Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **518,** 2971–2990 (2023) https://doi.org/10.1093/mnras/stac3288
Advance Access publication 2022 November 12

Towards optimal foreground mitigation strategies for interferometric H I intensity mapping in the low-redshift Universe

Zhaoting Chen ,★ Laura Wolz and Richard Battye
Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK

Accepted 2022 November 5. Received 2022 October 19; in original form 2022 May 16

ABSTRACT

We conduct the first case study towards developing optimal foreground mitigation strategies for neutral hydrogen (H I) intensity mapping using radio interferometers at low redshifts. A pipeline for simulation, foreground mitigation, and power spectrum estimation is built, which can be used for ongoing and future surveys using *MeerKAT* and Square Kilometre Array Observatory. It simulates realistic sky signals to generate visibility data-given instrument and observation specifications, which is subsequently used to perform foreground mitigation and power spectrum estimation. A quadratic estimator formalism is developed to estimate the temperature power spectrum in visibility space. Using *MeerKAT* telescope specifications for observations in the redshift range, $z \sim 0.25$–0.30, corresponding to the MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) survey, we present a case study, where we compare different approaches of foreground mitigation. We find that component separation in visibility space provides a more accurate estimation of H I clustering when compared with foreground avoidance, with the uncertainties being 30 per cent smaller. Power spectrum estimation from image is found to be less robust with larger bias and more information loss when compared with estimation in visibility. We conclude that for the considered sub-band of $z \sim 0.25$–0.30, the MIGHTEE survey will be capable of measuring the H I power spectrum from $k \sim 0.5$ to $k \sim 10\,\mathrm{Mpc}^{-1}$ with signal-to-noise ratio being ∼3. We are the first to show that, at low redshift, component separation in visibility space suppresses foreground contamination at large line-of-sight scales, allowing measurement of H I power spectrum closer to the foreground wedge, crucial for data analysis towards future detections.

Key words: (*cosmology*:) large-scale structure of Universe – radio lines: galaxies – techniques: interferometric.

1 INTRODUCTION

Measuring the distribution of dark matter in the Universe and its evolution is one of the most important objectives of observational cosmology. The clustering of dark matter on cosmological scales, i.e. the cosmic large-scale structure, can shed light on the nature of dark energy and dark matter (see e.g. Peebles 1980). A wide range of tracers of dark matter, i.e. observables that trace the underlying density of dark matter, can be used in order to probe this distribution. These probes follow the clustering of dark matter linearly at large cosmological scales. Such probes can be galaxies, such as galaxy number count (e.g. Alam et al. 2021), weak lensing of galaxies (e.g. Pandey et al. 2022), and more. Alternatively, one can also use a relatively new technique called intensity mapping (IM; e.g. Battye, Davies & Weller 2004; Chang et al. 2008; Mao et al. 2008; Wyithe & Loeb 2009; Battye et al. 2013; Kovetz et al. 2017). It uses the emission lines of elements that are abundant in the Universe as tracers of dark matter, most promisingly neutral hydrogen (H I). H I is initially the most abundant element in the Universe as predicted by big bang nucleosynthesis (see e.g. Dodelson 2003). The emission line caused by spin–flip transition of H I has a rest wavelength of around 21 cm, and thus can be observed in the radio band with low risks of line confusion. By mapping the distribution of the flux density across the sky, statistical inference on the underlying cosmology and astrophysics of H I sources can be made (e.g. Bull et al. 2015; Chen et al. 2021).

H I IM survey with large survey volume and coarse angular resolution can be used for cosmological measurements at low redshifts, such as baryon acoustic oscillations (Eisenstein & Hu 1998) and redshift-space distortions (RSDs; Kaiser 1987). Single-dish telescopes and dish/cylinder arrays operating in single-dish mode can thus be powerful tools for H I IM, such as *Five-Hundred-Meter Aperture Spherical Radio Telescope* (Hu et al. 2020), *MeerKAT* (Santos et al. 2017), and future Square Kilometre Array Observatory (SKAO; Square Kilometre Array Cosmology Science Working Group et al. 2020).

No detection of the autocorrelation of H I using single-dish mode has been claimed but, using single-dish telescopes, such as the *Green Bank Telescope* (*GBT*), statistically significant detections have been made by cross-correlation of the IM signal from H I with optical galaxies from the WiggleZ survey (Masui et al. 2013; Switzer et al. 2013). Results from cross-correlating the 2dF Galaxy Survey and H I maps from the *Parkes* radio telescope are used to confirm the relation between the star-forming properties of galaxies and its H I mass (Anderson et al. 2018). Single-dish observations are capable

★ E-mail: zhaoting.chen@manchester.ac.uk

© 2022 The Author(s).
Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

of providing precise measurements of H I properties beyond the local Universe, as demonstrated by the results from cross-correlating eBOSS galaxies with *GBT* data (Wolz et al. 2022). Cross-correlation using the Canadian Hydrogen Intensity Mapping Experiment interferometer has been claimed by stacking the IM signal using eBOSS optical galaxy catalogues (CHIME Collaboration 2022).

Although single-dish observations are expected to be the primary method for IM at low redshifts, interferometry can play an important role as well in probing smaller scales. Accessing non-linear scales of clustering helps further extract cosmological information (e.g. Padmanabhan, Refregier & Amara 2019; d'Amico et al. 2020), and the scales typically smaller than the size of dark matter haloes can yield information about astrophysics of H I galaxies (Wolz et al. 2019; Chen et al. 2021; Schaan & White 2021). The combination of cosmological and astrophysical information in the non-linear scales of H I clustering provides strong incentives to interferometric IM.

Surveys using radio interferometers, such as the MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) survey (Jarvis et al. 2016), can be used to measure the H I power spectrum. For surveys like MIGHTEE, it is believed that H I clustering can be measured with high precision across a range of redshifts (Paul et al. 2021).

H I can also be used to probe the epoch of reionization (EoR) at $z \gtrsim 6$. During this period, structures such as the first stars and galaxies form and emit high-energy photons to ionize the H I inside the intergalactic medium. The high-redshift and low-frequency range of EoR naturally call for antenna array interferometers (Madau, Meiksin & Rees 1997), such as Precision Array for Probing the Epoch of Reionization (Parsons et al. 2014), Murchison Widefield Array (Barry et al. 2019), Low-Frequency Array (Patil et al. 2017), and Hydrogen Epoch of Reionization Array (DeBoer et al. 2017). These arrays have relatively wide field-of-view (FoV) to balance the need for deep observations and large survey volumes.

Despite probing for very different scales and underlying signals, interferometric H I IM at low redshifts and H I observations of the EoR have very similar challenges. In the frequency range of the observation, the radio sky is dominated by the diffuse foreground emissions from our galaxy and the local Universe (Di Matteo et al. 2002). The foreground radiation is smooth in frequency and can be up to several orders of magnitude higher than the H I. The H I signal, on the other hand, is discrete in frequency. Therefore, we can use the smoothness of foreground in frequency by Fourier transforming the observed visibilities along the frequency axis, commonly called the 'delay transform' (Morales & Hewitt 2004; Parsons et al. 2012a, b). It separates smooth, large frequency structure of foreground and small oscillating frequency structure of H I. The smooth foreground mainly resides on the large frequency-scale modes, creating the 'foreground wedge' (Liu, Parsons & Trott 2014a) and an observation window outside these modes. Measuring the H I power spectrum outside the foreground wedge is, therefore, sometimes called foreground avoidance.

Apart from avoiding the foreground wedge, one can also try to subtract the foreground by using its smoothness in frequency and further extract information (e.g. Bowman, Morales & Hewitt 2009). The specific approaches can be generally split into two types, the parametric approaches that use polynomial fits to extract foreground (e.g. Santos, Cooray & Knox 2005; Bonaldi & Brown 2015) and non-parametric approaches that use statistical methods to separate foreground components. The most standard approach for component separation is Principle Component Analysis (PCA). More advanced methods can be built upon it, such as Fast Independent Component Analysis (FastICA; e.g. Chapman et al. 2012), Generalized Morphological Component Analysis (e.g. Chapman et al. 2016), and Gaussian Process Regression (GPR; e.g. Mertens, Ghosh & Koopmans 2018). The success of a detection of H I signal in visibility data also relies heavily on mitigation of various systematics through the calibration of the data (Barry et al. 2016). This requires a thorough understanding of the instrument (e.g. Thyagarajan et al. 2016) and the properties of the foreground (e.g. Nasirudin et al. 2020).

To tackle the difficult problem of measuring H I clustering in the post-ionization Universe at small angular scales, it is necessary to create robust simulations for low-redshift IM using radio interferometers and the data analysis pipeline for H I power spectrum estimation. In this paper, we present an end-to-end simulation pipeline for low-redshift interferometric IM that generates realistic foreground and H I signal, simulates interferometric observations, and applies robust foreground mitigation strategies, which are crucial for future detections for *MeerKAT* and SKAO. The visibility data are used to calculate the brightness temperature power spectrum using a quadratic estimator formalism. We use the configurations of *MeerKAT* telescope and observational specifications mimicking a typical pointing of MIGHTEE survey to present a case study. A detailed comparison of the effects of different foreground mitigation methods is conducted, including a direct comparison between foreground removal in visibility data and in images. The aim of this pipeline is to provide a detailed look into the topics of IM delay power spectrum that have been extensively discussed in the context of EoR, but have not yet thoroughly quantified for observations of the low-redshift Universe. It also allows more realistic simulations that will enable us to fully study the challenges of interferometric IM towards future detection.

The paper is organized as follows: In Section 2, the basics of delay power spectrum in analytical formalism is reviewed. The quadratic estimator formalism for converting the visibility power spectrum to the brightness temperature power spectrum is discussed in Section 3. We present the simulation of the sky signal input in Section 4. Foreground mitigation in visibility space is discussed in Section 5. Power spectrum estimation using interferometric images is presented in Section 6. Comparison between different methods with MIGHTEE-like noise level is made in Section 7. We conclude our findings in Section 8. Throughout this paper, we assume the Lambda cold dark matter cosmology from Planck Collaboration VI (2020).

2 H I CLUSTERING FROM VISIBILITY

In this section, we derive the connection between the visibility data and the power spectrum of cosmological H I. Note that for simplicity, we do not consider RSDs (Kaiser 1987), which, in the scales of our interest, are dominated by effects from peculiar velocities of H I galaxies (see e.g. Chung et al. 2021). The density of H I clustering is typically expressed as the brightness temperature $T_{\rm H\,I}(\boldsymbol{x})$,

$$T_{\rm H\,I}(\boldsymbol{x}) = \frac{1}{\mathbb{V}} \sum_i C_{\rm H\,I}(z^i) M^i_{\rm H\,I} \delta^3_{\rm D}(\boldsymbol{x} - \boldsymbol{x}^i), \tag{1}$$

where $\mathbb{V}$ is the survey volume, $M^i_{\rm H\,I}$ is the H I mass of each sources within the survey volume, z^i is the redshift each source is at, $\delta^3_{\rm D}$ is the Dirac delta function in comoving space, and

$$C_{\rm H\,I}(z) = \frac{3A_{12} h_{\rm P} c^3 (1+z)^2}{32\pi m_{\rm H} k_{\rm B} \nu_{21}^2 H(z)} \tag{2}$$

is the conversion factor from H I density to brightness temperature with $h_{\rm P}$ the Planck constant, $k_{\rm B}$ the Boltzmann constant, $m_{\rm H}$ the mass of the hydrogen atom, A_{12} the emission coefficient of the 21-cm line

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

transmission, ν_{21} the rest frequency of the 21-cm emission, and $H(z)$ the Hubble parameter at redshift z (Wolz et al. 2016).

The power spectrum is measured in $\boldsymbol{k}$ space, which is the Fourier pair of comoving space coordinate $\boldsymbol{x}$. The Fourier convention used for the brightness temperature field is

$$\tilde{T}_{\rm H\,I}(\boldsymbol{k}) = \int \frac{{\rm d}^3x}{\mathbb{V}} T_{\rm H\,I}(\boldsymbol{x})\, e^{-i\boldsymbol{kx}},$$
$$T_{\rm H\,I}(\boldsymbol{x}) = \frac{\mathbb{V}}{(2\pi)^3}\int {\rm d}^3k\, \tilde{T}_{\rm H\,I}(\boldsymbol{k})\, e^{i\boldsymbol{kx}}, \tag{3}$$

and the corresponding convention for Dirac $\delta_{\rm D}$ function is:

$$\int \frac{{\rm d}^3x}{\mathbb{V}} \delta^3_{\rm D}(\boldsymbol{x}-\boldsymbol{x_0}) f(\boldsymbol{x}) = f(\boldsymbol{x_0}),$$
$$\mathbb{V}\int \frac{{\rm d}^3k}{(2\pi)^3} \tilde{\delta}^3_{\rm D}(\boldsymbol{k}-\boldsymbol{k}_0)\tilde{f}(\boldsymbol{k}) = \tilde{f}(\boldsymbol{k_0}) \tag{4}$$

for arbitrary function $f(\boldsymbol{x})$ in comoving space and $\tilde{f}(\boldsymbol{k})$ in Fourier space.

Assuming homogeneity, the two-point correlation function can be written as

$$\xi(\boldsymbol{s}) = \langle T_{\rm H\,I}(\boldsymbol{x}) T_{\rm H\,I}(\boldsymbol{x}+\boldsymbol{s})\rangle = \int \frac{{\rm d}^3x}{\mathbb{V}} T_{\rm H\,I}(\boldsymbol{x}) T_{\rm H\,I}(\boldsymbol{x}+\boldsymbol{s}). \tag{5}$$

The brightness temperature power spectrum is the Fourier transform of the two-point correlation function

$$P_{21}(\boldsymbol{k}) = \int {\rm d}^3s\, \xi(\boldsymbol{s}) e^{-i\boldsymbol{ks}} = \int {\rm d}^3s \frac{{\rm d}^3x}{\mathbb{V}} T_{\rm H\,I}(\boldsymbol{x}) T_{\rm H\,I}(\boldsymbol{x}+\boldsymbol{s}) e^{-i\boldsymbol{ks}}, \tag{6}$$

which can be similarly written in Fourier space as

$$\mathbb{V}\langle \tilde{T}(\boldsymbol{k})\tilde{T}^*(\boldsymbol{k}')\rangle = \tilde{\delta}^3_{\rm D}(\boldsymbol{k}-\boldsymbol{k}')P_{21}(\boldsymbol{k}). \tag{7}$$

The distribution of the brightness temperature can be observed using radio interferometers by measuring the radio sky through the correlations of signals across different pairs of antennas. At any given time, each pair measures the sky signal using the difference between the received signal phases determined by the position vector between the pair, i.e. the baseline. The length of the baseline in the units of the observing wavelength, $\{u, v, w\}$, corresponds to the scale of fluctuations the baseline is measuring. For a set of baselines with u–v coordinates $\{\boldsymbol{b_\alpha}\} = \{(u_\alpha, v_\alpha, w_\alpha)\}c/f$, where c is the speed of light and f is the observing frequency, the visibility $V(u, v, w, f)$ generated on each baseline is related to the sky intensity distribution $I(l, m, f)$ (Condon & Ransom 2016):

$$V(u,v,w,f) = \int \frac{{\rm d}l\,{\rm d}m}{\sqrt{1-l^2-m^2}} I(l,m,f)A(l,m,f) \times \exp\big[-2\pi i\big(lu+mv+(1-n)w\big)\big], \tag{8}$$

where l, m are the sky coordinates on the celestial sphere, $n = \sqrt{1-l^2-m^2}$, and $A(l, m, f)$ is the beam response.

For wide-FoV instruments used in EoR observation, the curved sky poses a challenge in power spectrum estimation (e.g. Thyagarajan et al. 2015b). Here, we focus on IM in the low-redshift Universe, which is typically done using small-FoV dish arrays such as *MeerKAT* and SKAO-mid. For *MeerKAT L*-band receivers which are simulated in this paper, the beamwidth is $\sim$1 deg (Asad et al. 2021). Combined with the fact that beam properties of dish telescopes are relatively well-understood and yield more desirable features of foreground contamination (Thyagarajan et al. 2015a), we expect applying simple flat-sky approximation is good enough for IM at low redshifts. Using the flat-sky approximation and Fourier transforming the visibility along the frequency axis, we have

$$\tilde{V}(u,v,\eta) \approx \int {\rm d}l\,{\rm d}m\,{\rm d}f\, I(l,m,f)A(l,m,f) \times \exp\big[-2\pi i\big(lu+mv+f\eta\big)\big], \tag{9}$$

where η is the Fourier pair of frequency f.

In the flat-sky approximation, we can write down the transformation between the sky and comoving space:

$$l = \frac{r_x}{D_{\rm c}}, m = \frac{r_y}{D_{\rm c}}, f = \frac{f_{21}}{1+z}, \tag{10}$$

where we use $r_{x,\,y}$ to denote transverse scales. Note that the comoving distance $D_{\rm c}$ is the scale along the line-of-sight direction which we also denote as r_z.

The integral can be written as

$$\tilde{V} = \int {\rm d}^3r \frac{-2k_{\rm B} f_{21} H(z)}{D_{\rm c}^2(z)(1+z)^2 c\lambda^2}[AT](\boldsymbol{r}) \exp\big[-2\pi i\big(lu+mv+f\eta\big)\big] = \frac{-2k_{\rm B}}{\lambda_{21}^3}\int {\rm d}^3r \frac{H(z)}{D_{\rm c}^2(z)}[AT](\boldsymbol{r})\exp\big[-2\pi i\big(lu+mv+f\eta\big)\big], \tag{11}$$

where we use the Jacobian determinant to perform coordinate transformation and write $[AT](\boldsymbol{r}) = A(\boldsymbol{r})T(\boldsymbol{r})$ as the product of beam response and brightness temperature.

