



# Emission-line Stacking of 21 cm Intensity Maps with MeerKLASS: Inference Pipeline and Application to the *L*-band Deep-field Data

Zhaoting Chen (陈兆庭)<sup>1</sup> , Steven Cunnington<sup>2,3</sup> , Alkistis Pourtsidou<sup>1,4</sup> , Laura Wolz<sup>3</sup> , Marta Spinelli<sup>5,6</sup> , José Luis Bernal<sup>7</sup> , Matilde Barberi-Squarotti<sup>8,9,10</sup> , Stefano Camera<sup>11,12,13,14</sup> , Isabella P. Carucci<sup>15,16</sup> , José Fonseca<sup>6,17</sup> , Keith Grainge<sup>3</sup> , Melis O. Irfan<sup>18</sup> , Mario G. Santos<sup>6</sup> , and Jingying Wang (王婧颖)<sup>19,6</sup>

(MeerKLASS Collaboration)

<sup>1</sup> Institute for Astronomy, The University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK; [zhaoting.chen@roe.ac.uk](mailto:zhaoting.chen@roe.ac.uk)

<sup>2</sup> Institute of Cosmology & Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth, PO1 3FX, UK

<sup>3</sup> Jodrell Bank Centre for Astrophysics, Department of Physics & Astronomy, The University of Manchester, Manchester M13 9PL, UK

<sup>4</sup> Higgs Centre for Theoretical Physics, School of Physics and Astronomy, Edinburgh EH9 3FD, UK

<sup>5</sup> Observatoire de la Côte d'Azur, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice cedex 4, France

<sup>6</sup> Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Road, Cape Town 7535, South Africa

<sup>7</sup> Instituto de Física de Cantabria (IFCA), CSIC-University de Cantabria, Avda. de los Castros s/n, E-39005 Santander, Spain

<sup>8</sup> Università degli Studi di Milano, Via Celoria 16, 20133 Milan, Italy

<sup>9</sup> INAF—Istituto Nazionale di Astrofisica, Osservatorio Astrofisico di Brera-Merate, Via Brera 28, 20121, Milan, Italy

<sup>10</sup> INFN—Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Via Celoria 16, 20133 Milan, Italy

<sup>11</sup> Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy

<sup>12</sup> INFN—Istituto Nazionale di Fisica Nucleare, Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy

<sup>13</sup> INAF—Istituto Nazionale di Astrofisica, Osservatorio Astrofisico di Torino, Strada Osservatorio 20, 10025 Pino Torinese, Italy

<sup>14</sup> Department of Physics & Astronomy, University of the Western Cape, 7535 Cape Town, South Africa

<sup>15</sup> INAF—Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, 34131 Trieste, Italy

<sup>16</sup> IFPU—Institute for Fundamental Physics of the Universe, Via Beirut 2, 34151 Trieste, Italy

<sup>17</sup> Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto CAUP, Rua das Estrelas, PT4150-762 Porto, Portugal

<sup>18</sup> Institute of Astronomy, University of Cambridge, Cambridge, CB3 0HA, UK

<sup>19</sup> Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Road, Shanghai, 200030, People's Republic of China; [jywang@shao.ac.cn](mailto:jywang@shao.ac.cn)

Received 2025 April 4; revised 2025 May 6; accepted 2025 May 12; published 2025 June 30

## Abstract

We present a novel analysis of observational systematics through the emission-line stacking of the MeerKLASS *L*-band deep-field intensity maps, following the detection in MeerKLASS Collaboration et al. A stacking signal is obtained by stacking the 21 cm intensity map cubelets around the galaxy positions from the GAMA survey at  $0.39 \lesssim z \lesssim 0.46$ . An extensive simulation framework is built to study the viability of the stacking detection, the covariance estimation, and the model inference, which are then applied to the data. The statistical significance of the detection is  $8.66\sigma$  when averaged into an angular map, and  $7.45\sigma$  when averaged into a spectrum. The stacked spectrum exhibits an oscillating component of systematics, and we provide evidence that these systematics are a convolutional effect on the map data. The oscillation frequency matches the diffraction from the secondary reflector into the primary beam of the MeerKAT telescope. Bayesian inference can be used to constrain the systematics and the average HI emission of the galaxies. The fitting of the parameters gives a constraint on the systematics frequency  $\nu_{\text{sys}}$  [MHz] =  $17.90_{-4.27}^{+6.53}$ . The posterior of the systematics amplitude reaches the wide prior and gives  $A_{\text{sys}} = 0.50_{-0.33}^{+0.33}$ . A tentative measurement of the average HI mass of the sources is achieved at  $\log_{10}[(M_{\text{HI}})/M_{\odot}] = 9.84_{-0.59}^{+0.48}$ , which is an underestimation limited by the narrow redshift bin, the strong degeneracy with the systematics, and the low-density galaxy sample. These shortfalls will be resolved for future MeerKLASS data to enable accurate measurements of the HI density through stacking of intensity maps.

*Unified Astronomy Thesaurus concepts:* [Observational cosmology \(1146\)](#); [Radio astronomy \(1338\)](#); [H I line emission \(690\)](#)

## 1. Introduction

A primary goal of observational cosmology is to map the distribution of the cosmic large-scale structure (LSS) throughout the evolutionary history of the Universe. To probe the initial conditions and the subsequent growth of the cosmic structure, different tracers covering a wide range of cosmological redshifts are needed. Among them, neutral hydrogen (HI) intensity mapping (P. Madau et al. 1997; S. Bharadwaj et al. 2001; R. A. Battye et al. 2004; T.-C. Chang et al. 2008;

J. S. B. Wyithe et al. 2008) emerges as a unique probe. Instead of resolving the sources of HI, intensity mapping aims to map the flux density of the 21 cm emission line from the hyperfine transition of the HI (H. Hellwig et al. 1970) over large cosmological volumes. It can be used to probe the cosmic dark ages (A. Lewis & A. Challinor 2007), the cosmic dawn, and the subsequent epoch of reionization (S. R. Furlanetto et al. 2006). After the cosmic reionization is complete at  $z \lesssim 5.5$  (S. E. I. Bosman et al. 2022; Y. Zhu et al. 2022, 2024; B. Spina et al. 2024), the 21 cm line traces primarily the dark matter halos and structures therein (F. Villaescusa-Navarro et al. 2018). As the 21 cm line has a fixed rest frame frequency, intensity mapping surveys are spectroscopic in nature. Surveys in the near future with the Square Kilometre Array Observatory (SKAO)



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

will be able to measure the HI power spectrum and constrain the underlying cosmological model precisely, matching the precision of current optical galaxy surveys (SKAO Cosmology Science Working Group et al. 2020). Moreover, it has the unique advantage of probing the high-redshift, post-reionization Universe  $z \gtrsim 3.0$  with fine redshift resolution (Z. Chen et al. 2023a), and holds great synergy potential with the line-emission intensity mapping of other spectral lines (J. L. Bernal & E. D. Kovetz 2022).

