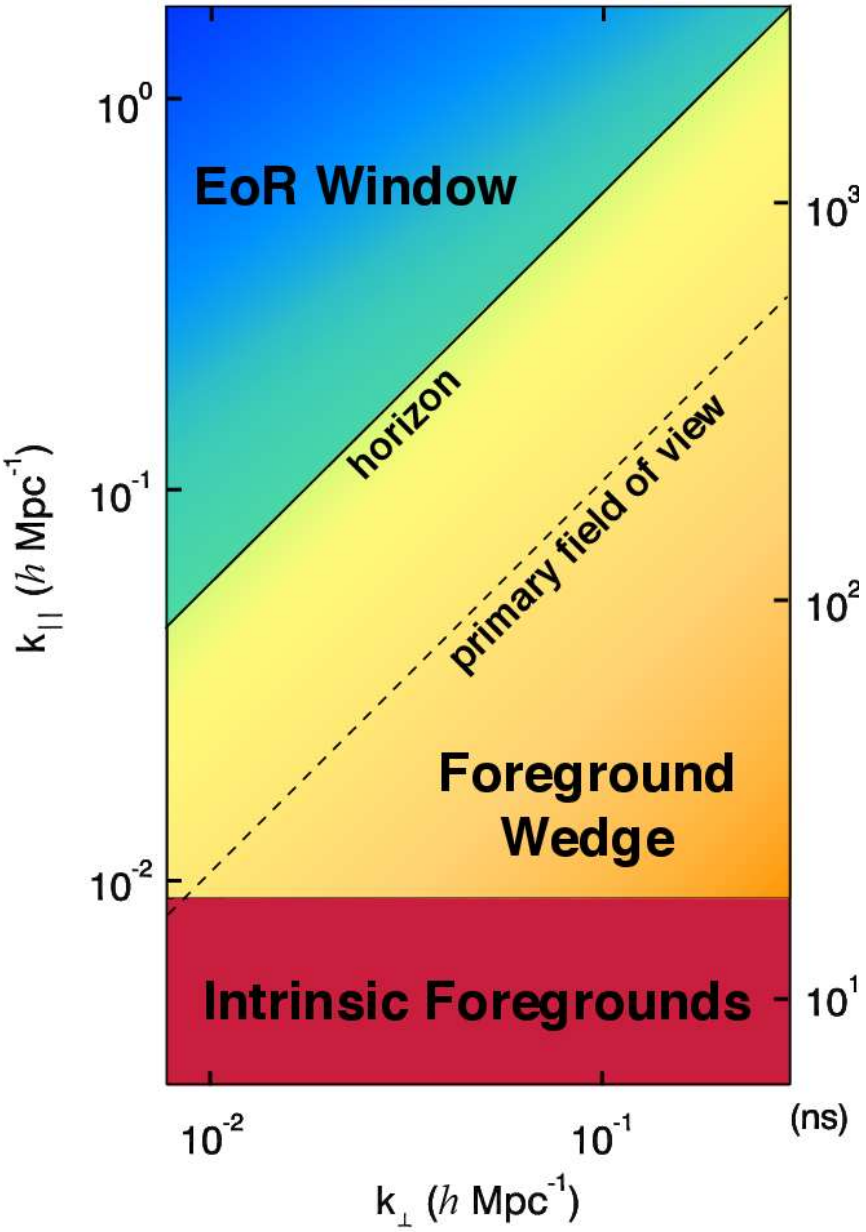

Fig. 2.5 A schematic picture of an expected 2D power spectrum showing the region of Fourier space that is accessible to the radio interferometric observations to measure the redshifted 21- cm signal. The high modes along the line of sight are limited by the spectral resolution and the low modes, for all the modes perpendicular to the line of sight, are essentially dominated by the spectrally smooth foregrounds. The chromatic response of the interferometers results in the leakage of these foregrounds to higher modes along the line of sight, leading to the foreground wedge. The primary field of view line and the horizon line are contamination limits dependent on how far off-axis sources are in the sky. The remaining region in Fourier space, which is free from the foreground contamination is termed the EoR window and is a promising region in which to pursue the first detection of the 21- cm power spectrum. The most sensitive, foreground-free measurement modes are expected to be in the lower, left-hand corner of the EoR window. The figure is reproduced from Barry et al. (2016).

estimator which can suppress the Galactic diffuse synchrotron power spectrum by a few hundred times.

The table 2.2 shows the pros and cons of the various foreground mitigation techniques discussed above. A significant amount of work is being done in the area of foreground removal, and new algorithms are being developed to mitigate the effect of foregrounds with better accuracy. Various foreground removal algorithms are investigated with the LOFAR-EoR data, such as the Fast Independent Component Analysis (FastICA), Generalized Morphological Component Analysis (GMCA), and Gaussian Process Regression (GPR) in Hothi et al. (2021).

The additional foreground removal techniques such as GPR might reduce the gain error effects discussed in this thesis to some extent. Especially GPR which accounts for the smooth component of foregrounds and the foreground component coming from the mode-mixing and models them using the Gaussian Process, but this restricts to the instrumental effects only which is coherent in frequency. Further investigations are required to quantify how accurately these techniques reduce residual gain error effects.

2.6 Discussion

In this chapter, we discuss briefly the H I power spectrum and show its relation with the dark matter density power spectrum. We present the basic measurement process of a radio interferometer and discuss how the visibilities, related to the specific intensity of the signal coming from the sky, are measured. Further presenting a discussion on " Earth Rotation Synthesis", we discuss the presence of statistical noise in visibilities and its origin. The power spectrum can be used to quantify the statistics of sky brightness distribution. We discuss in detail how visibility measurements can be used to estimate the H I power spectrum and present the visibility correlation-based power spectrum estimator. Foregrounds present a huge difficulty in the detection of redshifted H I 21- cm power

Method	Pros	Cons
Avoidance	Does not require an exquisite modeling and subtraction of the relevant foregrounds. Provides a foreground contamination free window to detect the EoR 21- cm. signal power spectrum	Limits the number of modes that can be probed, hence considerably reduces the sensitivity of the instruments. Ignoring the wedge can also introduce a bias in the recovered 21-cm signal power spectra. Also discarding the foreground cleaning in the wedge makes it harder to probe the redshift space distortion effects.
Subtraction	Aim to subtract the foregrounds completely giving full Fourier space for measurements.	Accuracy is often limited due to imperfect modeling of foreground and unmodeled sources.
Suppression	The response of bright foregrounds lying outside the primary beam is suppressed using a window function which reduces the power from the outlier sources significantly.	Because of the reduced field of view some information is lost at higher angular scales, also the cosmic variance for smaller angular modes increases. This limits the minimum ℓ at which the power spectrum can be estimated. Note that the suppression has to be used in the combination of avoidance or subtraction.

Table 2.2 Pros and Cons of the various foreground removal strategies used in experiments aiming to detect the 21- cm signal from EoR.

spectrum. We discuss various components of foregrounds and present a parametric model for the multifrequency angular power spectrum. In the end, we present various foreground removal techniques.

In all these, we assume that a complete and accurate calibration is performed, so all the gains are unity. We shall discuss the issues related to the calibration, and gains in the next chapter.

****