The delay power spectrum $P_{\rm d} = \langle|\tilde{V}(u, v, \eta)|^2\rangle$ can be written as

$$P_{\rm d} = \Big(\frac{2k_{\rm B}}{\lambda_{21}^3}\Big)^2 \int {\rm d}^3r\, {\rm d}^3s \frac{H(z_r)}{D_{\rm c}^2(z_r)}\frac{H(z_{r+s})}{D_{\rm c}^2(z_{r+s})}\langle[AT](\boldsymbol{r})[AT](\boldsymbol{r}+\boldsymbol{s})\rangle \times \exp\Bigg[-2\pi i\Bigg(u\Big[\frac{r_x+s_x}{D_{\rm c}(z_{r+s})}-\frac{r_x}{D_{\rm c}(z_r)}\Big] + v\Big[\frac{r_y+s_y}{D_{\rm c}(z_{r+s})}-\frac{r_y}{D_{\rm c}(z_r)}\Big] + \eta\Big[\frac{f_{21}}{1+z_{r+s}}-\frac{f_{21}}{1+z_r}\Big]\Bigg)\Bigg]. \tag{12}$$

From the previous equation one can see that the matching between the delay power spectrum and the H I brightness temperature power spectrum is not exact. By assuming a narrow redshift range of integration, we can effectively use one central redshift z_0, so that

$$\frac{f_{21}}{1+z_{r+s}} - \frac{f_{21}}{1+z_r} \approx -\frac{f_{21}\Delta z_s}{(1+z_0)^2} \approx -\frac{f_{21}s_z H(z_0)}{(1+z_0)^2 c}. \tag{13}$$

To make the equations more compact, we denote $X = D_{\rm c}(z_0)$ and $Y = \lambda_{21}(1+z_0)^2/H(z_0)$ with $H(z_0)$ being the Hubble parameter at z_0. This results in

$$\boldsymbol{k}_\perp = \frac{2\pi\boldsymbol{u}}{X},\ k_\parallel = -\frac{2\pi\eta}{Y}, \tag{14}$$

where $\boldsymbol{k}_\perp$ is the scale on the angular plane, $\boldsymbol{u} = \{u, v\}$ is the visibility space coordinates, and $k_\parallel$ is the scale along the line-of-sight. With this notation, we can write

$$P_{\rm d} = \Big(\frac{2k_B}{\lambda^2}\Big)^2 \frac{1}{X^4Y^2}\int {\rm d}^3r\, {\rm d}^3s\, \langle[AT](\boldsymbol{r})[AT](\boldsymbol{r}+\boldsymbol{s})\rangle \times \exp\Bigg[-i\Bigg(\frac{2\pi u}{D_{\rm c}(z_0)}s_x + \frac{2\pi v}{D_{\rm c}(z_0)}s_y - \frac{2\pi f_{21}H(z_0)\eta}{(1+z_0)^2 c}s_z\Bigg)\Bigg] = \Big(\frac{2k_B}{\lambda^2}\Big)^2 \frac{\mathbb{V}^4}{X^4Y^2}\int \frac{{\rm d}^3k'}{(2\pi)^3}\frac{{\rm d}^3k''}{(2\pi)^3}\tilde{A}(\boldsymbol{k}-\boldsymbol{k}')\tilde{A}^*(\boldsymbol{k}-\boldsymbol{k}'') \times \langle\tilde{T}(\boldsymbol{k}')\tilde{T}^*(\boldsymbol{k}'')\rangle, \tag{15}$$

where $\boldsymbol{k} = \{\boldsymbol{k}_\perp, k_\parallel\}$ is the three-dimensional (3D) wave vector and $\tilde{A}$ is the Fourier transform of the beam response.

Following equation (7), we can further simplify the previous equation to

$$P_{\rm d}(u,v,\eta) = \Big(\frac{2k_B}{\lambda^2}\Big)^2 \frac{\mathbb{V}^2}{X^4Y^2}\int \frac{{\rm d}^3k'}{(2\pi)^3}|\tilde{A}(\boldsymbol{k}-\boldsymbol{k}')|^2 P_{\rm T}(\boldsymbol{k}'). \tag{16}$$

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

The previous equation not only imposes flat-sky approximation, but also requires the evolution along the light-cone to be negligible. This assumption is reasonable for the simulation used in this paper. We leave the treatment to wide frequency ranges for future work.

3 QUADRATIC ESTIMATOR OF TEMPERATURE POWER SPECTRUM

In this section, we derive the implemented power spectrum estimator, with foreground mitigation and minimization of measurement uncertainties. In order to relate the foreground-dominated observations to the brightness temperature power spectrum of H I, the sky signal measured in visibility space will be processed through a data analysis framework. Such a framework should allow estimations of the brightness temperature power spectrum of H I clustering through appropriate means of statistics to mitigate foreground. These types of power spectrum frameworks using interferometric data have been extensively studied in the context of cosmic microwave background (e.g. Tegmark 1997; Myers et al. 2003) and, more recently, EoR (e.g. Liu & Tegmark 2011; Choudhuri et al. 2014; Morales et al. 2019) and high-redshift IM (e.g. Sarkar, Bharadwaj & Marthi 2018; Chatterjee, Bharadwaj & Marthi 2021). See Liu & Shaw (2020) for a review. Here, we focus on interferometric low-redshift IM, by constructing a quadratic estimator which explicitly includes the operation of foreground mitigation and deconvolves the mode mixing introduced by the primary beam attenuation.

The visibility data consist of $N_{\rm bl} \times N_{\rm ch} \times N_{\rm steps}$ elements, where $N_{\rm bl}$ is the number of baselines, $N_{\rm ch}$ is the number of frequency channels, and $N_{\rm steps}$ is the number of time-steps in the observations. By averaging the visibility data into u–v grids, a data vector $\boldsymbol{V}$ with a length of $N_{\rm ch} \times N_{\rm u}$ can be constructed, where $N_{\rm ch}$ is the number of frequency channels and $N_{\rm u}$ is the number of grids on the u–v plane. Note that $\boldsymbol{V}$ is a column vector that loops over both frequency channel and u–v grids. Its elements can be written as

$$\mathbf{V}^j = V(u_j, v_j, f_j), \tag{17}$$

with $V(u_j, v_j, f_j)$ being the visibility data at the specific u–v coordinate and frequency the $j^{\rm th}$ gridded baseline corresponds to. Similarly, we can also define a delay-transformed data vector

$$\tilde{\mathbf{V}}^k = \tilde{V}(u_k, v_k, \eta_k), \tag{18}$$

where η is the Fourier-inverse of the observing frequencies.

The visibility power spectrum of the 21-cm emission can be discretized into the summation of the bandpower

$$P_{\rm d}(\boldsymbol{u}, \eta) = \sum_\alpha \chi_\alpha(\boldsymbol{u}, \eta) P_{\rm d}(|\boldsymbol{u}|_\alpha, \eta_\alpha) = \sum_\alpha \chi_\alpha p^{\rm d}_\alpha, \tag{19}$$

with χ^α being the selection function, returning 1 if the baseline falls into the $\alpha^{\rm th}$ bin and 0 if not. We define $p^{\rm d}_\alpha \equiv P_{\rm d}(|\boldsymbol{u}|_\alpha, \eta_\alpha)$.

Similarly the temperature bandpower is defined as

$$P_{\rm T}(\boldsymbol{k}) = \sum_\beta \chi_\beta(\boldsymbol{k}) P_{\rm T}(\boldsymbol{k}_\beta) = \sum_\beta \chi_\beta(\boldsymbol{k}) p^{\rm T}_\beta. \tag{20}$$

The estimator of the bandpower $\hat{p}^{\rm d}_\alpha$ can be constructed as

$$\hat{p}^{\rm d}_\alpha = \mathbf{V}^\dagger \mathbf{E}^{\rm d}_\alpha \mathbf{V} - \hat{b}^{\rm d}_\alpha. \tag{21}$$

Here, $\mathbf{V}$ is the gridded visibility data vector, $\hat{b}^{\rm d}_\alpha$ is the correction term for bias, and $\mathbf{E}^{\rm d}_\alpha$ is the power spectrum estimation matrix which we will derive explicitly.

When estimating $\hat{p}^{\rm d}_\alpha$, one can choose $\mathbf{V}$ to include all the visibility data. In our case, $\mathbf{E}^{\rm d}_\alpha$ is a block matrix of size $N_{\rm u} \times N_{\rm u}$, with each element being a $N_{\rm ch} \times N_{\rm ch}$ matrix.

$$\mathbf{E} = \begin{pmatrix} \mathbf{E}^1_{11} & \cdots & \mathbf{E}^1_{1N_{\rm ch}} & & & & \\ \vdots & \ddots & \vdots & \cdots & & \cdots & \\ \mathbf{E}^1_{N_{\rm ch}} & \cdots & \mathbf{E}^1_{N_{\rm ch}N_{\rm ch}} & & & & \\ & \vdots & & \ddots & & \vdots & \\ & & & & \mathbf{E}^{N_{\rm u}}_{11} & \cdots & \mathbf{E}^{N_{\rm u}}_{1N_{\rm ch}} \\ & \cdots & & \cdots & \vdots & \ddots & \vdots \\ & & & & \mathbf{E}^{N_{\rm u}}_{N_{\rm ch}} & \cdots & \mathbf{E}^{N_{\rm u}}_{N_{\rm ch}N_{\rm ch}} \end{pmatrix}$$

The resulting $\mathbf{E}^{\rm d}_\alpha$ is computationally difficult to deal with due to its large size. Note that if different u–v grids do not correlate, the off-diagonal blocks will be empty. The block elements of $\mathbf{E}^{\rm d}_\alpha$ will then only operate on the $N_{\rm ch}$ data points in the corresponding u–v grid. Thus, we assume the off-diagonal part to be negligible and only construct the matrix for one u–v grid at a time, using an input data vector of size $N_{\rm ch}$ (the number of frequency channels). The beam mixes different u–v modes, whose correlation is present in the non-block-diagonal part of $\mathbf{E}^{\rm d}_\alpha$. However, this approximation provides massive speed-up for computational efficiency. We leave the full treatment of $\mathbf{E}^{\rm d}_\alpha$ to future work. Under this assumption, the covariance matrix of the data vector can be written as (Dillon et al. 2015)

$$C_{\boldsymbol{u}\boldsymbol{u}'ff'} \approx \delta^{\rm K}_{\boldsymbol{u}\boldsymbol{u}'} \hat{C}_{ff'}(|\boldsymbol{k}_\perp|), \tag{22}$$

where $\delta^{\rm K}$ is the Kronecker delta and $\hat{C}_{ff'}(|\boldsymbol{k}_\perp|)$ is the estimated data covariance in each annulus $|\boldsymbol{u}|$ bin, which is calculated from averaging the visibility covariance across the baselines that fall into the particular $|\boldsymbol{u}|$ bin.

For the data vector of one u–v grid, the conversion from frequency to the delay time domain can be written as

$$\tilde{\mathbf{V}}^k = \delta f \sum_j e^{-2\pi i \eta_k f_j} \mathbf{V}^j = \sum_j \mathbf{F}^k{}_j \mathbf{V}^j, \tag{23}$$

where $\mathbf{F}^k{}_j$ is the discrete Fourier transform kernel and δf is the channel bandwidth.

The covariance matrix of the data vector of one u–v grid can be written as

$$\mathbf{C} \equiv \langle \mathbf{V}\mathbf{V}^\dagger \rangle = \mathbf{C}_{\rm fg} + \mathbf{N} + \sum_\alpha p^{\rm d}_\alpha \mathbf{C}_{,\alpha}, \tag{24}$$

with $\mathbf{C}_{\rm fg}$ being the covariance matrix of the radio foreground in the frequency domain, $\mathbf{N}$ the noise covariance matrix, and $\sum_\alpha p^{\rm d}_\alpha \mathbf{C}_{,\alpha}$ the signal cross-correlation[1] which we decompose into some bandpower of visibility data $p^{\rm d}_\alpha$.

The elements of the signal covariance matrix can be written as

$$\begin{aligned} \left(\mathbf{C}_{\rm s}\right)_{ij} &= \langle V_{\rm H\,I}(u, v, f_i) V^*_{\rm H\,I}(u, v, f_j) \rangle \\ &= \int {\rm d}\eta_1 {\rm d}\eta_2 \exp\left[2\pi i (f_i \eta_1 - f_j \eta_2)\right] \langle \tilde{V}_i \tilde{V}^*_j \rangle \\ &= \int {\rm d}\eta_1 {\rm d}\eta_2 \exp\left[2\pi i (f_i \eta_1 - f_j \eta_2)\right] \tilde{\delta}_{\rm D}(\eta_1 - \eta_2) P_{\rm d} \\ &= \frac{1}{(N_{\rm ch}\delta f)^2} \sum_\alpha \exp\left[2\pi i (f_i - f_j)\eta_\alpha\right] p^{\rm d}_\alpha. \end{aligned} \tag{25}$$

[1]The signal cross-correlation is a linear combination of $p^{\rm d}_\alpha$. Therefore, the coefficient of the expansion $\mathbf{C}_{,\alpha} = \partial \mathbf{C} / \partial p^{\rm d}_\alpha$.

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

Therefore, we can write

$$\left(\mathbf{C}_{,\alpha}\right)_{ij} = \frac{1}{(N_{\rm ch}\delta f)^2}\exp\left[2\pi i(f_i - f_j)\eta_\alpha\right]. \tag{26}$$

If we aim to simply calculate the delay power of the visibility data instead of the temperature power and assume no foreground and noise contamination, the power spectrum estimation can be written as

$$\hat{p}^{\rm d}_\alpha|_{\rm no\,fg,n} = \frac{\sum_i \chi_\alpha(u_i, v_i, \eta_i)|\tilde{\mathbf{V}}^i|^2}{\sum_i \chi_\alpha(u_i, v_i, \eta_i)} = \sum_i (\mathbf{w}_\alpha)_i |\tilde{\mathbf{V}}^i|^2, \tag{27}$$

where i loops over all Fourier-transformed u–v grids, χ_α is the selection function, and we express the normalized selection function as $\mathbf{w}_\alpha$ to make the expression more compact. With the previous expression we can write

$$\hat{p}^{\rm d}_\alpha|_{\rm no\,fg,n} = \mathbf{V}^\dagger\mathbf{F}^\dagger{\rm diag}(\mathbf{w}_\alpha)\mathbf{F}\mathbf{V}. \tag{28}$$

Here, ${\rm diag}(\mathbf{w}_\alpha)$ is a diagonal matrix with the $i^{\rm th}$ diagonal element being $(\mathbf{w}_\alpha)_i$.

In the presence of thermal noise and foreground contamination, we can include some linear operation $\mathbf{R}$ in the estimator in order to mitigate the foreground. This operation $\mathbf{R}$ may include foreground removal, inverse covariance weighting, frequency tapering, etc. Taking this operator into consideration we can rewrite the power spectrum estimation matrix $\mathbf{E}^{\rm d}_\alpha$ as

$$\mathbf{E}^{\rm d}_\alpha = {\rm diag}(\mathbf{S}_\alpha)\mathbf{R}^\dagger\mathbf{F}^\dagger{\rm diag}(\mathbf{w}_\alpha)\mathbf{F}\,\mathbf{R}. \tag{29}$$

Here, $\mathbf{S}_\alpha$ is a normalization vector which we can solve for by taking the expectation value of the power spectrum estimator

$$\langle\hat{p}^{\rm d}_\alpha\rangle = \sum_\beta {\rm tr}\left[\mathbf{C}_{,\beta}\mathbf{E}_\alpha\right]p^{\rm d}_\beta + {\rm tr}\left[\left(\mathbf{N} + \mathbf{C}_{\rm fg}\right)\mathbf{E}^{\rm d}_\alpha\right] - \hat{b}^{\rm d}_\alpha. \tag{30}$$

To solve for $\mathbf{S}_\alpha$, we impose

$$\sum_\beta {\rm tr}\left[\mathbf{C}_{,\beta}\mathbf{E}_\alpha\right] = 1, \tag{31}$$

and the covariance of the estimation is

$$\Sigma_{\alpha\alpha'} = \langle\hat{p}^{\rm d}_\alpha\hat{p}^{\rm d}_{\alpha'}\rangle - \langle\hat{p}^{\rm d}_\alpha\rangle\langle\hat{p}^{\rm d}_{\alpha'}\rangle = 2{\rm tr}\left[\mathbf{C}\,\mathbf{E}^{\rm d}_\alpha\mathbf{C}\,\mathbf{E}^{\rm d}_{\alpha'}\right]. \tag{32}$$

Note that here we choose $\mathbf{S}_\alpha$ as a vector, or effectively as a diagonal normalization matrix, as discussed in Hamilton (1997) and Tegmark (1997). This leads to the correlation of the variance of different bandpowers, as the previous matrix $\Sigma_{\alpha\alpha'}$ will have non-diagonal components. To decorrelate the variance, one should choose the normalization matrix to be $\mathbf{F}^{-1/2}$ where $\mathbf{F}$ is the bandpower Fisher matrix (Tegmark 1998; Hamilton & Tegmark 2000). As mentioned earlier though, we calculate the bandpower of one u–v grid at a time for computational efficiency. Thus, despite using a normalization method that results in correlation between variances of different bandpowers, the non-diagonal part of $\Sigma_{\alpha\alpha'}$ is not included in our calculation. See Dillon et al. (2014) for a discussion on overcoming real-world obstacles such as large data volume and error properties.

In the ideal case where the foreground and noise covariance is known, one should always choose $\hat{b}^{\rm d}_\alpha = {\rm tr}\left[\left(\mathbf{N} + \mathbf{C}_{\rm fg}\right)\mathbf{E}^{\rm d}_\alpha\right]$ and use inverse covariance weighting $\mathbf{R} = \mathbf{C}^{-1}$, so that we have an unbiased estimation of the power spectrum with minimum uncertainty (i.e. the optimal estimator). In reality though, we only have a guess for the true covariance of the foreground and noise, and therefore the estimator can be written as

$$\hat{p}^{\rm d}_\alpha = \mathbf{V}^\dagger\mathbf{E}^{\rm d}_\alpha\mathbf{V} - {\rm tr}\left[\left(\mathbf{N}^{\rm psd} + \mathbf{C}^{\rm psd}_{\rm fg}\right)\mathbf{E}^{\rm d}_\alpha\right], \tag{33}$$

with a pseudo-covariance as our best guess of the true foreground and noise covariance.

If no extra data processing is applied, $\mathbf{R}$ is simply the identity matrix. Frequency tapering can be applied by choosing $\mathbf{R}$ to be a diagonal matrix with the diagonal components being the frequency window (see e.g. The HERA Collaboration 2022). For any foreground mitigation strategy $\mathbf{R}$ that perfectly removes the foreground components in the data vector, the resulting estimator should revert to the optimal case. Thus, we can decompose the operation matrix into

$$\mathbf{R}_{\rm fg} = (\mathbf{C} - \mathbf{C}^{\rm psd}_{\rm fg})^{-1}\mathbf{A}. \tag{34}$$

Here, $\mathbf{A}$ is the component separation matrix that should remove the contribution of the foregrounds. Since we believe the resulting data vector after the operation of $\mathbf{A}$ is free of foreground, a subsequent inverse covariance matrix is applied. See Kern & Liu (2021) for an example of this using GPR.