Tremendous efforts have been made toward measuring the post-reionization HI intensity mapping signal at large cosmological scales. Detections of the HI clustering in cross correlation with optical galaxies have been achieved, for example, using the Green Bank Telescope (K. W. Masui et al. 2013; E. R. Switzer et al. 2013; L. Wolz et al. 2022) and the Parkes telescope (C. J. Anderson et al. 2018). There are numerous current and forthcoming experiments conducting post-reionization 21 cm surveys, such as the Baryon Acoustic Oscillations from Integrated Neutral Gas Observations telescope (E. Abdalla et al. 2022), the Canadian Hydrogen Intensity Mapping Experiment (CHIME; CHIME Collaboration et al. 2022), the Canadian Hydrogen Observatory and Radio-transient Detector (K. Vanderlinde et al. 2019), the Five-hundred-meter Aperture Spherical radio Telescope (Y. Li et al. 2023), the Hydrogen Intensity and Real-Time Analysis Experiment (D. Crichton et al. 2022), the Tianlai array (S. Zuo et al. 2021), and the upgraded Giant Metrewave Radio Telescope (S. Pal et al. 2022). In this paper, we focus on the progress made by the MeerKAT Large Area Synoptic Survey (MeerKLASS; M. Santos et al. 2016) using the MeerKAT telescope. The MeerKAT telescope is the precursor to, and will be part of, the midfrequency array of the SKAO (SKA-Mid). The MeerKLASS survey has produced 21 cm intensity maps in the  $L$  band, which were used to detect the cross-power spectrum with optical galaxies (J. Wang et al. 2021; S. Cunnington et al. 2023a; I. P. Carucci et al. 2024). The MeerKLASS  $L$ -band deep-field observations, produced from 62 hr of observation over  $236 \text{ deg}^2$ , represent the deepest single-dish HI intensity maps to date, as presented in MeerKLASS Collaboration et al. (2025, hereafter MK25).

The measurement of the 21 cm signal in the intensity maps relies on the removal of foregrounds. While the 21 cm line is an emission-line signal, the foregrounds, such as Galactic synchrotron and extragalactic radio sources, are spectrally smooth and can therefore be separated using the technique of principal component analysis (PCA) with signal loss correction (e.g., S. Cunnington et al. 2023b). As a result, the residual signal is susceptible to systematics that break the assumption that the foregrounds are spectrally smooth. The origins of such systematics can be polarization leakage (D. Alonso et al. 2014; I. P. Carucci et al. 2020; S. Cunnington et al. 2021), chromaticity of the instrument beam (S. D. Matshawule et al. 2021), calibration errors (e.g., I. Heywood et al. 2020), etc. These effects induce contamination of the signal in the measured summary statistics, which is at the same order of magnitude as the HI signal for the MeerKLASS  $L$ -band intensity maps (see Figures 14 and 20 of MK25). Understanding and modeling the systematics are therefore important to enable the inference of the cosmological HI signal, analogous to the marginalization over nuisance parameters in optical galaxy surveys of LSS.

In this paper, we propose a novel way of modeling and constraining a type of multiplicative systematic induced by chromatic primary beam ripple, using the emission-line stacking of the 21 cm intensity maps. Stacking the emission-line signal onto positions of optical galaxies is in itself a powerful way of probing cosmic HI, and is one of the main scientific goals of interferometric MeerKAT observations. For example, the MeerKAT International Gigahertz Tiered Extragalactic Exploration survey (M. Jarvis et al. 2016; N. Maddox et al. 2021) has produced stacking measurements of HI galaxies at  $z \sim 0.37$  (F. Sinigaglia et al. 2022b, 2024; A. Bianchetti et al. 2025) to probe scaling relations and dependencies of HI mass on the LSS environment (see also A. A. Ponomareva et al. 2021, 2023; S. Ranchod et al. 2021; S. H. A. Rajohnson et al. 2022; M. N. Tudorache et al. 2022; H. Pan et al. 2023; I. Heywood et al. 2024; M. N. Tudorache et al. 2024 for direct detections of HI sources at  $z \lesssim 0.01$ ). These stacking measurements utilize the information on source positions from external catalogs with maps in  $\sim$ arcsecond resolution made from interferometric observations (see also, e.g., CHIME Collaboration et al. 2023). Similarly, single-dish intensity mapping observations can also be used for stacking, typically with maps in  $\sim$ arcminute resolution such as HI stacking using the Parkes telescope (D. Tramonte et al. 2019; D. Tramonte & Y.-Z. Ma 2020) and CO emission-line stacking using the CO Mapping Array Project data (D. A. Dunne et al. 2024). Detections of Ly $\alpha$  intensity mapping have also been achieved using the Hobby–Eberly Telescope Dark Energy Experiment observations (M. Lujan Niemeyer et al. 2022a, 2022b).

The stacking measurement presented in MK25, on the other hand, is made with intensity maps of  $\sim 1^\circ$  resolution. The resolution corresponds to large  $\sim 30$  Mpc scales, over which we expect several galaxies contributing to the stacking measurements. Therefore, modeling clustering beyond Poisson statistics is required for a precise prediction (J. L. Bernal 2024; see also P. Renard et al. 2024 for the case of stacking Ly $\alpha$  emission). Combined with the fact that the intensity map is affected by signal loss from foreground removal, forward modeling of the signal is required to infer the properties of the underlying HI sources as well as the systematics in the data, which we aim to demonstrate in this work.