Finally, after obtaining the estimation of the delay power spectrum on each u–v grid, we apply annulus bins in u–v space to $\hat{p}^{\rm d}_\alpha$. The temperature power spectrum can be estimated with the annulus-binned delay power spectrum via

$$\hat{p}^{\rm T}_\beta = \sum_\alpha \left(\mathcal{M}\right)^{-1}_{\beta\alpha}\hat{p}^{\rm d}_\alpha. \tag{35}$$

Here, $\mathcal{M}$ is a conversion matrix derived from equation (16) (see Appendix A for detailed derivation):

$$\left(\mathcal{M}\right)_{\alpha\beta} = \left(\frac{2k_{\rm B}}{\lambda^2}\right)^2\frac{N_{\rm ch}\,\delta f}{Y\sum_i \chi_\alpha(\boldsymbol{k}_i)}\sum_i\int\frac{{\rm d}^2\boldsymbol{k}_\perp}{(2\pi)^2}\chi_\alpha(\boldsymbol{k}_i) \times\left|\tilde{A}_\perp\left(\boldsymbol{k}^i_\perp - \boldsymbol{k}_\perp\right)\right|^2\chi_\beta(\boldsymbol{k}_\perp, k^i_\parallel), \tag{36}$$

where i loops over all Fourier-transformed u–v grids and $\tilde{A}_\perp$ is the Fourier-transformed beam response in the transverse plane defined in equation (A3). The variance of the estimation can also be propagated assuming each bandpower is an independent measurement of the clustering.

If we choose the same number of bins for $|\boldsymbol{u}|_\alpha$ and $|\boldsymbol{k}_\perp|_\beta$, we can calculate the previous square matrix and estimate the cylindrical temperature power spectrum $\left[\hat{P}^\gamma_{21}\right]_\beta$ via a matrix inversion. The resulting cylindrical power spectrum can be further averaged into 1D $\{k_i\}$ bins:

$$\hat{p}^{\rm 1d}_i = \frac{\sum_\beta w_i(\boldsymbol{k}^\beta_\perp, k^\beta_\parallel)\hat{p}^{\rm T}_\beta}{\sum_\beta w_i(\boldsymbol{k}^\beta_\perp, k^\beta_\parallel)}, \tag{37}$$

where w_i is a combination of the selection function and wedge criteria; w_i returns 1 if $(\boldsymbol{k}^\beta_\perp, k^\beta_\parallel)$ falls into the bin and is not avoided by the foreground wedge criteria.

4 SIMULATION OF THE RADIO SKY

In this section, we describe the simulations of the sky signal and the experimental set-up we use to generate visibility data consistent with *MeerKAT* observations. The *MeerKAT* array consists of 64-dish telescope. It observes the sky in the *L* band and the *UHF* band. In this paper, we are focusing on the *L*-band observations. The *L*-band receivers have 4096 frequency channels and observe the sky with a time resolution of 8 s. Also, they have a frequency resolution of 208.984 kHz. Following Paul et al. (2021), we choose 220 frequency channels centred at 1115.14 MHz and set the pointing centre at the Cosmic Evolution Survey (COSMOS) field (Scoville et al. 2007) at RA = 150.12 deg and Dec = 2.21 deg with an 11.2-h tracking. The

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

frequency range of the sub-band we choose covers a narrow redshift bin $z \sim 0.25$–0.3. The size of the input sky image is $3.7 \times 3.7\,\mathrm{deg}^2$, much larger than the *MeerKAT* dish FoV to ensure completeness.

4.1 Simulation of Galactic foregrounds

The biggest challenge for simulating interferometric observations is the extreme angular resolution, which is typically ∼arcsec. Note that we are only interested in the cosmological clustering and can therefore ignore the scales corresponding to the longest baselines. For $k \gtrsim 20\,\mathrm{Mpc}^{-1}$, the point source assumption of H I sources breaks down and the information within these small scales is beyond the interests of cosmology. Therefore, we simulate HEALPIX (Górski et al. 2005; Zonca et al. 2019) $\mathrm{N_{side}} = 8192$ maps, which corresponds to a pixel size of $(0.43\,\mathrm{arcmin})^2$ and clustering scales of $k \sim 40\,\mathrm{Mpc}^{-1}$ at $z \sim 0.25$–0.3.

At frequencies around 1 GHz, the dominant component of smooth Galactic foregrounds is the synchrotron radiation. The template we use to generate the signal is the 'Haslam map' at 408 MHz (Haslam et al. 1981, 1982). We use the de-sourced, de-striped version of the map described in Remazeilles et al. (2015). The pixel size of the original map is $(6.87\,\mathrm{arcmin})^2$ corresponding to HEALPIX $\mathrm{N_{side}} = 512$. To reduce the pixel size of the template, we follow the method in Remazeilles et al. (2015) and generate Gaussian random structure to fill in the small scales.

The spectral index of the emission can be extrapolated from observations of synchrotron-dominated radio sky at different wavelengths (e.g. Spinelli et al. 2021). Here, we use the 1.4 and 2.3 GHz maps from the Global Sky Model (Zheng et al. 2017), which are based on observations of Reich, Testori & Reich (2001) and Jonas, Baart & Nicolson (1998). The resulting input sky image at 1.4 GHz is presented at the top panel of Fig. 1. Note that upgrading the resolution using HEALPIX creates numerical artefacts at small scales seen in the map. As we show in Section 5.3, the synchrotron component has a trivial impact on power spectrum on small scales.

Apart from the synchrotron radiation, the free–free emission is another important component of the smooth foregrounds (e.g. Lian et al. 2020). It is believed to be well-approximated by a Gaussian distribution (e.g. Alonso, Ferreira & Santos 2014). Alternatively, one can also use existing H_α templates (e.g. Olivari et al. 2018). We use the H_α template of Finkbeiner (2003) and the conversion factor from unit Rayleigh to brightness temperature as in Dickinson, Davies & Davis (2003) assuming a constant spectral index extrapolated from 2.326 and 1.420 GHz. We upgrade the pixel size of the map from HEALPIX $\mathrm{N_{side}} = 1024$ to HEALPIX $\mathrm{N_{side}} = 8192$. The input sky image of the free–free emission at 1.4 GHz is shown in the middle panel of Fig. 1.

4.2 Simulation of extragalactic foreground

The discrete extragalactic radio sources dominate the power spectrum of the total sky signal and are the biggest source of foreground contamination on small scales. Therefore, a realistic modelling of the discrete radio sources is crucial to the accurate simulation of the foregrounds (e.g. Liu, Tegmark & Zaldarriaga 2009). Three approaches have been used: Gaussian realizations from a given angular power spectrum (e.g. Santos et al. 2005), a point source catalogue for the particular patch of the sky of interest (e.g. Paul et al. 2021), or Poisson realizations using flux count statistics (e.g. Battye et al. 2013). We simulate the sky in a stochastic fashion and choose the flux count approach.

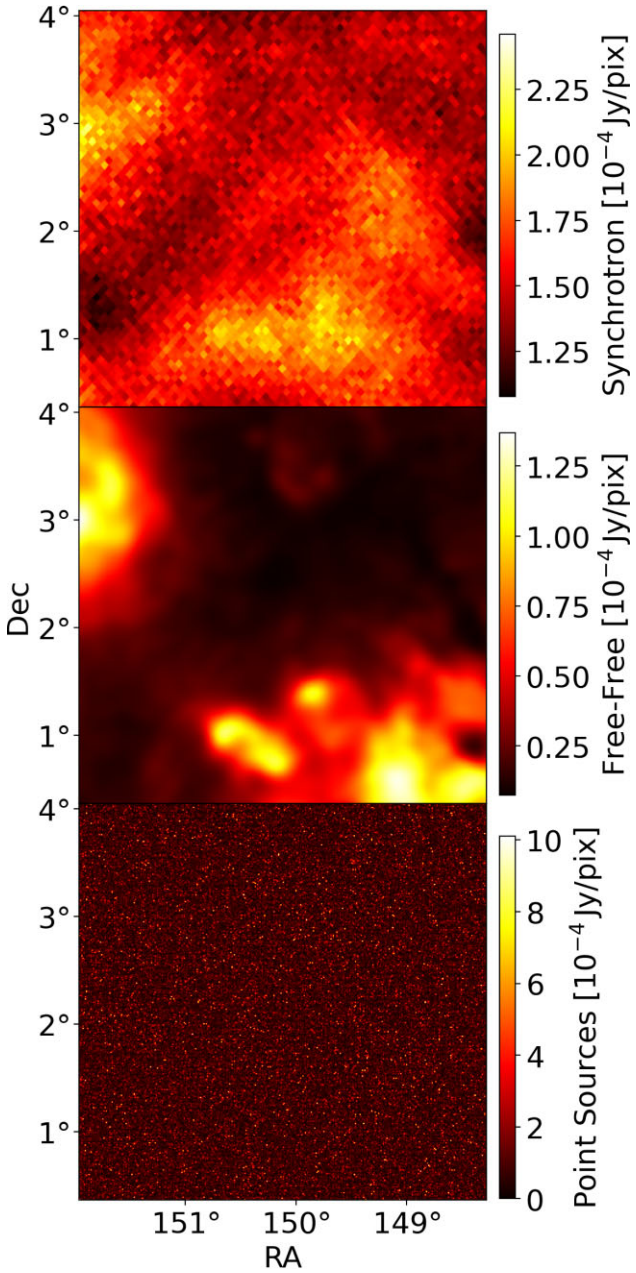

Figure 1. Top panel: The simulated synchrotron radiation around the pointing centre at RA = 150.12 deg and Dec = 2.21 deg, originally generated at 408 MHz and extrapolated to 1.4 GHz. Middle panel: The simulated free–free emission around the pointing centre at 1.4 GHz. Bottom panel: The simulated extragalactic point sources around the pointing centre at 1.4 GHz. Values above 10^{-3} Jy per pixel are set to 10^{-3} for better presentation. The pixel size of the images is set to $0.01 \times 0.01\,\mathrm{deg}^2$. The size of the input sky image is $3.7 \times 3.7\,\mathrm{deg}^2$.

Radio galaxies are the most important targets for observations in the radio band, and the source counts are well studied by many surveys, such as the NVSS survey (Condon et al. 1998), the VLA-Deep Field (Bondi et al. 2003), and more recently the *MeerKAT* DEEP2 field (Mauch et al. 2020) and the ongoing VLASS survey (Lacy et al. 2020). We follow the source count statistics at 1.4 GHz described in Matthews et al. (2021), which uses NVSS and DEEP2 to account for the bright and faint end of the distribution, respectively.

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

Two assumptions are made for the simulation of extragalactic sources. First, they are treated as point sources. Although large radio interferometers such as *MeerKAT* can resolve most of the sources observed, the arcsec angular scales are not of cosmological interest. Second, we assume that point sources above the flux density of 10^{-3} Jy can be efficiently removed or avoided. This is based on the fact that some of the fields that are being investigated, such as COSMOS and the DEEP2 field, bright radio sources are avoided or modelled as demonstrated in Paul et al. (2021). For other fields with more bright sources, we can expect the source 'peeling' to be efficient above this flux cut given the sensitivity of the MIGHTEE survey (Jarvis et al. 2016). Relaxing the flux cut can lead to the foreground power saturating all $k_{\parallel}$ scales of our interests. We leave a more careful treatment to the effects of imperfect peeling for future work.

The all-sky distribution of point sources is generated with the following steps:

(1) We grid the flux density range (10^{-7}–10^{-3} Jy) with 20 narrow, logarithmic bins. In each bin, a total number of sources is calculated based on the source count statistics. We find that using more bins has little impact on the results.
(2) For each bin, a random subsample of pixels is selected and each is assigned a source with the average flux density of the bin. Note this is different from simulations of single-dish observations due to the angular resolution of the observations. For lower resolution, in a flux bin, each pixel will have multiple sources on average. For each pixel a source number will be Poisson sampled. Here, the number of pixels will be larger than the number of sources, and therefore only a sample of pixels is uniformly selected.
(3) Following Matthews et al. (2021), for each pixel a Gaussian random spectral index with an average of −0.7 and standard deviation of 0.2 is assigned.

We present the simulated extragalactic foreground in the bottom panel of Fig. 1.

Note that the extragalactic point sources also cluster (e.g. Overzier et al. 2003; Hale et al. 2018; Siewert et al. 2020). It is straightforward to generate Gaussian fluctuations based on input angular power spectrum to account for the clustering component. It scales approximately as $w(\theta) \propto \theta^{-0.8}$ (Peebles 1974) and is therefore negligible on small scales of our interest.

4.3 Simulations of H I signal

In this subsection, we describe the simulations of H I used in this paper. The H I signal is intrinsically different to the foregrounds, since the foregrounds are smooth in frequency while the H I signal from a particular frequency corresponds to specific cosmological redshift. As a consequence of this, simulating all-sky maps of H I at different frequencies are computationally expensive and difficult. Instead, we use 3D simulations of H I in cubic boxes of cosmological volumes. While more sophisticated simulations of H I can be found (e.g. Crain et al. 2017; Villaescusa-Navarro et al. 2018), for our purpose we use halo-model-based (Cooray & Sheth 2002) log-normal simulations introduced in Wolz et al. (2019) using POWERBOX (Murray 2018). This formalism allows us to efficiently generate many realizations to test our pipeline. The simulation involves the following steps:

(1) Assuming the halo mass function of Tinker et al. (2008) and the halo bias of Tinker et al. (2010), we calculate the halo auto power spectrum using HALOMOD (Murray et al. 2021) for a fixed redshift at the centre of our redshift bin $z \sim 0.27$. All steps before light-cone construction assume this fixed redshift. The power spectrum is then used as an input to generate log-normal discrete samples of halo centres using POWERBOX in a 250^3 Mpc^3 box with 800^3 resolution, corresponding to $k_{\rm min} = 0.025\,{\rm Mpc}^{-1}$ and $k_{\rm max} = 10.0\,{\rm Mpc}^{-1}$.
(2) Each halo is then randomly assigned a halo mass based on the halo mass function of Tinker et al. (2008) using HMF (Murray, Power & Robotham 2013).
(3) Assuming the galaxy halo occupation distribution (HOD) of Zheng et al. (2005), a number of galaxies is assigned to each halo. By Bernoulli sampling with $p = \langle N^{\rm g}_{\rm cen}\rangle$, we determine whether each halo has a central galaxy. For haloes with central galaxies, the number of satellite galaxies is determined by a Poisson distribution with mean $\langle N^{\rm g}_{\rm sat}\rangle$.
(4) The positions of central galaxies are assigned to the halo centres. For satellite galaxies, we assume the positions follow the probability distribution of a Navarro-Frenk-White (NFW) profile (Navarro, Frenk & White 1996) and assign a random distance to the halo centre following that distribution. The angular positions of the galaxies with regard to the halo centres are then uniformly sampled and combined with the distance we can assign physical coordinates to the satellite galaxies.
(5) Each galaxy is assigned an H I mass following log-normal distributions. The central and satellite H I mass of each halo is calculated based on the H I HOD $\langle M^{\rm HI}_{\rm cen,sat}\rangle$ following Spinelli et al. (2020). The mean of the distribution is set to be $M_{\rm field} = \langle M^{\rm HI}_{\rm cen,sat}\rangle / \langle N^{\rm g}_{\rm cen,sat}\rangle$ and a standard deviation of $0.25 M_{\rm field}$.

The steps mentioned earlier generate a catalogue of H I galaxies with their positions and H I mass in the comoving space. To map it onto the sky to construct the light-cone, we perform the following steps:

(1) The centre of the simulation box is assumed to be the pointing centre and at the centre of the frequency range. Using that, we can assign a position vector to the pointing centre $X_{\rm cen}(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\phi)^{\rm T}$ relative to the observer, where $X_{\rm cen}$ is the comoving distance of the central redshift and (θ, ϕ) is the sky coordinate of the pointing centre.
(2) The position vector of each H I galaxy can be solved for. The modulus of each position vector is the comoving distance for each source. The comoving distances can be conversely used to assign redshifts and subsequently frequency channels to H I galaxies.
(3) The $i^{\rm th}$ galaxy is assigned a flux density following equation (B3) as discussed in Appendix B.

The resulting H I simulation can be passed to the visibility simulation described in the next subsection.

4.4 Simulation of instrument

In this subsection, we outline the simulation of visibility data using the sky input including foregrounds and H I. We use OSKAR (Mort et al. 2010) to generate visibility data. OSKAR is a C++-based simulation tool for radio interferometers, supporting GPU-accelerated computation for efficiency. It takes in a sky model, observation strategy and the telescope array specifications (including station placement and primary beam), to simulate visibility data as a measurement set file.[2]

We simulate observations by the 64-dish *MeerKAT* telescope array. We assume the beam of each dish is Gaussian with a full width at half-maximum of 57.5 arcmin at 1.5 GHz (Mauch et al. 2020).

[2]https://casadocs.readthedocs.io/en/latest/notebooks/casa-fundamentals.html

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

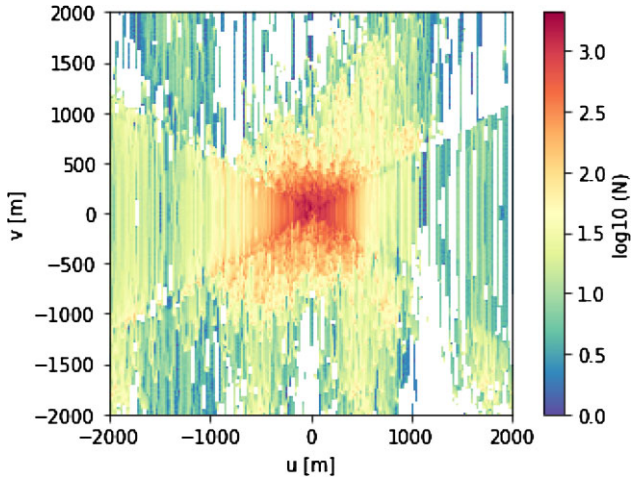

Figure 2. The u–v coverage of simulated baselines. The u–v plane is cut to only show the scales of our interest corresponding to $k \lesssim 20\,\mathrm{Mpc}^{-1}$, and gridded with a pixel size of $(20\mathrm{m})^2$ for presentation. The colour map denotes the number of baselines in each pixel.

Bright foreground sources in sidelobes can have a non-trivial effect on foreground cleaning. In interferometric observations, this can be resolved by the secondary and position-dependent calibration steps (see e.g. section 2 of Heywood et al. 2022) and we leave the treatment of beam sidelobes for future work. The pointing centre is set to be RA = 150.12 deg and Dec = 2.21 deg at the COSMOS field with an 11.2-h tracking. The *MeerKAT* telescope has a time resolution of 8 s, generating $\sim 10^7$ instantaneous baselines in one tracking. For computational efficiency, we choose the time resolution to be $\delta t =$ 40 s for the simulation and have verified that there is no visible difference when compared with a full time-resolution simulation. The simulated u–v coverage of the 11.2-h tracking is presented in Fig. 2.

We generate thermal noise per baseline following the radiometer equation (Condon & Ransom 2016)

$$\sigma_{\mathrm{N}} = \frac{2k_{\mathrm{B}}T_{\mathrm{sys}}}{A_{\mathrm{e}}\sqrt{\delta f\,\delta t}}, \tag{38}$$

where T_{sys} is the temperature of the receiver system, A_{e} is the effective aperture of the dish, $\delta f = 208.984$ kHz is the channel bandwidth, and $\delta t = 40$ s is the time resolution as mentioned earlier. We use $A_{\mathrm{e}}/T_{\mathrm{sys}} = 6.22\ \mathrm{m^2\,K^{-1}}$[3] and generate random Gaussian noise for the complex visibility.