The modeling of the stacked signal can incorporate modeling of systematic effects in the data. The stacked signal can be averaged into a map of stacked emission in the angular plane. The excess emission in the center region of the stacked map against the noise background can be used to describe the convolution of the instrument beam with the HI signal. Similar analysis of systematics can be found in weak lensing, for example, for cross-correlating the point-spread function with galaxy shapes (e.g., T. Zhang et al. 2023). On the other hand, the stacked signal can also be averaged into a spectrum along the frequency direction. The stacked spectrum can then be used to examine the chromatic structure of the data that affects the two-point statistics, such as the structure seen in the line-of-sight power spectrum (e.g., M. Spinelli et al. 2022). The complexity of the effects of the systematics requires detailed simulation and validation pipeline to study the viability of inference using emission-line stacking, which can be applied to the single-dish HI intensity mapping data using MeerKLASS and future SKAO. This work lays the foundation for the pipeline and, for the first time, applies model inference to the emission-line stacking using MeerKLASS data.

The rest of the paper is organized as follows: In Section 2, we describe the specifications of the MeerKLASS *L*-band deep-field data. In Section 3, we present the simulation pipeline of stacked HI signal. We validate the detectability of the stacked signal against the presence of thermal noise and foregrounds and establish the fact that the stacked signal needs to be forward modeled in Section 4. In Section 5, we discuss the method for covariance estimation. We then describe the data analysis pipeline to measure the stacked signal in the data in Section 6, which is an update of the one presented in MK25. In Section 7, we present an analysis that pinpoints the nature of the systematics in the data. We then proceed to parameterize the systematics to enable the modeling of the stacked HI signal, which we describe in Section 8. The modeling framework is used for parameter inference, and the results are presented in Section 9. We discuss the implications of our results for future HI intensity mapping surveys in Section 10, and conclude in Section 11. Throughout this paper, we assume a  $\Lambda$  cold dark matter cosmology with the values of the model parameters reported in Planck Collaboration et al. (2020).

## 2. The MeerKLASS *L*-band Deep-field Data

The MeerKLASS *L*-band deep-field data is observed from 41 observation blocks with a total of 62 hr of integration time per dish before flagging, spanning across  $236 \text{ deg}^2$  in terms of sky area in R.A.  $\sim (330^\circ, 360^\circ)$  and decl.  $\sim (-36^\circ, -25^\circ)$ . The details of the scanning strategy and observation time can be found in Table A1 of MK25.

The *L*-band data is observed from 900–1670 MHz with a frequency channel width of  $\delta\nu = 208.984375 \text{ kHz}$  and a time resolution of  $\delta t = 2 \text{ s}$ . Noise diodes are fired to each receiver for 0.585 s once every 19.5 s for relative reference calibration at the level of time-ordered data (TOD). Before and after each scan of around 100 minutes, the telescope is pointed to track a nearby celestial point source, either PKS 1934-638 or Pictor A, as a bandpass and absolute flux calibrator.

The data is then calibrated using the calibration pipeline KATCALI.<sup>20</sup> The details of the calibration process are described in J. Wang et al. (2021), and we briefly summarize it below. An initial radio-frequency interference (RFI) flagging is applied to the raw data. Removing channels with strong RFI contamination leaves the frequency ranges 971–1075 MHz and 1305–1504 MHz to be considered (see Figure 5 of J. Wang et al. 2021). The data is then modeled as a combination of several time- and frequency-dependent components,

$$T_{\text{obs}}(t, \nu) = g(t, \nu)[T_{\text{ps}} + T_{\text{diffuse}} + T_{\text{el}} + T_{\text{diode}} + T_{\text{rec}}](t, \nu), \quad (1)$$

where  $t$  is observation time,  $\nu$  is the observing frequency,  $g(t, \nu)$  is the gain,  $T_{\text{ps}}$  is the brightness temperature of the calibrator,  $T_{\text{diffuse}}$  is the diffuse foreground emission,  $T_{\text{el}}$  is the elevation dependent terrestrial emission from the Earth’s atmosphere and ground spillover,  $T_{\text{diode}}$  is the noise diode signal, and the residual temperature from the receiver as well as modeling errors is absorbed into  $T_{\text{rec}}$ . The modeling of each component of brightness temperature is described in J. Wang et al. (2021). The residual temperature  $T_{\text{rec}}$  is assumed to be slowly varying in time, and its time dependence can be

described by a Legendre expansion up to the third-order Legendre polynomial.

The gain is then solved independently at each frequency for each feed of each dish using the reference model. The solution is assumed to be smooth in time, described by a Legendre expansion up to the fourth order. A Bayesian fitting framework is applied to solve for the gain and model temperature parameters. A more detailed description of the fitting and examination on the calibration quality can be found in J. Wang et al. (2021) and Section 2.4 of MK25.

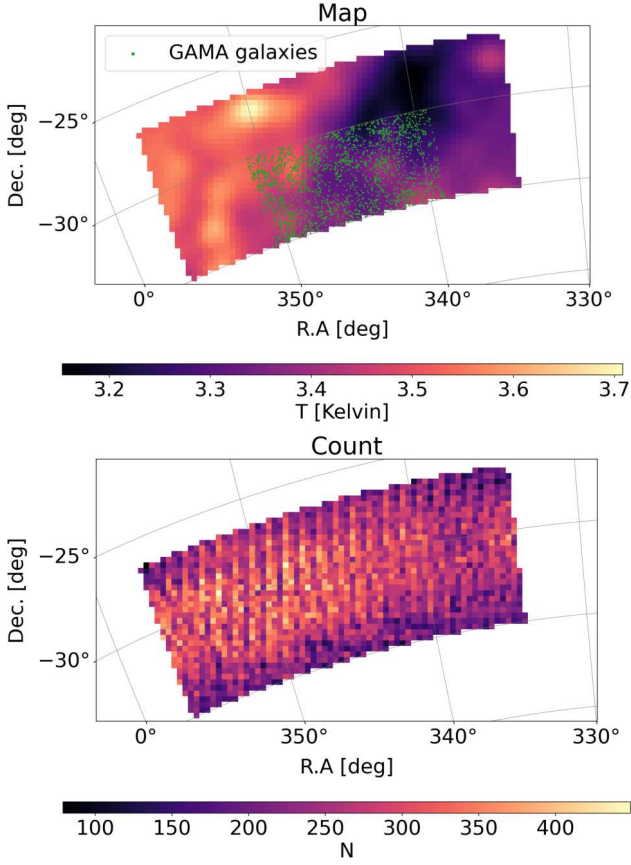
After applying the inverse of the gain solution, the data set is then flagged again to remove outliers along the frequency direction.  $T_{\text{el}}$  and  $T_{\text{rec}}$  are then subtracted from the TOD, which is subsequently converted from polarization to Stokes  $I$  intensity and gridded onto a sky map with azimuthal equal area projection. The angular size of the sky map pixels is chosen to be 0.3 deg. Another round of RFI flagging is then applied to the scans by comparing the TOD with the median value along the R.A. direction in the sky map. The TOD after flagging are then averaged again into the final sky map.

In Equation (1), the model components are mostly fixed from external measurements, for example, using the Python Sky Model (B. Thorne et al. 2017) for synchrotron emission in the modeling of the diffuse foregrounds. Any mismatch is then absorbed into  $T_{\text{rec}}$  and may induce nontrivial calibration errors. To further improve the gain solutions, the sky map obtained from the reference calibration is then passed back to Equation (1) to replace the original sky model of  $T_{\text{diffuse}}$ . The calibration process is then repeated to obtain an updated gain solution, which is then used to produce a new iteration of the final sky map. This self-calibration step is iterated 5 times for convergence to produce the intensity maps used for subsequent data analysis. Finally, by examining of flagging percentage of the data in each channel, the relatively clean sub-band of  $971.15 \text{ MHz} < \nu < 1023.61 \text{ MHz}$  is chosen for the data analysis, corresponding to  $0.39 \lesssim z \lesssim 0.46$  with an effective redshift  $z = 0.424$  at the center of the frequency sub-band. The frequency-averaged sky map after calibration, as reported in MK25, is shown in Figure 1 for illustration. For reference, the hit counts, i.e., number of time stamps averaged in each pixel, are also shown.

To perform the stacking analysis, we use an overlapping spectroscopic galaxy catalog from the Galaxy and Mass Assembly (GAMA) survey (S. P. Driver et al. 2009, 2011; J. Liske et al. 2015). The specific region we use in this paper is the 23 hr (G23) field (S. P. Driver et al. 2022) covering the area of  $339^\circ < \text{R.A.} < 351^\circ$  and  $-35^\circ < \text{decl.} < -30^\circ$ , as shown in the upper panel of Figure 1. The galaxy sample has a magnitude limit of  $i < 19.2$  (for target selection, see I. K. Baldry et al. 2010; J. Liske et al. 2015).

We note that, the GAMA Data Release 2 (J. Liske et al. 2015) only provides complete survey information and value-added catalogs for part of the total survey region, namely the G09, G12, and G15 regions. For simplicity, we assume that the G23 region has uniform survey geometry and constant redshift distribution in the redshift range of our interest, following MK25. For reference, we show the redshift distribution of the GAMA G23 catalog in Figure 2. As shown, the redshift bin  $0.39 \lesssim z \lesssim 0.46$  is at the tail of the redshift distribution. The galaxy catalog is therefore likely to be HI incomplete, with only the most-massive star-forming galaxies being selected. In total,  $N_g^{\text{GAMA}} = 2269$  galaxies are selected, resulting in a

<sup>20</sup> <https://github.com/meerclass/katcali>



**Figure 1.** Top panel: the frequency-averaged intensity map from the MeerKLASS  $L$ -band deep-field data reported in MK25. The maps are trimmed so that only pixels within  $334^\circ < \text{R.A.} < 357^\circ$  and  $-35^\circ < \text{decl.} < -26.5^\circ$  are kept. The green dots denote the positions of the overlapping GAMA galaxies within the redshift bin of the intensity mapping data. Bottom panel: the time-stamp counts of each pixel in the intensity map.

comoving number density of  $\sim 2.0 \times 10^{-4} \text{ Mpc}^{-3}$ . As the H I mass correlates with the stellar mass (e.g., H. Guo et al. 2021), it is expected that only some of the massive H I galaxies are present in the catalog, with a larger fraction of H I sources missing. The interpretation of stacking onto an incomplete catalog is discussed later in Section 4.

### 3. Simulation of Stacked H I Signal

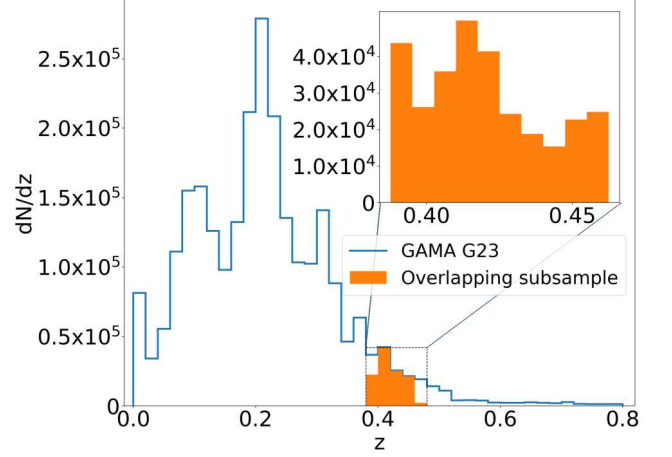
In this section, we describe the simulation pipeline used in this paper. The simulation pipeline validates the viability of detecting a stacked signal using the MeerKLASS  $L$ -band data, which will be presented in Section 4, and also instructs us on the optimal way of performing the stacking analysis later in Section 6.

#### 3.1. H I Emission

For an H I galaxy with total H I mass  $M_{\text{HI}}$ , the total 21 cm flux of the source is (M. Meyer et al. 2017)

$$S_{\text{HI}} = \frac{3h_P \nu_{21} A_{21}}{16\pi D_L^2} M_{\text{HI}}, \quad (2)$$

where  $h_P$  is the Planck constant,  $\nu_{21}$  is the rest frequency of the 21 cm line,  $A_{21}$  is the spontaneous emission rate of the H I atoms, and  $D_L$  is the luminosity distance of the source.



**Figure 2.** The redshift distribution of the GAMA G23 catalog. The blue solid line denotes the distribution for the entire catalog, whereas the orange filled region denotes the distribution of the subsample that overlaps with the intensity mapping data. The top-right corner shows the zoom-in of the overlapping redshift bin. Note that the values are different in the zoomed-in plot due to different choices of binning width in redshift.

Throughout this section, we assume an effective redshift  $z = 0.424$  when calculating background quantities such as  $D_L$  and the matter power spectrum discussed later.