The noise level of a single pointing for a small range of 220 channels at $z \sim 0.25$–0.3 is quite high, with the amplitude of the noise covariance comparable to the foregrounds. For the rest of the study, we consider the following two thermal noise scenarios. In order to isolate the effects of foregrounds from the thermal noise, we simulate visibility data with noise level scaled down by a factor of 40. From now on, this simulation is referred to as the 'low noise level' case. Note that we do not use the results of the low-noise level case for realistic forecasts, but to showcase the effects of different foreground mitigation strategies in Section 5.

For the second noise scenario, we aim to show the robustness of the methods for the realistic thermal noise level matching the corresponding sub-band of the MIGHTEE survey while simulating observations of one field for simplicity. The MIGHTEE survey consists of 52 pointings with a total observational time of ∼1920 h (Jarvis et al. 2016). To match the noise level of our simulation of 11.2 h to the entire survey, we first scale the thermal noise level down dividing by a factor of $\sqrt{(1920/52)/11.2}$, matching the total integration time for a single pointing. In real observations, the scaling of the thermal noise will be achieved through coherently averaging across different nights of visibility data (see e.g. Mertens et al. 2020). We then further reduce the thermal noise level dividing by a factor of $\sqrt[4]{52}$, so that the thermal noise power spectrum matches the incoherent averaging of all 52 pointings. From now on, the simulation with MIGHTEE-like noise level for $z \sim 0.25$–0.3 is referred to as the 'high-noise level' case.

[3] https://www.sarao.ac.za/science/meerkat/about-meerkat/

5 FOREGROUND MITIGATION IN VISIBILITY SPACE

In this section, we examine the effects of three different foreground mitigation strategies in visibility space on the power spectrum measurement. First, we validate our power spectrum estimator using H I-only simulations in Section 5.1. We confirm the wedge structure of the foreground power in our simulation in Section 5.2. We then apply foreground avoidance in Section 5.3. We further explore foreground subtraction by polynomial fitting the foreground covariance in Section 5.4. Component separation for the visibility data using PCA is presented in Section 5.5. As mentioned in Section 4.4, the results shown here are for the low-noise level case to isolate the effects of foregrounds from the thermal noise.

5.1 Validation of power spectrum estimator

In this subsection, we validate the power spectrum estimator described in Section 3. We simulate visibility data with only H I input, and pass the output data to the estimator. We choose to grid the u–v plane with $\Delta u = \Delta v = 10$ from 0 to 6000. Here, we apply gridding for computational efficiency, reducing the number of baselines from $N_{\mathrm{bl}} \sim 2 \times 10^6$ to $N_{\mathrm{bl}} \sim 3 \times 10^4$. The results from gridded visibility will be suboptimal (Liu et al. 2014a); however, in the case of noise covariance being dominant, the effects on the uncertainties are small.

The power spectrum is computed in cylindrical space with bandpowers further averaged into $\{|\boldsymbol{u}|\}$ bins. We choose the edges of the u–v annulus bins to be $\{|\boldsymbol{u}|_\alpha\} = [0, 100, 200, ..., 6000]$ and $\{|\boldsymbol{k}_\perp|_\beta\} = 2\pi\{|\boldsymbol{u}|_\alpha\}/X$ for the cylindrical power spectrum used in equations (35) and (37). From now on, we refer to $\{|\boldsymbol{u}|_\alpha\} = [0, 100, 200, ..., 6000]$ as the 'annulus $\{|\boldsymbol{u}|\}$ bins' used in our simulation. For the 1D power spectrum, we choose the edges of the k bins to have $\Delta k = 0.5\,\mathrm{Mpc}^{-1}$ from 0.5 to 20 Mpc^{-1}. The delay power spectrum is calculated using equation (28) and converted to the temperature power spectrum using equations (35) and (37).

We simulate 20 realizations with different dark matter halo and H I content and show the mean and the standard deviation of the power spectrum results in the top panel of Fig. 3. The blue line shows the mean of input H I boxes averaged over the realizations ('input'). The blue-shaded region represents the standard deviation of the input power spectrum. The averaged output of the power spectrum estimator is shown as the orange-dashed line. As one can see, the results from the estimator agree tightly with the input. The standard deviation of the output power spectrum is shown as the orange-shaded region. The large variance is due to the fact that the input H I box is much larger than the telescope FoV. At scales of 1-halo correlation and shot noise, the variance of the power spectrum is large for the small volume of one pointing. The number density of massive H I galaxies in the telescope FoV fluctuates from one point

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

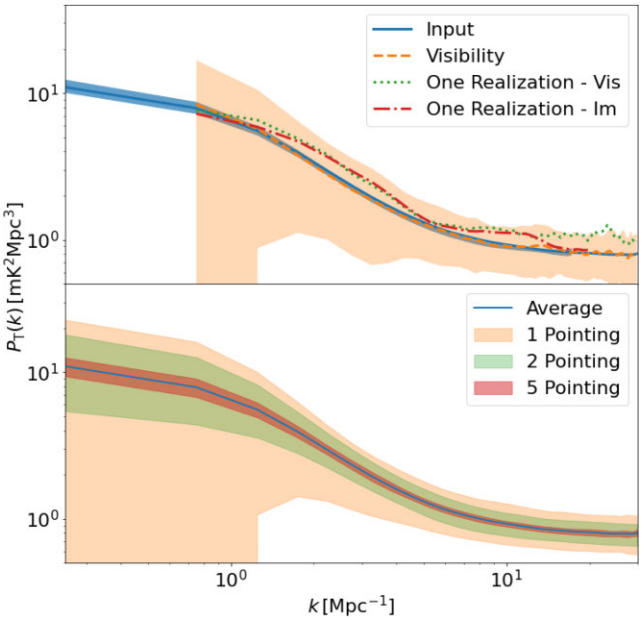

Figure 3. Top panel: The brightness temperature power spectrum of input H I signal ('Input'), compared with the results using the power spectrum estimator described in Section 3, averaged over 20 different H I-only realizations ('Visibility'). The shaded blue region shows the standard deviation of the input power spectrum. The shaded orange region shows the standard deviation of the output power spectrum. Note that for the power spectrum from simulated visibility, we cut the first $|\boldsymbol{u}|$ bin and consequently first k bin for reasons discussed in Section 5.3. Therefore, the orange region starts at larger k than the input. The green, dotted line shows the result of one specific realization using the power spectrum estimator. The red, dash–dotted line shows the input H I power spectrum from the simulated light-cone within the telescope FoV from the same realization, corresponding to the power spectrum of the input image within the FoV ('One Realization – Im'). Bottom panel: The variance of H I power spectrum due to the limited survey volume for 1, 2, and 5 pointings as shown in the shaded regions compared with the average H I power spectrum of the entire box.

to another. To illustrate this variance, we calculate the variance of the input power spectrum using bootstrapping by dividing the input H I box into sub-boxes matching the survey volume of 1, 2, and 5 pointings and present the results in the bottom panel of Fig. 3. The variance from the volume of 1 pointing agrees well with the variance of the output. The variance decreases as the survey volume goes up. For the MIGHTEE survey with ∼50 pointings, the effect of this variance will be negligible.

We have shown that the output power spectrum for one realization and one pointing will not correspond to the input simulation box due to small survey volume, but agrees very well after being averaged over multiple pointings. One may think that instead of the entire H I box, the output power spectrum is the power spectrum of the part of the box within the telescope FoV. We further clarify that the output H I power spectrum from one realization does not correspond exactly to 'the input image', i.e. the H I signal within the telescope primary-beam FoV as well. The H I sources outside the telescope FoV, though being heavily attenuated, still contribute to the visibility data, and therefore, the output from our estimator differs from 'the input image'. In the top panel of Fig. 3, we show, for one realization, the output power spectrum ('One Realization – Vis') against the power spectrum of the H I signal within the telescope FoV ('One Realization – Im').

As discussed in Section 4.4, we aim to forecast for the MIGHTEE survey using one pointing. We choose the specific H I realization shown in Fig. 3, with the foreground signal and thermal noise added, to present our results for the rest of this section.

5.2 Foreground wedge

In this subsection, we present the cylindrical delay power spectrum $P_{\rm d}(\boldsymbol{k}_\perp, k_\parallel)$ for each component of the simulation. $P_{\rm d}(\boldsymbol{k}_\perp, k_\parallel)$ contains information on the clustering of H I on both the transverse plane and along the frequency direction. Since H I is a cosmological signal, there is no preferred direction for the H I power spectrum except for the finger-of-god (FoG) effect. The foregrounds, on the other hand, are smooth in frequency. Therefore, the foreground power resides mostly in low-$k_\parallel$ modes and decreases sharply as $k_\parallel$ increases. It is important to use simulations to understand which $(\boldsymbol{k}_\perp, k_\parallel)$ modes are contaminated by the foregrounds in order to extract information on H I.

We use the simulated visibility data of different components without the thermal noise to visualize the 'foreground wedge' (Liu, Parsons & Trott 2014b). In this subsection, we calculate the delay power spectrum of different components using equation (33), with uniform weighting $\mathbf{R} = \mathbf{I}$ to calculate $\mathbf{E}^{\rm d}_\alpha$ in equation (29). We do not include frequency tapering, because as shown later the foreground contamination in our case is severe, and we find that the tapering does not have significant effects on containing the foreground wedge. We also do not perform inverse covariance weighting for this subsection, the reasons for which will be discussed in Section 5.3. The cylindrical power spectra shown throughout this paper are the outputs of equation (33), with no $k_\parallel$ filter applied as we only filter out small-$k_\parallel$ scales when converting to 1D power as in equation (37).

We present the cylindrical power spectrum for each component of the signal in Fig. 4. The top panels, from left to right, show the delay power spectra of point source, synchrotron and free–free emissions, respectively. The sum of these three components is shown in the lower left panel. The foreground power is significantly larger in low-$k_\parallel$ modes than in high-$k_\parallel$ modes. The dominant component of the foregrounds is extragalactic radio sources, especially for large $|\boldsymbol{k}_\perp| > 1\,{\rm Mpc}^{-1}$ where it is at least 2 orders of magnitude larger than synchrotron and the free–free emission. The point source power spectrum increases and leaks more into high-$k_\parallel$ modes when considering smaller angular scales, while synchrotron and the free–free emission behave in the opposite way. On short baselines corresponding to $|\boldsymbol{u}| < 100$ and $|\boldsymbol{k}_\perp| \lesssim 0.5\,{\rm Mpc}^{-1}$, the delay power spectrum of foreground is exceptionally large when compared with other $|\boldsymbol{k}_\perp|$ bins. This is for several reasons. First, at the particular pointing we are simulating which is close to the Galactic plane, the synchrotron radiation is bright at large scales; for $|\boldsymbol{k}_\perp| \lesssim 0.5\,{\rm Mpc}^{-1}$, the synchrotron radiation is at least an order of magnitude larger than other components. Second, the first u–v bin, namely $|\boldsymbol{u}|$ from 0 to 100, contains contributions from the shortest baselines $|\boldsymbol{u}| \sim 10-100$, covering an order of magnitude in angular scale. The rapidly decreasing power spectrum at these scales means the overall contribution to the first $|\boldsymbol{u}|$ bin is overestimated. Finally, the simulation of a limited patch of the sky introduces a windowing effect, leaking the power of the monopole into low-$|\boldsymbol{k}_\perp|$ modes. When we convert the delay power spectrum to the temperature power spectrum, the conversion matrix mixes angular modes that leads to further overestimation for nearby $|\boldsymbol{u}|$ bins. Therefore, we discard the first $|\boldsymbol{u}|$ bin and start from $|\boldsymbol{u}| > 100$.

The bottom-centre panel of Fig. 4 shows the delay power spectrum of H I. As expected, the H I power spectrum only depends on the

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

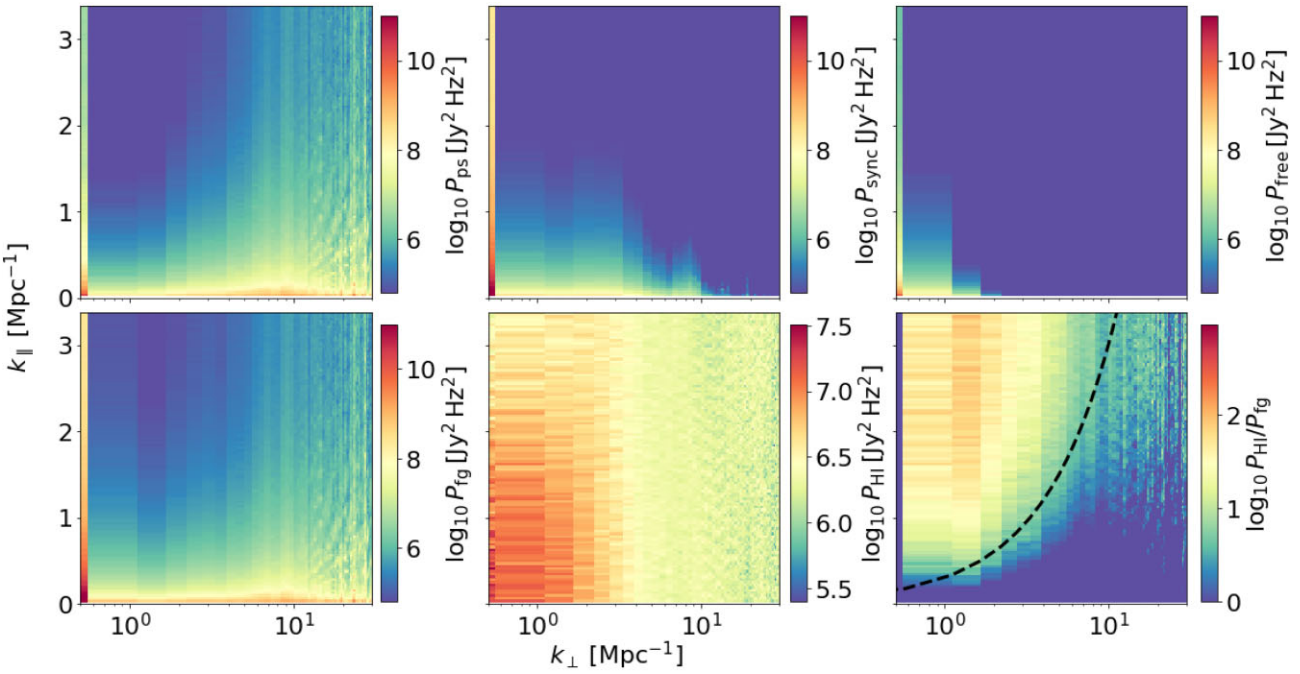

Figure 4. The cylindrical delay power spectrum for different components of the simulation. The delay time γ and the radius of annulus u–v bins $|\boldsymbol{u}|$ have been converted to k-space coordinates following equation (14). All delay power spectra are in Jy^2 Hz^2. Top-left panel: The delay power spectrum of the extragalactic point sources. Top-centre panel: The synchrotron radiation. Top-right panel: The free–free emission. Lower left panel: The total foreground emission. Lower centre panel: The H I signal. Lower right panel: The ratio of the H I power spectrum and the total foreground power spectrum. Values below 1 are masked for clearer view of the foreground wedge. The black dashed line shows a schematic illustration of the foreground wedge. Note that this line is simply for illustration and the actual criteria used to avoid foreground in cylindrical k-space are discussed in Section 5.3. The colour bars for top three panels and the bottom-left panel are kept the same for direct comparison.

1D wavenumber $k = \sqrt{|\boldsymbol{k}_\perp|^2 + k_\parallel^2}$ as we do not include RSDs. Comparing with the foreground power spectrum, H I can be orders of magnitude larger for high-$k_\parallel$ modes. We show the ratio of H I and total foreground power in the bottom-right panel. The 'wedge' where foreground contamination is most severe is at low-$k_\parallel$ scales and gets larger at longer baselines where the power spectrum of point sources increases. It leaves out a potential observation window for H I power spectrum. At a given $|\boldsymbol{k}_\perp|$, we can filter out low-$k_\parallel$ modes and only take the measurement of power spectrum at higher $k_\parallel$. This method of excluding foreground contaminated modes is called foreground avoidance, which we further explore in the next subsection.

5.3 Foreground avoidance

In this subsection, we include the contributions from foregrounds and thermal noise and use a foreground avoidance strategy to measure the H I power spectrum from our simulation. The estimator described in Section 3 applies foreground avoidance in equation (37). Namely, it includes $w_i(\boldsymbol{k}_\perp, k_\parallel)$ to encode the wedge criteria (defined later), which has the impact that when $k_\parallel$ is too small, w_i is zero and the foreground contaminated modes are filtered out. We can then calculate the resulting 1D H I power spectrum.

To estimate the temperature power spectrum from visibility data, we follow equation (33). Since we do not analytically model the foreground covariance and simply avoid contaminated modes, $\mathbf{N}^{\rm psd} + \mathbf{C}_{\rm fg}^{\rm psd}$ in equation (33) is the noise covariance for a single u–v grid point p in our simulation

$$\mathbf{N}_{ij}^{p} = \delta_{ij}\sigma_{\rm N}^2/N_i^p, \tag{39}$$

where N_i is the number of baselines for the $i^{\rm th}$ frequency channel in the $p^{\rm th}$ u–v grid. Throughout this paper, the data covariance $\mathbf{C}$ is taken to be the empirical data covariance from visibility data in each annulus $|\boldsymbol{u}|$ bin used to estimate the uncertainty following equation (32). The empirical data covariance $\mathbf{C}$ is calculated from the visibility data following equation (22) as discussed in Section 3. Note that here $\hat{C}_{ff'}(|\boldsymbol{k}_\perp|)$ is calculated from the visibility data before gridding. We then subtract the noise variance $\sigma_{\rm N}^2$ and add the noise covariance of the gridded data following equation (39). Throughout this paper, $\sigma_{\rm N}^2$ is treated as a known quantity and, therefore, the noise covariance is also known.

Real observations are likely to be noise dominated with the exact position of the foreground wedge also unknown. Assuming that we are working blind, we start with the generic criteria to be used in equation (37). The selection function $w_i(\boldsymbol{k}_\perp, k_\parallel)$ returns 1 only if $(\boldsymbol{k}_\perp, k_\parallel)$ falls into the 1D k bin and satisfies

$$k_\parallel \; > \; c_k\, k_\perp, \tag{40}$$

where c_k is a coefficient describing the position of the wedge.