The flux is distributed across the frequencies into an emission-line profile, due to the inner velocity dispersion of the source. The emission-line profile of the H I flux density can be described by a busy function (T. Westmeier et al. 2014)

$$F_\nu(\Delta\nu) = \frac{a_\nu}{2} \times (\text{erf}[b_\nu(w_\nu^2 - \Delta\nu^2)] + 1) \times (c_\nu \Delta\nu^2 + 1), \quad (3)$$

where “erf” denotes the Gauss error function,  $a_\nu$ ,  $b_\nu$ ,  $c_\nu$ ,  $w_\nu$  are parameters of the busy function,  $\Delta\nu = \nu - \nu_0$  is the difference between the observed frequency  $\nu$  and the central frequency  $\nu_0$  of the H I emission of the source, and the profile  $F_\nu$  is the observed flux density of the source so that  $\int d\Delta\nu F_\nu = S_{\text{HI}}$ . The observed flux density is discretized at each frequency channel. For a given busy function profile of source  $i$ , the flux density at a specific observing frequency  $\nu$  is

$$I_\nu^i(\nu - \nu_i) = \frac{1}{\delta\nu} \int_{\nu - \delta\nu/2}^{\nu + \delta\nu/2} d\nu' F_\nu(\nu' - \nu_i), \quad (4)$$

where  $\nu_i$  is the center frequency of the H I profile for source  $i$ .

The busy function encodes the velocity width of the H I galaxy. In particular,  $w_\nu$  controls the positions of the double peaks of the busy function and effectively describes the overall width of the profile. If the emission-line profile has a width of  $w_\nu$  in the observed spectrum, the corresponding velocity width in the source rest frame is

$$w_V = \frac{c}{\nu_{\text{obs}}} w_\nu, \quad (5)$$

where  $c$  is the speed of light, and  $\nu_{\text{obs}}$  is the observing frequency.

We note that Equation (3) is a simplified version of the busy function, with the emission-line profile being symmetric along  $\Delta\nu = 0$ . In our case of stacking, the emission-line profiles are averaged across many sources, which will symmetrize the underlying H I signal. Furthermore, as we discuss later in

Section 4, the effect of clustering greatly stretches the stacked spectrum and wipes out information on the velocity width. As a result, the effect of assuming a simplified busy function is negligible. For the same reason, we also do not consider the peculiar velocity of the observer and the impact of no Doppler correction on the map data.

Throughout this paper, HI galaxies are treated as point sources in the angular plane, as the intensity maps are of  $\sim 1^\circ$  resolution. The observed 21 cm intensity can then be described as

$$I(l, m, \nu) = B(l, m, \nu) \otimes \sum_i [\delta_D(\mathbf{l} - \mathbf{l}_i) I_i^i(\nu - \nu_i)], \quad (6)$$

where  $\mathbf{l} = (l, m)$  is the position on the sky,  $B(l, m, \nu)$  is the beam of the instrument,  $\otimes$  denotes convolution along the angular plane,  $i$  loops over each source, and  $\delta_D$  is the Dirac delta function. Equation (6) indicates that the simulation of HI signal requires the position, the HI mass, and the velocity width of the sources. We describe the routine for simulating each of these ingredients as follows.

We first determine the number density of HI galaxies. Given a population of HI galaxies, the distribution of the HI mass, the HI mass function (HIMF), can be described by a Schechter function (P. Schechter 1976)

$$\begin{aligned} \phi(M_{\text{HI}}) &= \frac{dn}{d\log_{10}(M_{\text{HI}})} \\ &= \ln(10) \phi_* \left( \frac{M_{\text{HI}}}{M_*} \right)^{\alpha+1} \exp \left[ -\frac{M_{\text{HI}}}{M_*} \right], \end{aligned} \quad (7)$$

where  $n$  is the number density of HI galaxies, and  $(\phi_*, M_*, \alpha)$  are the HIMF parameters.

In this paper, we adopt the values reported in M. G. Jones et al. (2018) at  $z \sim 0$ . It is expected that the values of the HIMF parameters evolve slowly over redshift (H. Xi et al. 2021; A. Gogate 2022; A. A. Ponomareva et al. 2023), although the evolution may become significant for our data at  $z \sim 0.4$  (A. Bera et al. 2022). Since the HIMF is not well understood beyond the local Universe, we resort to using the values at  $z \sim 0$ . Note that future measurements of HI stacking with Bayesian inference may help probe the HIMF over large samples of galaxies (H. Pan et al. 2020; J. Wang et al. 2025). For a given HIMF, we can calculate the number density  $\bar{n}_{\text{HI}}$ , and average density  $\bar{\rho}_{\text{HI}}$  of HI sources that have HI mass larger than  $M_{\text{HI}}^{\min}$  as

$$\bar{n}_{\text{HI}} = \int_{M_{\text{HI}}^{\min}} d\log_{10}(M_{\text{HI}}) \phi(M_{\text{HI}}), \quad (8)$$

$$\bar{\rho}_{\text{HI}} = \int_{M_{\text{HI}}^{\min}} d\log_{10}(M_{\text{HI}}) M_{\text{HI}} \phi(M_{\text{HI}}). \quad (9)$$

We find that for the HIMF parameter values in M. G. Jones et al. (2018), the HI density  $\bar{\rho}_{\text{HI}}$  is 97% complete for  $M_{\text{HI}}^{\min} = 10^8 M_\odot$ , which we choose as the lower limit. This gives a number density of  $\bar{n}_{\text{HI}} = 0.031 \text{ Mpc}^{-3}$  (note that this is much larger than the GAMA galaxy number density discussed in Section 2).

Second, we generate the clustering of galaxy positions by simulating a lognormal realization of the galaxy overdensity

field,  $\delta_g$ , following a model power spectrum,

$$\begin{aligned} P_g(\mathbf{k}) &= V |\tilde{\delta}_g(\mathbf{k})|^2 \\ &= b_g^2 (1 + f\mu^2/b_g)^2 P_m(\mathbf{k}) \end{aligned} \quad (10)$$

where  $V$  is the survey volume,  $\tilde{\delta}_g(\mathbf{k})$  is the galaxy overdensity in Fourier space,  $b_g$  is the galaxy bias,  $f$  is the growth rate,  $\mu = k_{\parallel}/|k|$ , and  $P_m$  is the matter power spectrum in real space. We choose  $b_g = 1.9$  matching the auto-power spectrum of the GAMA galaxies as discussed in MK25. Note that redshift space distortions are applied to the galaxy power spectrum using the Kaiser effect (N. Kaiser 1987) without the Finger-of-God effect, since the velocity dispersion is included in the simulation of the HI profile.