To determine the value of c_k, we start with $c_k = 0$ and increase c_k iteratively with a step size of 0.05. We find that for $c_k = 0.2$ and $c_k = 0.25$, the difference in power spectrum estimation is within 5 per cent. The convergence suggests that the foreground power is largely avoided. The H I power spectrum estimated using foreground avoidance with $c_k = 0.25$ ('Avoidance Uniform') is shown in Fig. 5. Note that the standard 'horizon criteria' (see equation 13 of Liu et al. 2014a), assuming a maximum angular extent $\sin\theta_0 = 1$, corresponds to $c_k \sim 0.232$, suggesting that the foreground contamination is likely to be the dominant constraint for MIGHTEE observations.

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

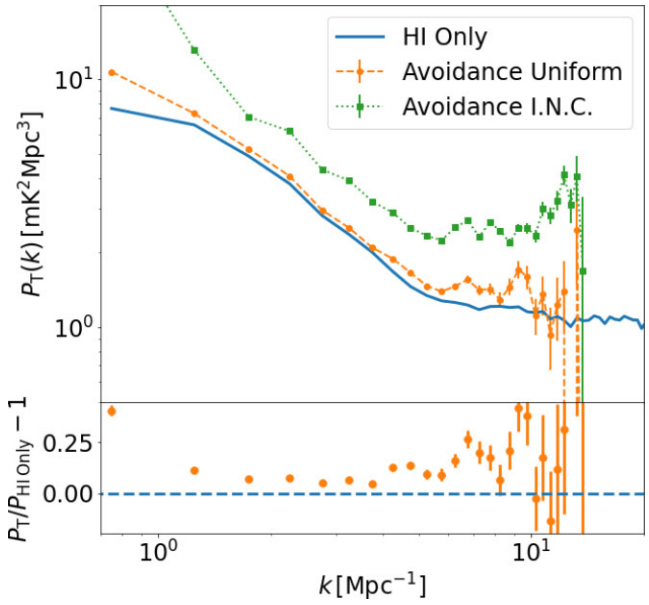

Figure 5. Top panel: The 1D brightness temperature power spectrum from H I-only visibility data ('HI only'), low thermal noise simulation using foreground avoidance method described in Section 5.3 ('Avoidance Uniform'), and the same avoidance method but with inverse noise covariance weighting ('Avoidance I.N.C.'). Bottom panel: The fractional difference between the 'HI only' and the 'Avoidance Uniform' results.

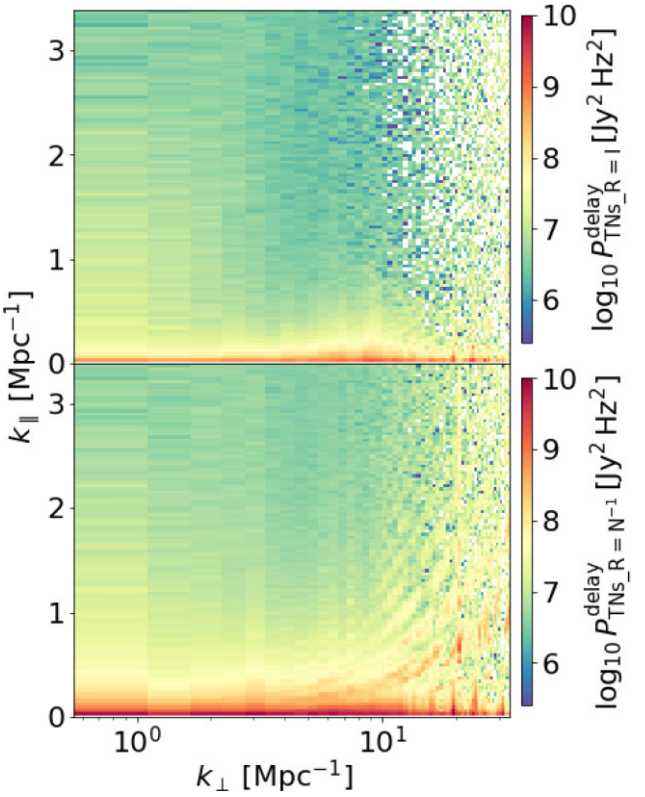

Figure 6. Top panel: The cylindrical delay power spectrum for the uniform weighting case, with noise covariance subtracted for the low-noise level case ($P^{delay}_{TNs_R=I}$). Bottom panel: The cylindrical delay power spectrum for the inverse noise covariance weighting case, with noise covariance subtracted for the low-noise level case as described in Section 5.3 ($P^{delay}_{TNs_R=N^{-1}}$). The empty (white) regions at high $|\boldsymbol{k}_\perp|$ in the figure are for negative power due to thermal noise covariance subtraction. The red and orange regions ($P^{delay}_{TNs} > 10^9$ Jy^2 Hz^2) in the bottom panel are much larger than ones in the upper panel, suggesting much more severe foreground leakage into higher $k_\parallel$ modes.

As seen by the fact that the estimated power is always higher when compared with the input, the foreground power contamination is present for all k modes and results in an overestimation of the power spectrum by around 10 per cent. For wavenumbers $k \sim 1\,\mathrm{Mpc}^{-1}$, the contamination is more severe because at larger angular scales synchrotron and free–free emissions have a large effect on the delay power spectrum as shown in Fig. 4. The contamination is also more severe for $k > 5\,\mathrm{Mpc}^{-1}$ modes; since at higher $k_\perp$, the power spectrum of point sources is much larger and leaks more into the window.

The noise level of a u–v grid point varies across frequencies due to the change in observing wavelengths and therefore applying inverse covariance weighting gives uneven weights across frequency channels, mixing different $k_\parallel$ modes in the estimator. The mixture of different $k_\parallel$ modes can lead to a spillover of the foreground power into the observational window (see also Cheng et al. 2018). To illustrate this, in Fig. 5, we also show the power spectrum estimation made by choosing inverse noise covariance weighting $\mathbf{R} = \mathbf{N}^{-1}$ ('Avoidance I.N.C.'), keeping everything else the same as the uniform weighting $\mathbf{R} = \mathbf{I}$ case.

Note that here we do not choose the inverse of the total data covariance, since in this low-thermal noise case foregrounds contribute a substantial fraction of the total covariance. In realistic observations, on the other hand, the noise covariance is expected to dominate. Therefore, we choose $\mathbf{R}$ in equation (29) to be the inverse noise covariance to illustrate the mode mixing. As one can see from Fig. 5, when applying $\mathbf{R} = \mathbf{N}^{-1}$, the power spectrum is significantly overestimated. Our claim that the foreground leakage into higher $k_\parallel$ modes is responsible can be verified by investigating the cylindrical delay power spectrum, as we show in Fig. 6. For $\mathbf{R} = \mathbf{I}$ case in the top panel, the foreground wedge is most visible at $k_\parallel \lesssim 0.3\,\mathrm{Mpc}^{-1}$ as the red and orange regions, where it is 2–4 orders of magnitude larger than the H I signal. For $\mathbf{R} = \mathbf{N}^{-1}$, on the other hand, the red and orange regions can be seen at $k_\parallel \sim 0.5\,\mathrm{Mpc}^{-1}$. In short, if foreground contamination is severe, additional weighting further mixes different $k_\parallel$ modes, further contaminating high-$k_\parallel$ modes which leads to the overestimation in the H I power spectrum.

5.4 Subtracting foreground using fitted foreground covariance

In this subsection, we explore polynomial fitting of the foreground covariance in visibility space, utilizing the fact that the foreground emission is smooth in frequency. The polynomial fitting of the covariance is similar to polynomial fitting of the signal along the line-of-sight direction in the image domain (e.g. Bowman et al. 2009) or in u–v space (e.g. Harker et al. 2010). The main difference is that we perform the fitting on the empirical data covariance instead of directly on the visibility data. If the subtraction is perfect, i.e. the resulting fit is the foreground covariance, there should be no increase in the measurement errors.

We split the visibility data into the same $\{|\boldsymbol{u}|_\alpha\}$ bins for the power spectrum estimation defined in Section 5.1 and calculate the empirical data covariance using equation (22). Note that for most $|\boldsymbol{u}|$ bins, the dominant foreground component is the point sources which are believed to be approximately Poisson distributed and assumed to be so in our simulation. In the limit of a narrow frequency range, the covariance matrix of the Poisson foreground should be real-valued

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

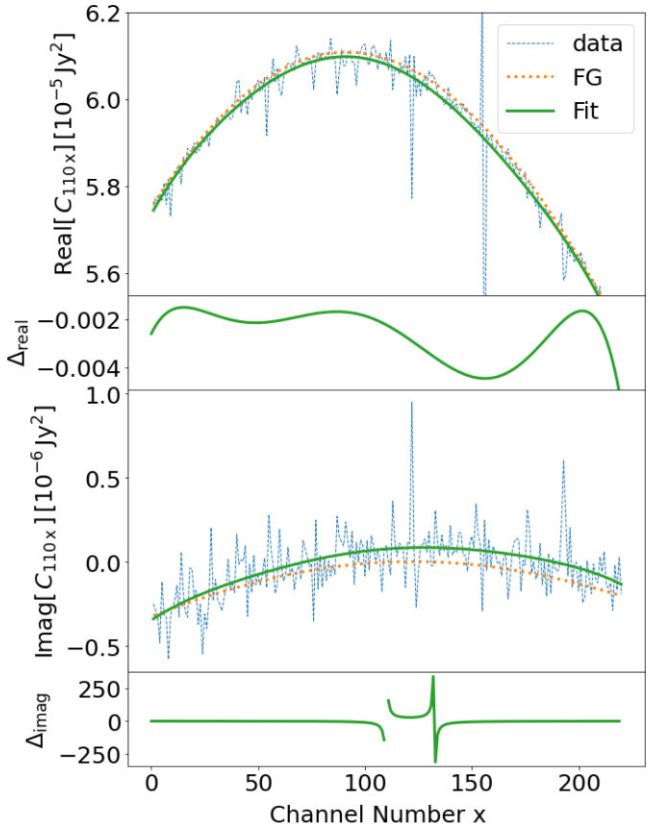

Figure 7. Top panel: The real part of the 110$^{\rm th}$ row (the centre frequency row) of the empirical data covariance matrix in the 6$^{\rm th}$ $|\boldsymbol{u}|$ bin corresponding to $k_\perp \sim 3.5$ Mpc^{-1} ('data'). The same row in the same bin of the covariance matrix for foreground component-only visibility is shown for comparison ('FG'). The extracted smooth component using 6$^{\rm th}$ order polynomial fits is also shown ('Fit'). Below it is the fractional difference between the real part of the fitted covariance and the actual foreground covariance on the 110$^{\rm th}$ row in the 6$^{\rm th}$ $|\boldsymbol{u}|$ bin. Bottom panel: The same with the top panel, except for the imaginary part of the covariance matrix.

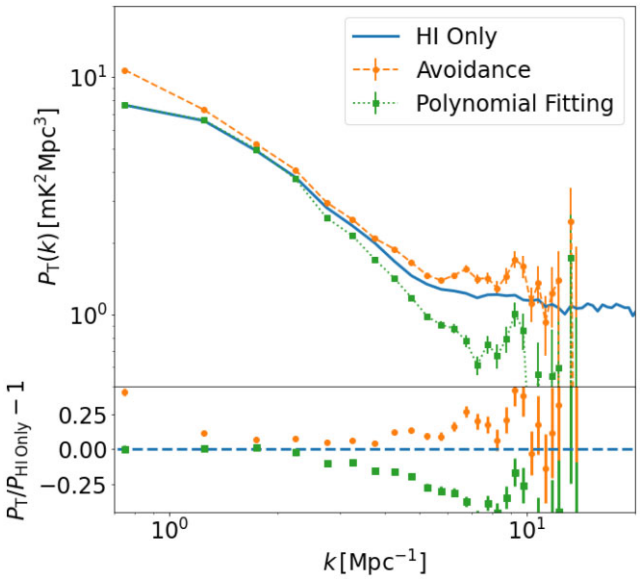

Figure 8. The 1D brightness temperature power spectrum from the H I-only visibility data ('HI Only'), the low-thermal noise simulation using foreground avoidance method ('Avoidance'), and the foreground subtraction method using polynomial fitting of the empirical data covariance ('Polynomial Fitting'). Shown also is the fractional difference among the 'HI Only', 'Avoidance', and 'Polynomial Fitting' results. While the avoidance undersubtracts the foregrounds, the polynomial fitting oversubtracts on small scales.

(see derivations in section 4 of Murray, Trott & Jordan 2017). We verify that this is indeed the case for our visibility data, where the real part of the each element in the covariance matrix is at least an order of magnitude larger than the imaginary part.

The resulting data covariance can then be processed to extract the foreground by applying a $k_\parallel$ filter to remove the rapidly oscillating H I component. We do this by applying polynomial fitting to each row of the covariance matrix, aiming to extract the smoothly varying part. We find that 6$^{\rm th}$ order fitting respectively for real and imaginary part of the empirical covariance matrix, is sufficient, as showcased in Fig. 7.

We take the fitting results as $\mathbf{C}_{\rm fg}^{\rm psd}$ and use equation (39) to calculate the noise covariance to subtract noise and foreground covariances via equation (33). We choose uniform weighting $\mathbf{R} = \mathbf{I}$ with the same $c_k = 0.25$ as in Section 5.3 and present our result in Fig. 8. As shown in the figure, when compared with the direct avoidance, subtracting the fitted foreground covariance corrects the amplitude at large $k \lesssim 1\,{\rm Mpc}^{-1}$ scales. However, it overcleans on $k \gtrsim 2\,{\rm Mpc}^{-1}$, which results in a $>$ 10-per cent signal loss. The overcleaning is due to the fact that despite in most cases the fitted covariance matches the foreground covariance up to one-per cent level as shown in the top panel of Fig. 7, it still wrongly subtracts H I features. This can be seen in the bottom panel of Fig. 7, where for the imaginary part of the covariance, foreground is relatively small. Comparing the fitted covariance with the actual foreground covariance, we can see that some additional structure over a large frequency range is mistaken to be a contribution from the foregrounds. This results in overcleaning, typically in relatively low-$k_\parallel$ modes.

We also show the cylindrical delay power spectrum to further verify this in Fig. 9. Comparing with direct avoidance in the upper panel, the subtraction overcleans signal at lower $k_\parallel$, which results in lower and even negative power as shown in the bottom panel of Fig. 9.

5.5 Foreground removal with PCA

In this section, we examine component separation in visibility data to mitigate foreground contamination. From Section 5.3, we see that to exclude the foreground power leakage into the observation window, a foreground removal method is required. We found that direct subtraction is likely to result in signal loss on small scales despite correcting the amplitude of the power spectrum well at large scales. It highlights the need for component separation techniques.

Here, we provide a case study of the most standard technique of PCA. PCA is proven to be very robust and works similarly well comparing with methods such as FastICA (see e.g. Cunnington et al. 2021). For sky maps/images, maps in different frequency channels are mean-centred and then an empirical frequency–frequency covariance matrix can be constructed for data analysis (e.g. Bigot-Sazy et al. 2015). For visibility data on the other hand, each baseline corresponds to a different Fourier mode and cannot be processed in the same way. As discussed in Section 5.3, we follow Dillon et al. (2015) to calculate empirical covariance using equation (22) in annulus $|\boldsymbol{u}|$ bins and perform PCA in each bin independently.

The first 20 eigenvalues of the empirical covariance matrices in four different bins are presented in Fig. 10. The largest component is

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

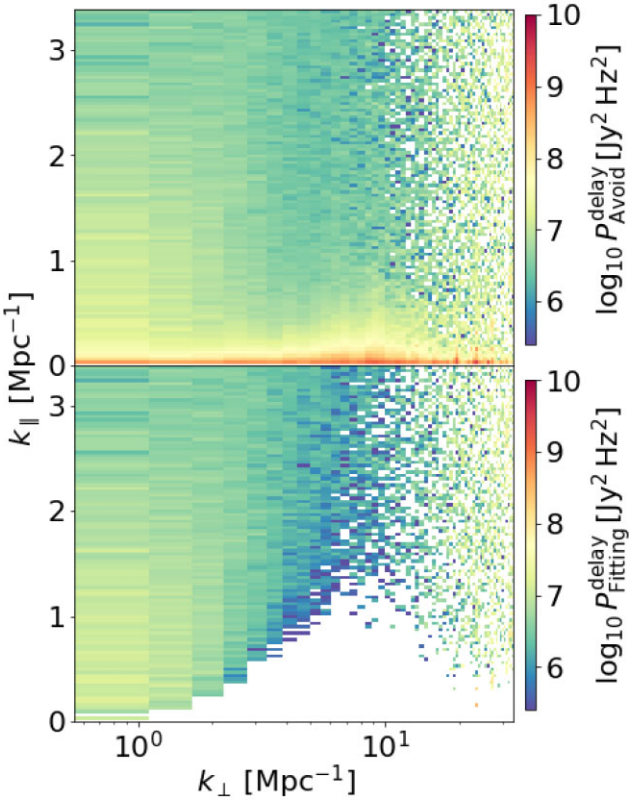

Figure 9. Top panel: The cylindrical delay power spectrum with noise covariance subtracted using foreground avoidance method ($P^{\rm delay}_{\rm Avoid}$). Bottom panel: The cylindrical delay power spectrum with noise covariance subtracted using polynomial fitting ($P^{\rm delay}_{\rm Fitting}$). The empty (white) regions indicate overcleaning, which results in negative power.