The nonlinear matter power spectrum is calculated using CAMB (A. Lewis & A. Challinor 2011) with HALOFIT (R. E. Smith et al. 2003; R. Takahashi et al. 2012). A lognormal realization of the galaxy overdensity field is then generated using POWERBOX (S. G. Murray 2018) based on the formalism described in F. Beutler et al. (2011). The lognormal simulation is motivated by its similarities to a Gaussian distribution as well as desirable physical properties, such as ensuring that the overdensity is always larger than  $-1$  (P. Coles & B. Jones 1991). The covariance of lognormal simulations is found to be more accurate than Gaussian realizations (e.g., S. Hilbert et al. 2011), and is sufficient for our signal-to-noise ratio in the data. The overdensity field is then converted to the number density field,

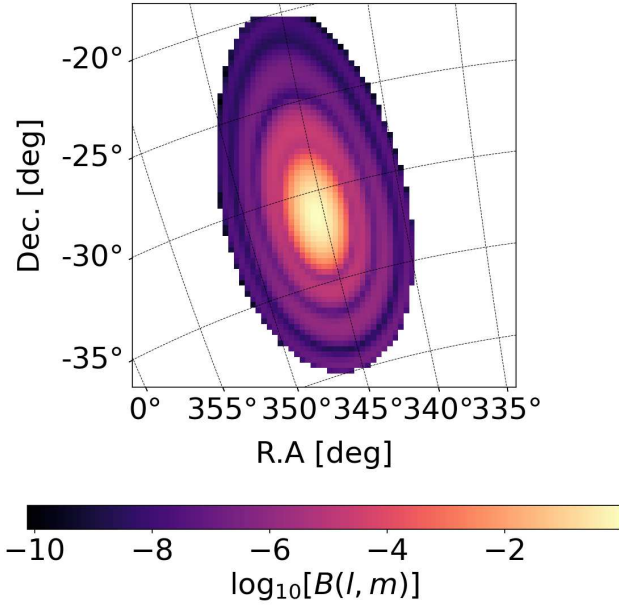
$$n_g(\mathbf{x}) = \bar{n}_{\text{HI}} (\delta_g(\mathbf{x}) + 1) W(\mathbf{x}), \quad (11)$$

where the survey selection function  $W(\mathbf{x})$  is 1 inside the 21 cm survey volume and 0 otherwise in our case. The galaxy positions are then Poisson sampled and projected onto the sky. Note that the comoving box used for generating galaxy positions is larger than the survey volume, as the beam smoothing requires information outside the survey area.

Third, we sample the input HIMF to assign HI mass to each galaxy. Note that the random assignment of HI mass is uncorrelated with the galaxy positions. Therefore, the bias of the HI map, as well as the bias of the stacking subsample described later, is the same as the input  $b_g = 1.9$ . The width of the emission profile is then calculated by first converting the HI mass to a velocity dispersion through the Tully–Fisher (T-F) relation (R. B. Tully & J. R. Fisher 1977). We choose the slope and the intercept of the T-F relation to be the measured values in A. A. Ponomareva et al. (2021). The intrinsic velocity dispersion is that projected to the line-of-sight direction by assuming a random inclination angle,  $\theta_{\text{incl}}$ , so that

$$w_\nu = \sin(\theta_{\text{incl}}) v_{\text{T-F}}, \quad (12)$$

where  $v_{\text{T-F}}$  is the velocity calculated from the T-F relation, and we randomly sample the inclination from a uniform distribution  $\theta_{\text{incl}} \in [0, 2\pi]$  for each mock galaxy. The velocity dispersion  $w_\nu$  is then converted to  $w_\nu$  in frequency according to Equation (5). For the other two busy function parameters  $b_\nu$  and  $c_\nu$ , we assume wide flat priors, so that  $b_\nu \in [0.01, 1] \text{ km}^{-2} \text{ s}^2$  and  $c_\nu \in [0.001, 0.01] \text{ km}^{-2} \text{ s}^2$ , and sample random values of  $b$  and  $c$  for each source. Note that the amplitude parameter  $a_\nu$  is not a free parameter and is instead set by the shape of the profile and the total HI mass of the galaxy. The emission-line profile can then be calculated

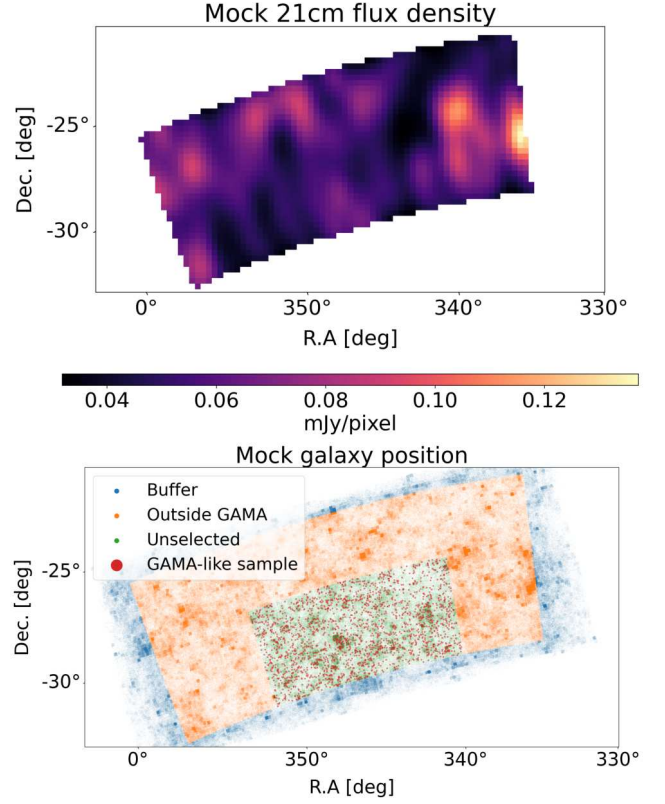


**Figure 3.** The EIDOS primary beam model (K. M. B. Asad et al. 2021) for the MeerKAT telescope. For the visualization, the beam is pointed toward the center of the survey area (R.A., decl.) = (345°:4, -31°:4) at the center of the frequency sub-band  $\nu = 997.38$  MHz. Note that the apparent ellipticity and orientation of the beam are due to projecting the sky coordinates onto the Cartesian grids. The angular grid lines are shown as black dotted lines to further illustrate the projection.

according to Equations (2) and (3). As we show later in Section 4, the stacked signal is dominated by the effects of double counting and clustering, and the shape of the emission-line profile has negligible effects in our modeling.