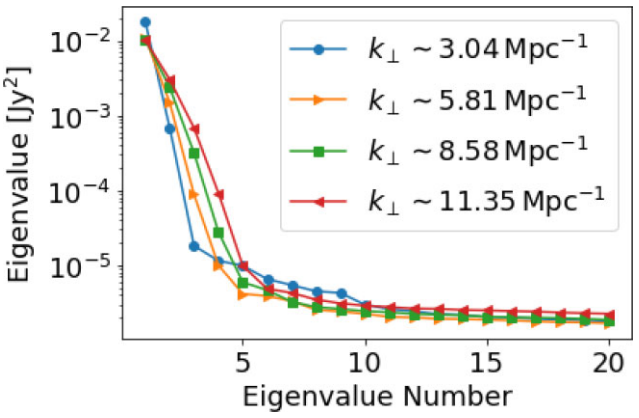

Figure 10. The first 20 eigenvalues, ranked from the biggest to the smallest, of the empirical data covariance matrices for the $k_\perp \sim 3.04$, $k_\perp \sim 5.81$, $k_\perp \sim 8.58$, and $k_\perp \sim 11.35\,{\rm Mpc}^{-1}$ bins. Note the significant drop-off in the first few bins and then the plateau beyond the fifth eigenvalue.

typically an order of magnitude higher than the second eigenvalue, and the eigenvalues hit the plateau at about the fifth eigenvalue, suggesting there is a mixture of H I and foregrounds in these modes. This is also supported by the fact that for higher $k_\perp$ the eigenvalues hit the plateau later, due to the more severe foreground contamination at higher $k_\perp$ we saw in Fig. 4. We find that choosing $N_{\rm fg} = 5$ removes foreground at small $k_\parallel$ modes but overcleans the H I over a wide range

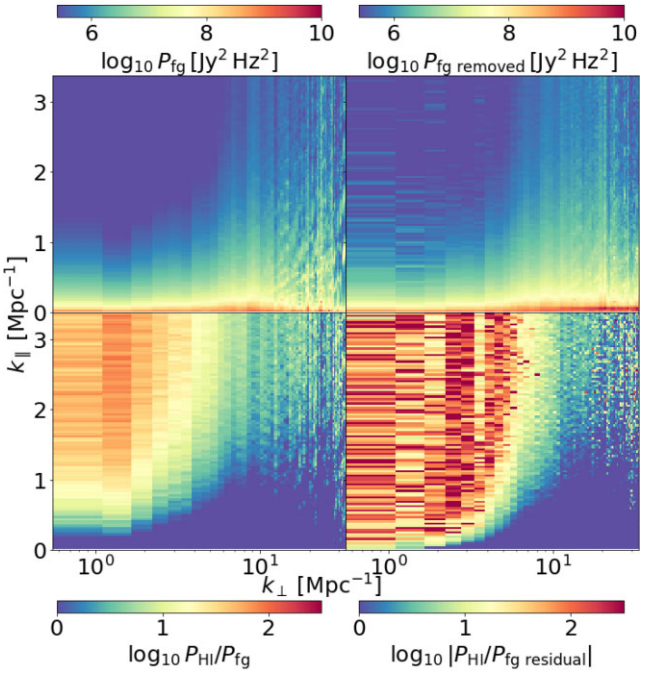

Figure 11. Top-left panel: The cylindrical delay power spectrum of total foreground in Jy^2 Hz^2. Top-right panel: The cylindrical delay power spectrum of the removed component by PCA for low thermal noise simulation. The colour bar is kept the same to the upper left panel for direct comparison. Bottom-left panel: The ratio of the cylindrical delay power spectrum of the H I component of visibility data and the power spectrum of the foregrounds. Bottom-right panel: The absolute value of ratio of the cylindrical delay power spectrum of the H I component of visibility data and the power spectrum of the residual of the foregrounds. Values below 1 are set to be 1 for better presentation in bottom-left and bottom-right panels.

of scales, something that we would like to avoid. We instead find $N_{\rm fg} = 2$ suits our purpose best. It still leaves a supressed foreground wedge, which we can avoid by applying a 'loose wedge criterion'.

PCA returns the source mixing matrix $\hat{\mathbf{A}}$, which extracts the foreground components. Note that following the framework of Section 3, we should include $\hat{\mathbf{A}}$ in $\mathbf{R}$ to construct our estimator. However, the component separation here is performed on the full visibility data to ensure sufficient cleaning; whereas, for power spectrum, estimation is applied to the gridded visibilities for computational efficiency. Therefore, we do not include $\hat{\mathbf{A}}$ in $\mathbf{R}$. Here, the foreground is subtracted before gridding and a new gridded data vector $\mathbf{V}'$ with foreground components removed is used, with the assumption that the operations of gridding and foreground subtraction are commutable.

We present the cylindrical delay power spectrum with uniform weighting $\mathbf{R} = \mathbf{I}$ in Fig. 11. For the $N_{\rm fg} = 2$ case we can see that the reconstructed foreground matches the clustering of the actual foreground quite well by comparing the upper panels of Fig. 11. PCA overcleans the foregrounds at low-$k_\perp$, high-$k_\parallel$ modes as seen by comparing the top-left and top-right panels of Fig. 11. Slight overcleaning can be acceptable, as long as the residual does not leak too much negative power into the observation window. The amplitude of the negative residual power is orders of magnitude smaller than H I, as shown in the lower right panel of Fig. 11. Comparing the bottom two panels, we see that PCA widens the observation window at lower $k_\perp$ by efficiently cleaning the foreground power, while not having visible improvements on higher $k_\perp$ modes. Comparing with foreground avoidance, a loose selection criterion can be applied for

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

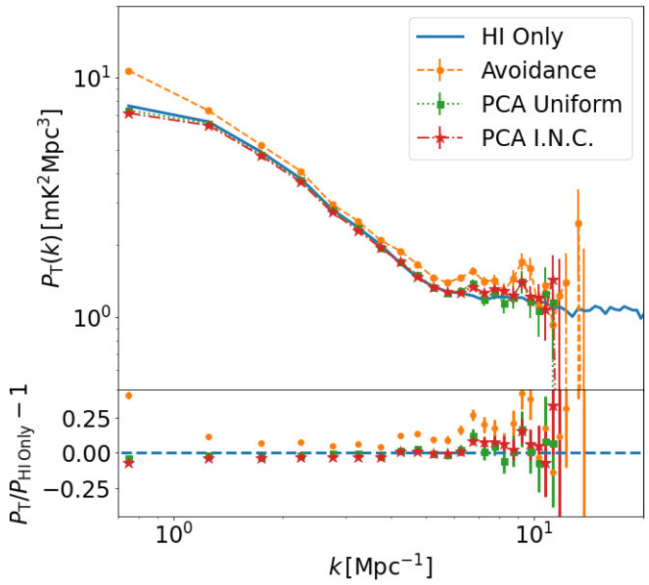

Figure 12. The 1D brightness temperature power spectrum from the H I-only visibility data ('HI Only'), simulation with low-thermal noise level using foreground avoidance method described in Section 5.3 ('Avoidance'), simulation with low-thermal noise level using PCA foreground removal with uniform weighting ('PCA Uniform'), and with inverse noise covariance weighting ('PCA I.N.C.'). In addition, we have provided the fractional difference between the estimated 1D brightness temperature power spectrum and the H I-only simulation for low-thermal noise case.

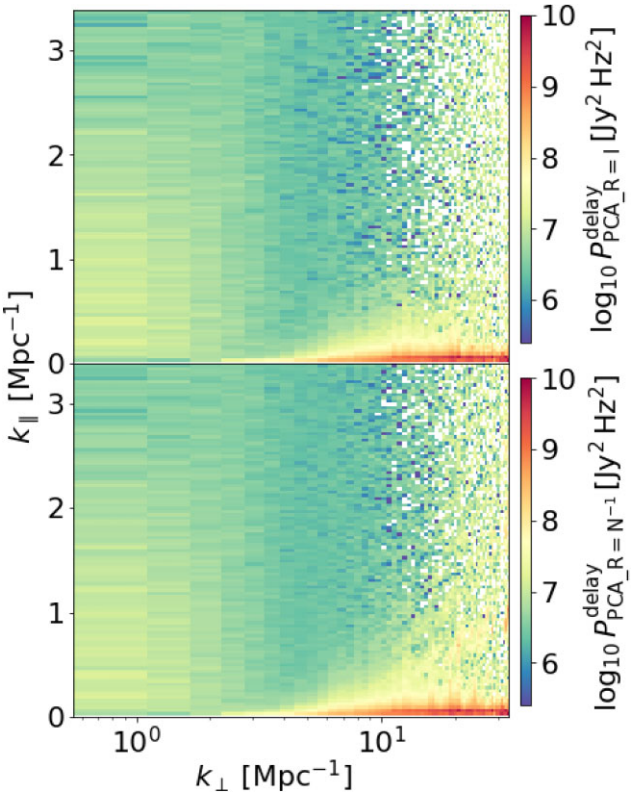

Figure 13. Top panel: The cylindrical delay power spectrum with noise covariance subtracted using PCA and uniform weighting described in Section 5.5 for the low-thermal noise simulation ($P^{delay}_{PCA_R=I}$). Bottom panel: The same with the top panel but with noise covariance subtracted using PCA and inverse noise covariance weighting for the low-thermal noise simulation ($P^{delay}_{PCA_R=N^{-1}}$).

small $k_\perp$. We find that the power spectrum converges on all scales when $c_k = 0.04$ for $k_\perp < 5\,\mathrm{Mpc}^{-1}$ and $c_k = 0.3$ for $k_\perp > 5\,\mathrm{Mpc}^{-1}$.

We show the results for the 1D temperature power spectrum in Fig. 12. Foreground removal using PCA accounts for the leakage of the foreground power into the observation window and returns a result which agrees relatively tightly with the H I-only case. Note that for $k > 5\,\mathrm{Mpc}^{-1}$, simple avoidance results in a 10-per cent level overestimation due to foreground contamination; however, for PCA the power spectrum estimation is relatively accurate up to $k \sim 10\,\mathrm{Mpc}^{-1}$. Moreover, as discussed in Section 5.3, the foreground contamination prevents us from using the inverse covariance weighting as it will further mixes different $k_\parallel$ modes resulting in more contamination. For the PCA case where the foregrounds are sufficiently removed, we can re-apply inverse covariance weighting. We show the results with $\mathbf{R} = \mathbf{N}^{-1}$ in Fig. 12 as well. The power spectrum for PCA method with inverse covariance weighting matches the H I-only case and produces smaller uncertainties. We also show the cylindrical delay power spectrum in Fig. 13 to verify there is no visible foreground leakage into higher $k_\parallel$. Comparing the top panel for uniform weighting and the bottom panel for inverse noise covariance weighting, the mode mixing is most visible at higher $k_\perp$ outside the observation window, whereas the difference at lower $k_\perp$ is negligible.

6 FOREGROUND MITIGATION IN IMAGE SPACE

In this section, we temporarily depart from visibility-only approach and investigate foreground cleaning by applying PCA directly on the image and perform power spectrum estimation in image space similar to Hothi et al. (2021) in order to compare with methods in visibility space. We use CASA *tclean* (Högbom 1974; McMullin et al. 2007) to produce the images. The pixel size of the images is set to (25.37 arcsec)^2, corresponding to $k \sim 45\,\mathrm{Mpc}^{-1}$ with 192×192 standard grids to match the size of the primary beam. Only dirty images are generated, since we find that iterative cleaning takes out part of the H I signal.

The dirty images are in the units of Jy per point spread function (PSF). We calculate the area of PSF around its centre to convert intensity to temperature unit. Starting with the centre and neighbouring pixels, we iteratively expand the integration area until the PSF area decreases, suggesting that the effects of sidelobe structures are starting to dominate. We then rescale the images from Jy per PSF to Jy per pixel using the PSF area calculated. The primary beam effect is then removed by dividing by the beam attenuation term. The processed dirty image is then converted to temperature unit and used to calculate the temperature power spectrum. We do not deconvolve the shape of the PSF, which leaves a scale-dependent attenuation effect in the power spectrum. Therefore, when investigating the effects of PCA, we compare the power spectrum results with the dirty images of H I and thermal noise simulation. We leave the full treatment of the PSF for future work.

We aim to present the imaging approach as a qualitative comparison to the visibility approach and do not derive in detail how to subtract thermal noise covariance and estimate uncertainties. We simulate a thermal noise-only image and use it to directly subtract the thermal noise power spectrum. The uncertainties are estimated using the sampling variance in k bins.

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

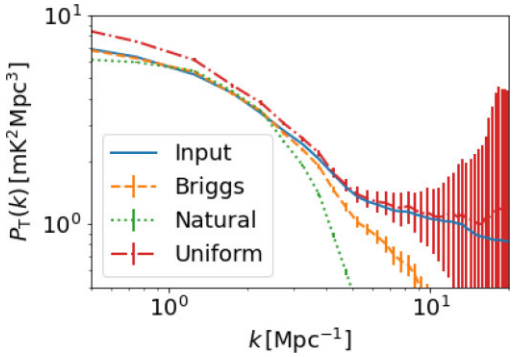

Figure 14. The power spectrum of input H I corresponding to the image size used in Section 6 ('Input'), compared with the power spectrum of images using H I visibilities produced by natural weighting ('Natural'), uniform weighting ('Uniform'), and Briggs weighting with robustness parameter equal to 0.5 ('Briggs').

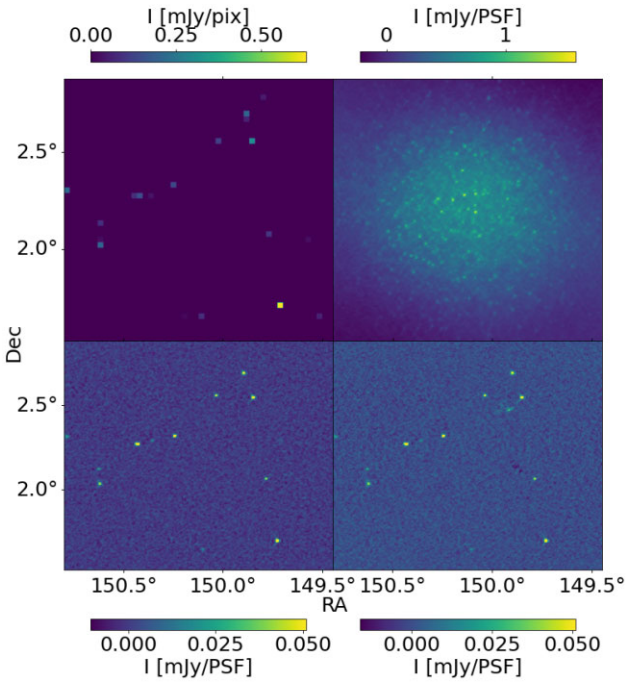

Figure 15. Top-left panel: The input image of H I signal in the 20$^{\text{th}}$ frequency bin around 1.096 GHz in Jy per pixel. The pixel size of the image is set to (100 arcsec)^2 to show clearly the position of the H I sources. Top-right panel: The output image of the simulation in the same frequency bin, with H I, foreground and thermal noise. Bottom-left panel: The output image for H I and thermal noise-only visibility simulation image without the foreground. Bottom-right panel: The image shown in the top-right panel after PCA cleaning. The image after PCA cleaning (bottom-right panel) matches the H I and thermal noise-only simulation well (bottom-left panel). In bottom panels, pixels brighter than 0.05 mJy per PSF are set to 0.05 mJy per PSF for better presentation.

We test our imaging and power spectrum estimation pipeline with H I and thermal noise-only visibility data and compare it with the input image within the telescope FoV. We find that the weighting of baselines can have a major effect on the resulting power spectrum as shown in Fig. 14. When uniform weighting in imaging is used, the narrow PSF results in overestimation of the power spectrum amplitude at small k. On the other hand, natural weighting achieves

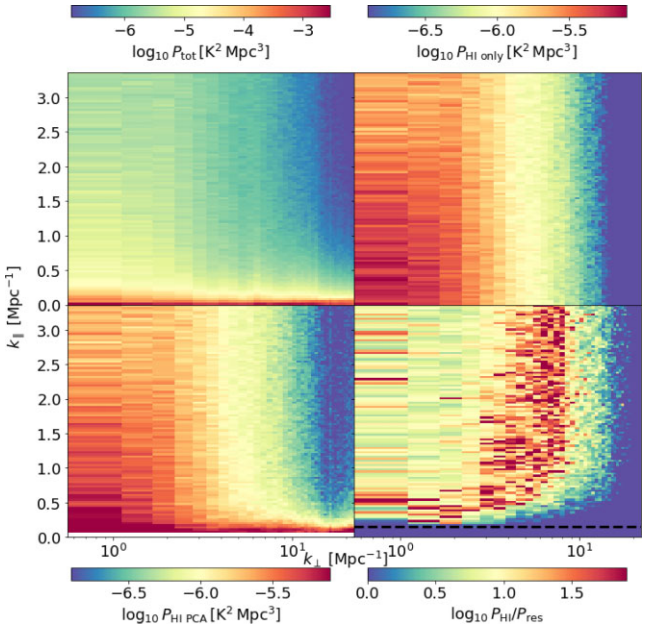

Figure 16. Top-left panel: The cylindrical temperature power spectrum of the images of the total intensity with thermal noise subtracted. Values under 10^{-7} K^2 Mpc^3 are masked for better presentation. Top-right panel: The cylindrical temperature power spectrum of the images of the H I signal. Values under 10^{-7} K^2 Mpc^3 are masked. Bottom-left panel: The cylindrical temperature power spectrum of the images of total intensity after PCA cleaning with thermal noise subtracted. Values under 10^{-7} K^2 Mpc^3 are masked. Bottom-right panel: The ratio of H I and the residual foreground power spectrum. Values under 1 are set to 1 for better presentation.

maximum sensitivity but underestimates the power spectrum due to the large PSF confusing different sources. Using Briggs weighting (Briggs 1995) and testing different robust parameters, we find that when the robustness parameter is set to 0.5, the resulting power spectrum achieves a balance between accuracy and sensitivity. The power spectrum of the image matches the input from $k_\parallel \sim 0.5$ up to $k_\parallel \sim 5\,\text{Mpc}^{-1}$, but suffers signal loss from the PSF convolution at smaller scales.

We generate image cubes for the simulated visibilities. The image cube is then mean-centred and subsequently a PCA is performed. We find that in image space, the mixture of H I signal with the thermal noise and foreground is more severe due to the convolution of the PSF and loss of information by gridding. More modes need to be removed for image space PCA when compared with the visibility space PCA in order to subtract the foreground at small $k_\parallel$ and widen the observation window. We find removing a total number of seven PCA modes gives the best result. We find that choosing more modes leads to overcleaning in the high-$k_\parallel$ modes while choosing less modes does not sufficiently remove the foregrounds at low $k_\parallel$. We show the results for the 10$^{\text{th}}$ frequency bin in Fig. 15. Comparing the lower panels, one can see PCA removes most of the foreground signal. The undercleaning of some foreground structure leaves residual sources, which are smooth in frequency. The cleaned images are then converted from Jy per PSF to Jy per pixel and corrected for primary beam attenuation.

We verify the frequency smoothness of residual foreground structure by calculating the cylindrical temperature power spectrum presented in Fig. 16. Comparing the power spectrum of images of H I and images of total intensity after PCA cleaning, the foreground residual mainly resides in low-$k_\parallel$ ranges. We find

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

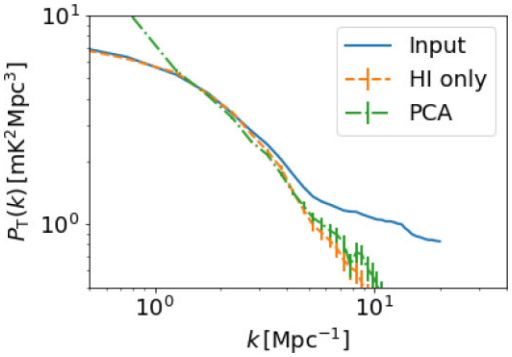

Figure 17. The power spectrum of input H I corresponding to the image size used in Section 6 ('Input'), compared with the power spectrum of images using H I visibilities ('H I only') and power spectrum of total intensity images of simulation with low-thermal noise level after PCA cleaning ('PCA').

that the observation window probes the region $k_\parallel > 0.2\,\mathrm{Mpc}^{-1}$ by checking the foreground wedge. We verify that raising this threshold does not have a significant impact on the resulting power spectrum.

The 1D temperature power spectrum result is presented in Fig. 17. Compared with the H I-only power spectrum, the PCA results overestimate the large $k \lesssim 1\,\mathrm{Mpc}^{-1}$ scales. This suggests that large-scale information is lost due to the excess gridding and imaging, as we do not see such effects in Section 5.5 when we apply PCA to the visibility data. Furthermore, scales $k \gtrsim 5\,\mathrm{Mpc}^{-1}$ are not recovered, due to the effects of weighting and PSF discussed previously.

In conclusion, we find that power spectrum estimation in image space is sensitive to the choice of baseline weighting and subsequent deconvolution of PSF, which requires more careful treatment. Applying component separation to the image cube recovers H I clustering at scales of roughly $1-5\,\mathrm{Mpc}^{-1}$. We compare it with PCA in the visibility data in the next section.

7 COMPARING FOREGROUND MITIGATION METHODS WITH MIGHTEE-LIKE NOISE LEVEL

In this section, we present a direct comparison of the foreground mitigation methods for power spectrum estimation using our simulation with realistic thermal noise level. As mentioned in Section 4.4, we generate MIGHTEE-like thermal noise consistent with the noise level of the entire MIGHTEE survey of 52 pointings and 1920 h. We then apply the foreground mitigation methods investigated in the previous sections. The edges of the 1D k bins are set to be logarithmically distributed from 0.1 to 30 Mpc^{-1} with 14 bins due to the lower signal-to-noise ratio.

Based on the previous discussion, we adopt the foreground avoidance described in Section 5.3, foreground removal using PCA in visibility with $\mathbf{R} = \mathbf{N}^{-1}$ described in Section 5.5, and in image space described in Section 6. The resulting projected power spectrum recovery is shown in Fig. 18. We find that for direct avoidance method, the foreground wedge criterion remains the same at $c_k = 0.25$. For PCA, the same $n_{\mathrm{fg}} = 2$ number of modes are removed and we find that power spectrum result converges when $c_k = 0.1$ for $k_\perp < 5\,\mathrm{Mpc}^{-1}$ and $c_k = 0.25$ for $k_\perp > 5\,\mathrm{Mpc}^{-1}$, tighter than the low-thermal noise case in Section 5.5. For the image output, we apply the exact same process in Section 6 with $n_{\mathrm{fg}} = 7$ number of modes removed. Overall, the methods used are robust to a substantially increased noise level.

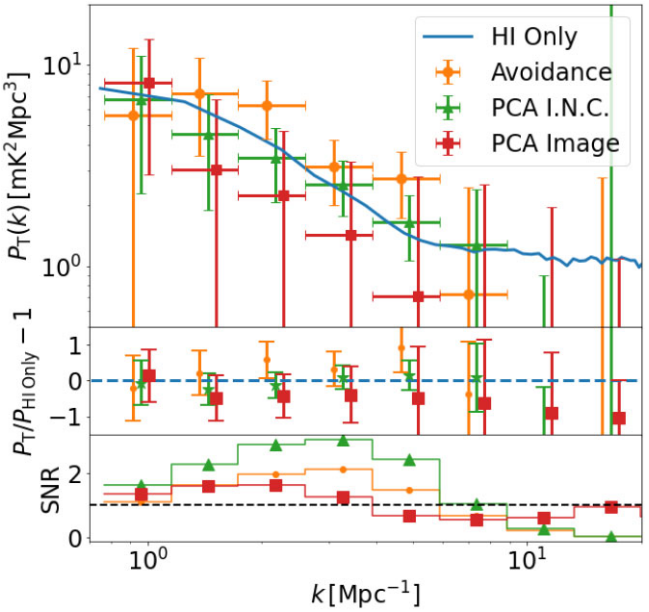

Figure 18. The 1D brightness temperature power spectrum from the H I-only visibility data ('HI Only'), simulation with MIGHTEE-like noise using foreground avoidance method described in Section 5.3 ('Avoidance'), simulation using PCA in visibility space with inverse noise covariance weighting ('PCA I.N.C.'), and simulation using PCA in image space ('PCA Image'). In addition, the fractional difference between the estimated 1D brightness temperature power spectrum and the H I-only simulation is presented below. The centres of the k bins for 'Avoidance' and 'PCA Image' are misplaced by 5 per cent for better comparison. The signal-to-noise ratios of the power spectrum measurements are shown at the bottom.

Comparing the results from avoidance and the PCA method, we find that PCA in visibility space provides less biased estimation of the H I power spectrum with projected uncertainties roughly 30 per cent smaller than direct avoidance. This is due to the larger observation window and the enabling of the inverse covariance weighting as discussed in Section 5. Overall, PCA in visibility space gives measurements of the H I power spectrum with signal-to-noise ratio $\lesssim 3$ from $k \sim 0.5$ up to $k \sim 5\,\mathrm{Mpc}^{-1}$ with $\Delta k \sim 0.5\,\mathrm{Mpc}^{-1}$.

Comparing the projected power spectrum using the image space PCA method in Fig. 18 and in Section 6, there is further signal loss in the presence of a higher thermal noise level. Overall, PCA in image space results in an overestimation of the H I power spectrum at $k < 1.0\,\mathrm{Mpc}^{-1}$. Compared with the visibility space PCA case, the measurement errors in the image space PCA case are $\sim$50-per cent larger and the signal-to-noise ratio is $\lesssim 2$ from $k \sim 0.5$ up to $k \sim 5\,\mathrm{Mpc}^{-1}$.

The fractional differences of the projected power spectrum measurements and the input as shown in Fig. 18 provide a direct comparison of the methods used. The visibility space PCA method has least bias on all scales $0.5\,\mathrm{Mpc}^{-1} \lesssim k \lesssim 10\,\mathrm{Mpc}^{-1}$ and the true signal is within the 1σ uncertainty. Foreground avoidance in visibility space and PCA in image space have larger bias in power spectrum estimation. We find that the projected uncertainties for PCA in image space are 50 per cent to 2 times larger than PCA in visibility space. This is due to three reasons. First, information is lost due to u–v gridding and limited image size which does not use all the baselines. Second, we weight the baselines with Briggs weighting robustness equal to 0.5 for reasons discussed in Section 6, which is far from

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

optimal sensitivity using natural weighting. Finally, the mixture of H I and foregrounds in principal components is more severe in image space due to the convolution of PSF, making foreground cleaning less effective. We also emphasize that the uncertainties for images are estimated using sampling variance in k space while the ones for visibilities are estimated using the quadratic estimator discussed in Section 3, which may not be a fair comparison (see e.g. Tan et al. 2021).

8 CONCLUSION

In this paper, we present an end-to-end pipeline of realistic signal modelling, observation simulation, and data analysis for low-redshift H I IM using radio interferometers. We have built a powerful simulation tool capable of generating input signal of H I, foregrounds and thermal noise for a given field of observation. The generated sky signal is used to simulate visibility data for any given observation strategy. We have developed a quadratic estimator for H I power spectrum estimation using visibility data. Multiple foreground mitigation strategies are examined.

We generate simulations of visibility data mimicking a typical tracking of *MeerKAT* for 220 frequency channels at $z \sim 0.25$–0.30, consistent with existing observations of deep fields such as COSMOS and DEEP2 to validate our estimation pipeline. By calculating the cylindrical power spectrum and comparing the contribution of different components, we find:

(1) The Galactic foreground signal, including the synchrotron radiation and the free–free emission, mainly affects angular scales larger than ∼0.1 degree. Its power drops significantly at smaller angular scales and is much smaller than the H I signal.

(2) Extragalactic radio sources dominate the foreground signal on smaller angular scales and increase its leakage into higher $k_\parallel$ modes. Comparing with the H I signal, the foreground contamination is severe and leaves a limited observation window for detection even for a deep field with sufficient source peeling.

(3) For scales probed by interferometry at low redshifts, the foreground can be well described by the covariance. This is due to the fact that at the scales of our interest, it is dominated by contributions from Poisson point sources.

(4) For observations of one field using small-FOV arrays such as *MeerKAT*, the limited survey volume induces large variance in the H I signal. We find that the variance due to survey volume becomes trivial for volumes larger than that of 5 MIGHTEE-like fields for a narrow frequency range, $V \sim 5 \times 20 \times 20 \times 200\,\mathrm{Mpc}^3$.

The foreground contamination calls for careful treatment of foreground mitigation methods. We compare, in detail, different ways to mitigate the impact of foregrounds using visibility data and conclude:

(1) When bright point sources are sufficiently removed, there exists an observation window at large $k_\parallel$ in which we can directly estimate the H I power spectrum. It provides biased estimation of the power spectrum, with an overestimation of 10 per cent. It is due to the leakage of the foregrounds into large $k_\parallel$ that cannot be completely excluded. Furthermore, data analysis techniques, such as inverse covariance weighting, become difficult as different $k_\parallel$ modes further mix under a non-uniform weighting, leading to more foreground contamination.

(2) Fitting the empirical covariance to extract out the smooth part for different annulus $|\boldsymbol{u}|$ bins, we find that subtracting the foreground covariance corrects for the overestimation at large $k < 1\,\mathrm{Mpc}^{-1}$ scales. It overcleans the foreground and results in signal loss at smaller scales. Extracting the smooth structure of data covariance falsely includes structure of H I covariance over large frequency ranges, which leads to overcleaning at small $k_\parallel$.

(3) By binning the visibility in annulus $|\boldsymbol{u}|$ bins and performing PCA in each bin, we find that the foreground contamination is reduced. The observation window widens due to improved cleaning of low-$k_\parallel$ modes and this works better with short baselines up to $k \sim 5\,\mathrm{Mpc}^{-1}$. It allows accurate estimation of H I power spectrum from $k \sim 0.5$ to $k \sim 10\,\mathrm{Mpc}^{-1}$. The reduced contamination allows inverse covariance weighting with little extra spillover of the foregrounds. Comparing with direct avoidance, it does not have the overestimation bias and has uncertainties that are ∼30-per cent smaller.

(4) We find that foreground mitigation in visibility space is robust to high levels of thermal noise consistent with the noise level of MIGHTEE for all 52 pointings and a redshift bin of $z \sim 0.25$–0.3. PCA in visibility space gives projected measurement of the power spectrum with signal-to-noise ratio ∼3 up to $k \sim 5\,\mathrm{Mpc}^{-1}$ and possible detection up to $k \sim 10\,\mathrm{Mpc}^{-1}$. Due to radio frequency interference (RFI) contaminations, data blocks from the MIGHTEE survey are usually divided into narrow-frequency sub-bands. Therefore, our results will likely apply to any sub-band from observations in the L band.

Our findings suggest the feasibility of using the visibility power spectrum and foreground extraction to measure the clustering of H I. At the angular scales of our interest ∼arcmin, the foreground components are largely stochastic and therefore can be removed using covariance-based methods such as PCA. Using surveys such as MIGHTEE, interferometric H I IM will be able to map the evolution of H I clustering at inner halo scales.

We have investigated the important question of choosing visibility space or image space to measure the H I power spectrum by directly comparing these two qualitatively. We generate image cubes of our visibility data to investigate power spectrum estimation in image space. We find:

(1) The power spectrum of the images relies heavily on the choice of baseline weighting, with uniform weighting overestimating the power spectrum and natural weighting causing signal loss due to the large PSF.

(2) We find that using Briggs weighting and performing PCA for the image cube, we can recover the H I power spectrum from $k \sim 1$ to $k \sim 5\,\mathrm{Mpc}^{-1}$. The component separation is less robust in the presence of large thermal noise.

(3) Due to the suboptimal weighting and information loss from visibility to image space, power spectrum estimation in image space performs relatively poorly when compared with the visibility approach, resulting in a more biased estimation with larger error bars. Comparing with PCA in visibility space, PCA in image space leads to reconstructed uncertainties that are at least 50-per cent larger.

Overall, we find that for future surveys, such as MIGHTEE, H I IM using visibility data is capable of measuring the H I power spectrum across a wide range of scales with high accuracy in narrow redshift bins, probing the redshift evolution of H I inside dark matter haloes. We provide a proof of concept study for data analysis using interferometric IM. Results from our pipeline strongly favour component separation and power spectrum estimation directly in visibility data without imaging. The pipeline can further include realistic beam model, polarization leakage, calibration error, and more. It will enable more studies in this topic towards future detection and constraints of cosmic H I in the near future.

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

ACKNOWLEDGEMENTS

We thank the anonymous referee for useful comments which improved our manuscript. We thank Steven Cunnington for discussions. LW is a UK Research and Innovation Future Leaders Fellow (grant MR/V026437/1). Apart from aforementioned packages, this work also uses PYTORCH (Paszke et al. 2019), NUMPY (Harris et al. 2020), SCIPY (Virtanen et al. 2020), ASTROPY (Astropy Collaboration 2018), CAMB (Lewis, Challinor & Lasenby 2000), NUMBA (Lam, Pitrou & Seibert 2015), and MATPLOTLIB (Hunter 2007). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

DATA AVAILABILITY

Data underlying this paper will be shared on reasonable request to the corresponding author.

REFERENCES

Alam S. et al., 2021, Phys. Rev. D, 103, 083533
Alonso D., Ferreira P. G., Santos M. G., 2014, MNRAS, 444, 3183
Anderson C. J. et al., 2018, MNRAS, 476, 3382
Asad K. M. B. et al., 2021, MNRAS, 502, 2970
Astropy Collaboration, 2018, AJ, 156, 123
Barry N., Hazelton B., Sullivan I., Morales M. F., Pober J. C., 2016, MNRAS, 461, 3135
Barry N. et al., 2019, ApJ, 884, 1
Battye R. A., Davies R. D., Weller J., 2004, MNRAS, 355, 1339
Battye R. A., Browne I. W. A., Dickinson C., Heron G., Maffei B., Pourtsidou A., 2013, MNRAS, 434, 1239
Bigot-Sazy M. A. et al., 2015, MNRAS, 454, 3240
Bonaldi A., Brown M. L., 2015, MNRAS, 447, 1973
Bondi M. et al., 2003, A&A, 403, 857
Bowman J. D., Morales M. F., Hewitt J. N., 2009, ApJ, 695, 183
Briggs D. S., 1995, AAS Meeting Abstracts, 112.02
Bull P., Ferreira P. G., Patel P., Santos M. G., 2015, ApJ, 803, 21
Chang T.-C., Pen U.-L., Peterson J. B., McDonald P., 2008, Phys. Rev. Lett., 100, 091303
Chapman E. et al., 2012, MNRAS, 423, 2518
Chapman E., Zaroubi S., Abdalla F. B., Dulwich F., Jelić V., Mort B., 2016, MNRAS, 458, 2928
Chatterjee S., Bharadwaj S., Marthi V. R., 2021, MNRAS, 500, 4398
Chen Z., Wolz L., Spinelli M., Murray S. G., 2021, MNRAS, 502, 5259
Cheng C. et al., 2018, ApJ, 868, 26
CHIME Collaboration, 2022, preprint (arXiv:2202.01242)
Choudhuri S., Bharadwaj S., Ghosh A., Ali S. S., 2014, MNRAS, 445, 4351
Chung D. T. et al., 2021, ApJ, 923, 188
Condon J. J., Ransom S. M., 2016, Essential Radio Astronomy. Princeton Univ. Press, Princeton, NJ
Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, AJ, 115, 1693
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Cornwell T. J., Golap K., Bhatnagar S., 2008, IEEE J. Sel. Top. Signal Process., 2, 647
Crain R. A. et al., 2017, MNRAS, 464, 4204
Cunnington S., Irfan M. O., Carucci I. P., Pourtsidou A., Bobin J., 2021, MNRAS, 504, 208
d'Amico G., Gleyzes J., Kokron N., Markovic K., Senatore L., Zhang P., Beutler F., Gil-Marín H., 2020, J. Cosmol. Astropart. Phys., 2020, 005
DeBoer D. R. et al., 2017, PASP, 129, 045001
Di Matteo T., Perna R., Abel T., Rees M. J., 2002, ApJ, 564, 576
Dickinson C., Davies R. D., Davis R. J., 2003, MNRAS, 341, 369
Dillon J. S. et al., 2014, Phys. Rev. D, 89, 023002
Dillon J. S. et al., 2015, Phys. Rev. D, 91, 123011
Dodelson S., 2003, Modern Cosmology. Academic Press, New York
Eisenstein D. J., Hu W., 1998, ApJ, 496, 605
Finkbeiner D. P., 2003, ApJS, 146, 407
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Hale C. L., Jarvis M. J., Delvecchio I., Hatfield P. W., Novak M., Smolčić V., Zamorani G., 2018, MNRAS, 474, 4133
Hamilton A. J. S., 1997, MNRAS, 289, 285
Hamilton A. J. S., Tegmark M., 2000, MNRAS, 312, 285
Harker G. et al., 2010, MNRAS, 405, 2492
Harris C. R. et al., 2020, Nature, 585, 357
Haslam C. G. T., Klein U., Salter C. J., Stoffel H., Wilson W. E., Cleary M. N., Cooke D. J., Thomasson P., 1981, A&A, 100, 209
Haslam C. G. T., Salter C. J., Stoffel H., Wilson W. E., 1982, A&AS, 47, 1
Heywood I. et al., 2022, MNRAS, 509, 2150
Högbom J. A., 1974, A&AS, 15, 417
Hothi I. et al., 2021, MNRAS, 500, 2264
Hu W., Wang X., Wu F., Wang Y., Zhang P., Chen X., 2020, MNRAS, 493, 5854
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Jarvis M. et al., 2016, Proc. MeerKAT Science: On the Pathway to the SKA, The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey. p. 6
Jonas J. L., Baart E. E., Nicolson G. D., 1998, MNRAS, 297, 977
Kaiser N., 1987, MNRAS, 227, 1
Kern N. S., Liu A., 2021, MNRAS, 501, 1463
Kovetz E. D. et al., 2017, preprint (arXiv:1709.09066)
Lacy M. et al., 2020, PASP, 132, 035001
Lam S. K., Pitrou A., Seibert S., 2015, Second Workshop on the LLVM Compiler Infrastructure in HPC (LLVM'15), Numba: A LLVM-based Python JIT Compiler. Assoc. Comput. Mach., NY, USA
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Lian X., Xu H., Zhu Z., Hu D., 2020, MNRAS, 496, 1232
Liu A., Shaw J. R., 2020, PASP, 132, 062001
Liu A., Tegmark M., 2011, Phys. Rev. D, 83, 103006
Liu A., Tegmark M., Zaldarriaga M., 2009, MNRAS, 394, 1575
Liu A., Parsons A. R., Trott C. M., 2014a, Phys. Rev. D, 90, 023018
Liu A., Parsons A. R., Trott C. M., 2014b, Phys. Rev. D, 90, 023019
McMullin J. P., Waters B., Schiebel D., Young W., Golap K., 2007, in Shaw R. A., Hill F., Bell D. J., eds, ASP Conf. Ser. Vol. 376, Astronomical Data Analysis Software and Systems XVI. Astron. Soc. Pac., San Francisco, p. 127
Madau P., Meiksin A., Rees M. J., 1997, ApJ, 475, 429
Mao Y., Tegmark M., McQuinn M., Zaldarriaga M., Zahn O., 2008, Phys. Rev. D, 78, 023529
Masui K. W. et al., 2013, ApJ, 763, L20
Matthews A. M., Condon J. J., Cotton W. D., Mauch T., 2021, ApJ, 909, 193
Mauch T. et al., 2020, ApJ, 888, 61
Mertens F. G., Ghosh A., Koopmans L. V. E., 2018, MNRAS, 478, 3640
Mertens F. G. et al., 2020, MNRAS, 493, 1662
Meyer M., Robotham A., Obreschkow D., Westmeier T., Duffy A. R., Staveley-Smith L., 2017, Publ. Astron. Soc. Aust., 34, 52
Morales M. F., Hewitt J., 2004, ApJ, 615, 7
Morales M. F., Beardsley A., Pober J., Barry N., Hazelton B., Jacobs D., Sullivan I., 2019, MNRAS, 483, 2207
Mort B. J., Dulwich F., Salvini S., Adami K. Z., Jones M. E., 2010, Proc. 2010 IEEE Int. Symp. Phased Array Systems and Technology. IEEE, NJ, p. 690
Murray S. G., 2018, J. Open Source Softw., 3, 850
Murray S. G., Power C., Robotham A. S. G., 2013, Astron. Comput., 3, 23
Murray S. G., Trott C. M., Jordan C. H., 2017, ApJ, 845, 7
Murray S. G., Diemer B., Chen Z., Neuhold A. G., Schnapp M. A., Peruzzi T., Blevins D., Engelman T., 2021, Astron. Comput., 36, 100487
Myers S. T. et al., 2003, ApJ, 591, 575
Nasirudin A., Murray S. G., Trott C. M., Greig B., Joseph R. C., Power C., 2020, ApJ, 893, 118
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Olivari L. C., Dickinson C., Battye R. A., Ma Y. Z., Costa A. A., Remazeilles M., Harper S., 2018, MNRAS, 473, 4242