Finally, based on the emission-line profiles and the galaxy positions, an intensity map can be generated as described in Equation (6). We use EIDOS, which is based on the astro-holographic measurement of the MeerKAT beam (K. M. B. Asad et al. 2021), to generate the input beam model  $B(l, m, \nu)$  for the convolution that we show in Figure 3. An illustration of one realization of the simulated HI intensity maps is shown in Figure 4.

After generating the HI galaxies, we still need to select the subsample in the mock that resembles the GAMA galaxy catalog for the stacking. As mentioned in Section 2, the galaxy data catalog is incomplete, with little information on derived quantities to be used for simulating the survey selection. We assume the GAMA galaxy catalog to contain the most-massive HI galaxies, and we select the  $N_g^{\text{GAMA}} = 2269$  most-massive mock galaxies inside the GAMA survey area to be the mock subsample. The process is illustrated in Figure 4. Note that we can also use the number density of the data catalog to recalculate a lower limit  $M_{\text{HI}}^{\text{min,GAMA}}$  by inverting Equation (8), and randomly select among the mock HI galaxies with HI mass larger than the limit. However, given the extremely small number density of the data, we find the lower limit  $M_{\text{HI}}^{\text{min,GAMA}} = 10^{10.235} M_\odot$ , at the very tail of the HIMF distribution. If we choose among mock galaxies with masses larger than  $M_{\text{HI}}^{\text{min,GAMA}}$ , the number of selected galaxies will fluctuate a lot around the expected  $N_g^{\text{GAMA}}$  due to sampling at the tail of the HIMF distribution. To compensate for the low number of galaxies in the mock stacking subsample, we then still have to select the most-massive mock galaxies below



**Figure 4.** Upper panel: the frequency-averaged flux density map of mock HI signal for one realization. The flux density is shown in the unit of  $\text{Jy pixel}^{-1}$ , where the pixel area is  $0.3 \times 0.3 \text{ deg}^2$ . Lower panel: the mock galaxy positions in one realization. Blue dots show galaxies that are generated in the comoving box that are outside the MeerKLASS survey area (“Buffer”). They are included in the HI signal simulation as required by the beam smoothing. Orange dots show galaxies that are inside the MeerKLASS survey area but outside the GAMA region (“Outside GAMA”). Green dots show the galaxies inside the GAMA region that have relatively small HI mass and are not included in the subsample for stacking (“Unselected”). The red dots show the galaxies selected for stacking (“GAMA-like sample”). Note the high number density of galaxies results in pixelated clumps seen in the panel. They correspond to the resolution over which the galaxy overdensity is generated, which is much smaller than the resolution of the 21 cm maps.

$M_{\text{HI}}^{\text{min,GAMA}}$ , effectively performing the same selection as simply choosing the  $N_g^{\text{GAMA}}$  most-massive mock galaxies.

As we discuss later, in the model fitting, we use the simplified assumption that the HI emission comes entirely from the galaxy catalog, and all galaxies in the catalog have the same HI mass. This is because in the data analysis, we have no knowledge of the underlying distribution of HI galaxies outside the GAMA galaxy sample. To mimic this lack of information in the mock, for each realization, we generate an alternative HI simulation following the simplified assumption. In each realization of the mock, we use the original simulation of mock HI galaxies, and calculate the total HI mass within the GAMA survey region. We then exclude all other galaxies from the simulations except the GAMA-like mock subsample (see Figure 4). For the GAMA-like subsample, we then assign the same HI mass to each galaxy and keep the total HI mass within the GAMA region the same, so that

$$M_{\text{HI}}^{\text{GAMA,alt}} = \sum_i M_{\text{HI}}^i / N_g^{\text{GAMA}}, \quad (13)$$

where  $i$  iterates over all mock galaxies in the original simulation that is inside the GAMA survey area. We then use the updated mock galaxy catalog and the HI mass to reassign random velocity profiles. An alternative mock HI map is then generated.

Comparing the original simulation with the full galaxy sample following an HIMF distribution and the simulation based on the simplified assumption,<sup>21</sup> we can quantify the biasing of the stacked signal from the simplified assumption in Section 4, which we then use to forward model the signal for inference in Section 8.

The mock simulations are generated with multiple realizations. Foregrounds and thermal noise are then added to the mock HI signal, which we discuss for the rest of this section.

### 3.2. Foregrounds

In 21 cm experiments, foregrounds originate from Galactic and extragalactic radio sources that have continuous spectra in frequency. For the  $L$ -band observation with  $\sim 1^\circ$  angular resolution, it is expected that Galactic synchrotron radiation will dominate the foreground emission. For simplicity, we only consider synchrotron in the simulation. For the actual data, foreground removal is performed and validated by the null tests described in MK25, and further validation tests are presented later in Sections 6.1 and 6.2. The foreground signal is therefore not included in the forward modeling for parameter fitting, and instead, the effects of PCA from the actual data are propagated into the mock HI as we describe later in Section 8. Here, foreground simulation is only used in the mock to examine the effect of PCA foreground removal on the stacked spectrum qualitatively.