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

Overzier R. A., Röttgering H. J. A., Rengelink R. B., Wilman R. J., 2003, A&A, 405, 53
Padmanabhan H., Refregier A., Amara A., 2019, MNRAS, 485, 4060
Pandey S. et al., 2022, Phys. Rev. D, 106, 043520
Parsons A., Pober J., McQuinn M., Jacobs D., Aguirre J., 2012a, ApJ, 753, 81
Parsons A. R., Pober J. C., Aguirre J. E., Carilli C. L., Jacobs D. C., Moore D. F., 2012b, ApJ, 756, 165
Parsons A. R. et al., 2014, ApJ, 788, 106
Paszke A. et al., 2019, in Wallach H., Larochelle H., Beygelzimer A., d'Alché-Buc F., Fox E., Garnett R., eds, Advances in Neural Information Processing Systems 32. Curran Associates, Inc., NY, USA, p. 8024
Patil A. H. et al., 2017, ApJ, 838, 65
Paul S., Santos M. G., Townsend J., Jarvis M. J., Maddox N., Collier J. D., Frank B. S., Taylor R., 2021, MNRAS, 505, 2039
Peebles P. J. E., 1974, ApJ, 189, L51
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Planck Collaboration VI, 2020, A&A, 641, A6
Reich P., Testori J. C., Reich W., 2001, A&A, 376, 861
Remazeilles M., Dickinson C., Banday A. J., Bigot-Sazy M. A., Ghosh T., 2015, MNRAS, 451, 4311
Santos M. G., Cooray A., Knox L., 2005, ApJ, 625, 575
Santos M. G. et al., 2017, Proc. Sci., MeerKAT Science: On the Pathway to the SKA. PoS#32
Sarkar A. K., Bharadwaj S., Marthi V. R., 2018, MNRAS, 473, 261
Schaan E., White M., 2021, J. Cosmol. Astropart. Phys., 2021, 068
Scoville N. et al., 2007, ApJS, 172, 1
Siewert T. M. et al., 2020, A&A, 643, A100
Spinelli M., Zoldan A., De Lucia G., Xie L., Viel M., 2020, MNRAS, 493, 5434
Spinelli M., Bernardi G., Garsden H., Greenhill L. J., Fialkov A., Dowell J., Price D. C., 2021, MNRAS, 505, 1575
Square Kilometre Array Cosmology Science Working Group et al., 2020, Publ. Astron. Soc. Aust., 37, e007
Switzer E. R. et al., 2013, MNRAS, 434, L46
Tan J. et al., 2021, ApJS, 255, 26
Tegmark M., 1997, Phys. Rev. D, 55, 5895
Tegmark M., 1998, in Olinto A. V., Frieman J. A., Schramm D. N., eds, Eighteenth Texas Symposium on Relativistic Astrophysics and Cosmology. World Scientific, NJ, USA, p. 270
The HERA Collaboration, 2022, ApJ, 925, 221
Thyagarajan N. et al., 2015a, ApJ, 804, 14
Thyagarajan N. et al., 2015b, ApJ, 807, L28
Thyagarajan N., Parsons A. R., DeBoer D. R., Bowman J. D., Ewall-Wice A. M., Neben A. R., Patra N., 2016, ApJ, 825, 9
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878
Trott C. M. et al., 2016, ApJ, 818, 139
Villaescusa-Navarro F. et al., 2018, ApJ, 866, 135
Virtanen P. et al., 2020, Nature Methods, 17, 261
Wolz L., Tonini C., Blake C., Wyithe J. S. B., 2016, MNRAS, 458, 3399
Wolz L., Murray S. G., Blake C., Wyithe J. S., 2019, MNRAS, 484, 1007
Wolz L. et al., 2022, MNRAS, 510, 3495
Wyithe J. S. B., Loeb A., 2009, MNRAS, 397, 1926
Zheng Z. et al., 2005, ApJ, 633, 791
Zheng H. et al., 2017, MNRAS, 464, 3486
Zonca A., Singer L., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K., 2019, J. Open Source Softw., 4, 1298

APPENDIX A: BRIGHTNESS TEMPERATURE POWER SPECTRUM

In this section, we present a detailed derivation of equation (36). We start first with the Dirac δ function in comoving space for transverse and line-of-sight directions. We define Dirac δ function $\tilde{\delta}_{\rm D}(k_\parallel)$ along the $k_\parallel$ axis such that

$$\Delta X \int \frac{{\rm d}k}{2\pi} \tilde{\delta}_{\rm D}(k-k_\parallel)\tilde{f}(k) = \tilde{f}(k_\parallel) \tag{A1}$$

for arbitrary function $\tilde{f}$, where ΔX is the line-of-sight length of the survey volume and can be written as $\Delta X = Y\delta f N_{\rm ch}$ with δf being the frequency bandwidth and $N_{\rm ch}$ the number of channels.

The survey volume $\mathbb{V}$ can be written as $\mathbb{V} = \mathbb{S}\cdot\Delta X$ where $\mathbb{S}$ is the survey area. As we will see later, both $\mathbb{V}$ and $\mathbb{S}$ are just abstract quantities to account for the physical units, and will be cancelled out in the final equation. We can then write down the definition for 2D Dirac delta function

$$\mathbb{S}\int \frac{{\rm d}^2k_\perp}{(2\pi)^2}\tilde{\delta}^2_{\rm D}(\boldsymbol{k}_\perp - \boldsymbol{k}'_\perp)\tilde{f}(\boldsymbol{k}_\perp) = \tilde{f}(\boldsymbol{k}'_\perp) \tag{A2}$$

for arbitrary function $\tilde{f}(\boldsymbol{k}_\perp)$.

For a frequency-independent beam response $A(l, m, f) \equiv A(l, m)$, we can write its Fourier transform as

$$\tilde{A}(\boldsymbol{k}) = \frac{X^2\Delta X}{\mathbb{V}}\tilde{\delta}_{\rm D}(k_\parallel)\tilde{A}_\perp\left(\boldsymbol{k}_\perp\right). \tag{A3}$$

Using the previous equation we can express equation (15) as 2D integrals:

$$\begin{aligned} P_{\rm d} = &\left(\frac{2k_B}{\lambda^2}\right)^2\frac{\mathbb{V}^4}{X^4Y^2}\int\frac{{\rm d}^3k'}{(2\pi)^3}\frac{{\rm d}^3k''}{(2\pi)^3}\tilde{A}(\boldsymbol{k}-\boldsymbol{k}')\tilde{A}^*(\boldsymbol{k}-\boldsymbol{k}'') \\ &\times\langle\tilde{T}(\boldsymbol{k}')\tilde{T}^*(\boldsymbol{k}'')\rangle = \left(\frac{2k_B}{\lambda^2}\right)^2\frac{\mathbb{V}^2}{Y^2} \\ &\int\frac{{\rm d}^2k'_\perp}{(2\pi)^2}\frac{{\rm d}^2k''_\perp}{(2\pi)^2}\tilde{A}_\perp\left(\boldsymbol{k}_\perp-\boldsymbol{k}'_\perp\right)\tilde{A}^*_\perp\left(\boldsymbol{k}_\perp-\boldsymbol{k}''_\perp\right) \\ &\times\langle\tilde{T}(\boldsymbol{k}'_\perp,k_\parallel)\tilde{T}^*(\boldsymbol{k}''_\perp,k_\parallel)\rangle. \end{aligned} \tag{A4}$$

The brightness power spectrum can be written as

$$\mathbb{V}\langle\tilde{T}(\boldsymbol{k}'_\perp,k_\parallel)\tilde{T}^*(\boldsymbol{k}''_\perp,k_\parallel)\rangle = \tilde{\delta}^2_{\rm D}(\boldsymbol{k}'_\perp-\boldsymbol{k}''_\perp)P_{\rm T}(\boldsymbol{k}'_\perp,k_\parallel), \tag{A5}$$

which leads to

$$\begin{aligned} P_{\rm d} = &\left(\frac{2k_B}{\lambda^2}\right)^2\frac{\mathbb{V}}{Y^2}\int\frac{{\rm d}^2k'_\perp}{(2\pi)^2}\frac{{\rm d}^2k''_\perp}{(2\pi)^2}\tilde{A}_\perp\left(\boldsymbol{k}_\perp-\boldsymbol{k}'_\perp\right)\tilde{A}^*_\perp\left(\boldsymbol{k}_\perp-\boldsymbol{k}''_\perp\right) \\ &\times\tilde{\delta}^2_{\rm D}(\boldsymbol{k}'_\perp-\boldsymbol{k}''_\perp)P_{\rm T}(\boldsymbol{k}'_\perp,k_\parallel) = \left(\frac{2k_B}{\lambda^2}\right)^2\frac{\mathbb{V}}{\mathbb{S}Y^2} \\ &\int\frac{{\rm d}^2k'_\perp}{(2\pi)^2}\left|\tilde{A}_\perp\left(\boldsymbol{k}_\perp-\boldsymbol{k}'_\perp\right)\right|^2P_{\rm T}(\boldsymbol{k}'_\perp,k_\parallel) \\ = &\left(\frac{2k_B}{\lambda^2}\right)^2\frac{\Delta X}{Y^2}\int\frac{{\rm d}^2k'_\perp}{(2\pi)^2}\left|\tilde{A}_\perp\left(\boldsymbol{k}_\perp-\boldsymbol{k}'_\perp\right)\right|^2P_{\rm T}(\boldsymbol{k}'_\perp,k_\parallel). \end{aligned} \tag{A6}$$

Equation (A6) shows that the delay power spectrum mixes different modes of $P_{\rm T}$ through the beam response. For wide-FOV arrays, the large beam corresponds to a narrow Fourier pair $\tilde{A}$ so that $P_{\rm T}$ can be extracted out of the integral. Alternatively, one can also deconvolve the beam response together with the w-projection kernel (Cornwell, Golap & Bhatnagar 2008), as shown for example in Trott et al. (2016). Here, we are dealing with dish arrays with narrow beams so we take the mode-mixing effects of beam response into account explicitly later in this section while avoiding the computationally consuming deconvolution.

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024

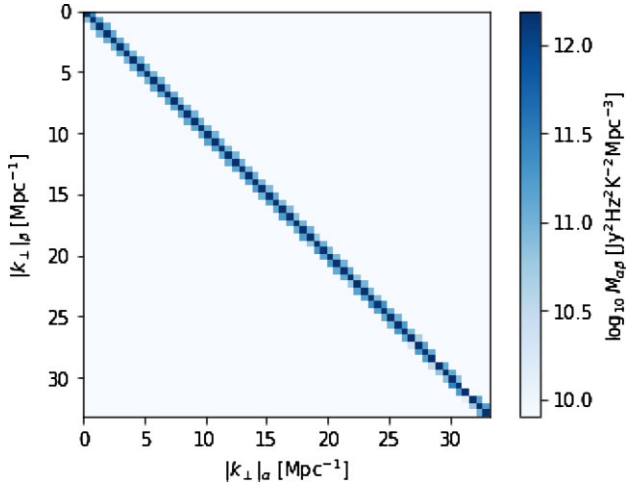

Figure A1. The mode-mixing matrix $\mathcal{M}_{\alpha\beta}$ used in this paper calculated according to equation (36) with the choice of annulus u–v bins mentioned in Section 3. Values below 10^{10} Jy^2 Hz^2 K^{-2} Mpc^{-3} are masked for better presentation.

Recalling equation (27), we can substitute the $|\tilde{\mathbf{V}}^i|^2$ with $P_{\rm d}$ in the previous equation to get

$$\hat{p}^{\rm d}_\alpha = \frac{\sum_i \chi_\alpha(\boldsymbol{u}_i,\eta_i)|\tilde{\mathbf{V}}^i|^2}{\sum_i \chi_\alpha(\boldsymbol{u}_i,\eta_i)} = \left(\frac{2k_B}{\lambda^2}\right)^2 \frac{\Delta X}{Y^2\sum_i \chi^i_\alpha} \times \sum_i \chi^i_\alpha \int \frac{{\rm d}^2k_\perp}{(2\pi)^2}\left|\tilde{A}_\perp\left(\boldsymbol{k}^i_\perp - \boldsymbol{k}_\perp\right)\right|^2 P_{\rm T}(\boldsymbol{k}_\perp, k^i_\parallel), \tag{A7}$$

where $\boldsymbol{k}^i_\perp = 2\pi\boldsymbol{u}_i/X$ and i loops over all Fourier-transformed u–v grids. Now recalling equation (20), we can further expand the previous equation:

$$\hat{p}^{\rm d}_\alpha = \left(\frac{2k_B}{\lambda^2}\right)^2 \frac{\Delta X}{Y^2\sum_i \chi^i_\alpha}\sum_i \int \frac{{\rm d}^2\boldsymbol{k}_\perp}{(2\pi)^2}\chi^i_\alpha \left|\tilde{A}_\perp\left(\boldsymbol{k}^i_\perp - \boldsymbol{k}_\perp\right)\right|^2 \times \sum_\beta \chi^i_\beta \hat{P}_{\rm T}\left(|\boldsymbol{k}_\perp|_\beta, k^i_\parallel\right) = \left(\sum_\beta\right)\mathcal{M}_{\alpha\beta}\,\hat{p}^{\rm T}_\beta. \tag{A8}$$

Finally, using $\Delta X = Y\delta f N_{\rm ch}$, we show that

$$(\mathcal{M})_{\alpha\beta} = \left(\frac{2k_{\rm B}}{\lambda^2}\right)^2 \frac{N_{\rm ch}\,\delta f}{Y\sum_i \chi_\alpha(\boldsymbol{k}_i)}\sum_i \int \frac{{\rm d}^2\boldsymbol{k}_\perp}{(2\pi)^2}\chi_\alpha(\boldsymbol{k}_i) \times \left|\tilde{A}_\perp\left(\boldsymbol{k}^i_\perp - \boldsymbol{k}_\perp\right)\right|^2 \chi_\beta(\boldsymbol{k}_\perp, k^i_\parallel), \tag{A9}$$

which is equation (36). Note that the previous equation depends on the summation of each baseline i. It takes into account of the sampling of u–v plane assuming equal data weights for each baseline, similar to 'natural weighting' in interferometric imaging. One can also define the selection function $\chi_{\alpha,\beta}$ differently to change the weighting. Therefore, the matrix formalism can effectively deconvolve the primary beam response and PSF simultaneously.

As mentioned in Section 4, we grid u–v plane into annulus bins with the edges of the bin being [0,100,200,...,6000]. The resulting $\mathcal{M}_{\alpha\beta}$ is shown in Fig. A1. As shown, the mixture of different angular modes due to the size of the beam mainly affects bins near the diagonal, with the width of $\Delta|\boldsymbol{k}_\perp| \sim 0.5\,{\rm Mpc}^{-1}$.

APPENDIX B: CONVERTING H I MASS TO FLUX DENSITY

In this section, we briefly derive the conversion between H I mass to flux density for cosmological simulations.

Suppose we grid the sky with equal area pixels, each with an area of $\Omega_{\rm pix}$ (this area is purely pedagogical). In one frequency channel with comoving distance X and line-of-sight scale of ΔX, we have

$$T^{\rm pix}_{\rm HI} = \sum_i \frac{C^i_{\rm HI}M^i_{\rm HI}}{V_{\rm pix}},\quad V_{\rm pix} = X^2\Omega_{\rm pix}\Delta X, \tag{B1}$$

where i loops over the H I sources within the pixel.

The flux density from one pixel can then be calculated as

$$I_{\rm pix} = \sum_i \frac{C^i_{\rm HI}M^i_{\rm HI}}{X^2\Omega_{\rm pix}\Delta X}\frac{2k_{\rm B}}{\lambda_i^2}\Omega_{\rm pix} = \sum_i \frac{2k_{\rm B}}{\lambda_i^2}\frac{C^i_{\rm HI}M^i_{\rm HI}}{X^2\Delta X}, \tag{B2}$$

and the flux density of each source is

$$I_i = \frac{2k_{\rm B}}{\lambda_i^2}\frac{C^i_{\rm HI}M^i_{\rm HI}}{X^2\Delta X}. \tag{B3}$$

In reality, the flux density of the sources depends on the peculiar velocity (see e.g. Meyer et al. 2017). Our result instead depends on the bandwidth of the frequency channel. It is assumed that the frequency displacement caused by peculiar velocity does not misplace H I galaxies into other frequency bins and that the width of the emission profile is negligible when compared with the frequency channel bandwidth.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article/518/2/2971/6825493 by EMBL user on 17 December 2024