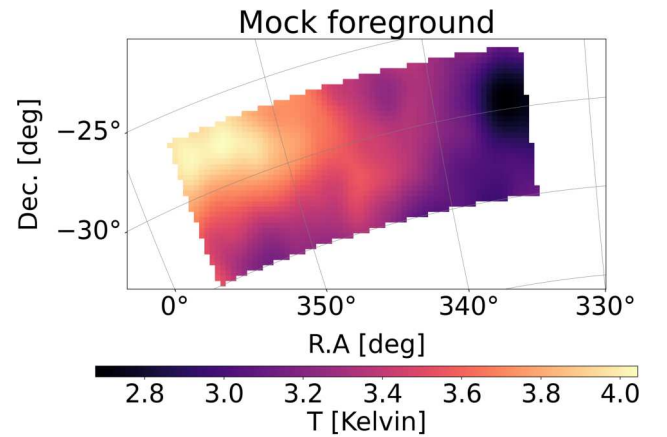
We use the improved version of the 408 MHz measurement of the synchrotron map (C. G. T. Haslam et al. 1982; M. Remazeilles et al. 2015) as the template. The synchrotron map is then extrapolated to the observing frequencies by assigning spectral indices to each pixel. The spectral indices are calculated based on the sky model at 1.4 GHz and 2.3 GHz provided in the Global Sky Model (H. Zheng et al. 2017), based on the observations of P. Reich et al. (2001) and J. L. Jonas et al. (1998). The resulting synchrotron maps are then smoothed by the primary beam using the EIDOS model. An illustration of the mock foreground map is shown in Figure 5.

Note that, the assumption of a fixed spectral index for each pixel on the sky ignores the secondary curvature of the synchrotron spectrum (e.g., M. O. Irfan et al. 2022). Consistency between low-frequency and  $L$ -band measurements of Galactic foregrounds is also contested (M. J. Wilensky et al. 2025). In fact, this can be seen by visually comparing the maps shown in Figures 1 and 5. As mentioned, a more realistic foreground simulation is outside the scope of this work.

### 3.3. Thermal Noise

The radiometer equation states that for Stokes  $I$  intensity at map level, the standard deviation of the thermal noise

<sup>21</sup> From now on, the ‘‘simplified assumption’’ always refers to the scenario where we only consider 21 cm signals from the galaxy catalog with the same HI mass for all galaxies in the forward modeling, ignoring the incompleteness and the mass distribution.



**Figure 5.** The frequency-averaged brightness temperature map for the mock foreground signal.

fluctuation is

$$\sigma_N = \frac{T_{\text{sys}}(\nu)}{\sqrt{2\delta\nu\delta t N_{\text{hits}}}}, \quad (14)$$

where  $\delta\nu$  is the frequency resolution,  $\delta t$  is the time resolution,  $N_{\text{hits}}$  is the number of time stamps in the pixel, and, following MK25, we model the temperature of the system as

$$T_{\text{sys}} = T_{\text{rx}} + T_{\text{el}} + T_{\text{CMB}} + T_{\text{gal}}, \quad (15)$$

where  $T_{\text{rx}}$  is the receiver temperature,  $T_{\text{el}}$  is the spillover introduced in Equation (1),  $T_{\text{CMB}}$  is the CMB temperature, and  $T_{\text{gal}}$  is the temperature of Galactic synchrotron. For the narrow frequency sub-band of our data, the system temperature is found to be near constant  $T_{\text{sys}} \approx 16$  K (see MK25). In reality, the measured fluctuation from residual temperature in the calibration model is typically higher than expected. We enlarge the system temperature by a factor of 1.2 so that  $T_{\text{sys}} = 19.2$  K, following MK25. The thermal noise is then generated randomly following a Gaussian distribution at each pixel.

The different components of the mock observation are then added together to produce the mock intensity maps. For now, we leave out the component of the systematics caused by beam ripple as seen in the data, which will be discussed in detail later in Section 7.

## 4. Viability of the Stacking Measurement

In this section, we use the mock observations to validate the fact that the MeerKLASS  $L$ -band intensity maps can be used to achieve a stacking detection. From the mock stacked cubes, we can average the signal into the angular plane or into a spectrum, to inform us on the optimal region around a source that goes into the averaging for optimal signal-to-noise ratio, which can be used for the data. We then discuss how the signal can be modeled and how the covariance of the stacked signal can be estimated. All results in this section are averaged over 100 realizations.

### 4.1. Double Counting in Stacked Cubelets

To investigate the validity of the stacking measurement, we need to first define how a stacked signal is obtained from the

intensity maps. The stacked signal is calculated as

$$I(\Delta\alpha, \Delta\phi, \Delta\nu) = \frac{\sum_i I(\alpha_i + \Delta\alpha, \phi_i + \Delta\phi, \nu_i + \Delta\nu) w_i}{\sum_i w_i}, \quad (16)$$

where  $(\alpha_i, \phi_i)$  are the R.A. and decl. of the map pixel in which the  $i$ th galaxy resides,  $I$  is the flux density, and  $w_i = w_{\text{HI}}(\alpha_i + \Delta\alpha, \phi_i + \Delta\phi, \nu_i + \Delta\nu)$  is the weight of each pixel.  $(\Delta\alpha, \Delta\phi, \Delta\nu)$  is the position of the stacked signal relative to the source position in terms of R.A., decl., and frequency. Throughout this paper, we adopt inverse noise variance weighting so that  $w_i = N_{\text{hits}}^i$ , as shown in the lower panel of Figure 1.

It is common to express the stacked signal in terms of velocity instead of frequency. In our work, since we do not consider the Doppler correction, the peculiar velocity of the observer is not accounted for. Therefore, there is an ambiguity between the frequency offset  $\Delta\nu$  and the velocity it should correspond to, so the stacked signal is not in actual velocity units. Nevertheless, when showing the results along the spectral direction, we also express  $\Delta\nu$  in an effective velocity unit so that<sup>22</sup>

$$v = \frac{c}{\nu_{\text{obs}}} \Delta\nu, \quad (17)$$

where  $\nu_{\text{obs}} = 997.38$  MHz is the central frequency of the frequency sub-band.

The frequency offset  $\Delta\nu$  can also be transformed in to an approximate comoving scale,  $\Delta D_c$ , along the line of sight so that

$$\Delta D_c \approx \frac{c \nu_{21}}{H\nu_{\text{obs}}^2} \Delta\nu, \quad (18)$$

where  $H$  is the Hubble parameter at the observing redshift. In our frequency sub-band, a frequency offset of  $\Delta\nu = 1$  MHz corresponds to  $\Delta D_c \approx 5$  Mpc.

Equation (16) states that the stacked signal is a weighted average of the nearby volume of each source position. We refer to the final stacked signal as the stacked cube, and the contribution of each source as the stacked cubelet. This expression also can be understood as equivalent to a cross-correlation function between the galaxy positions and the HI line-intensity fluctuations.

The stacked cube of flux density can then be summed along the frequency direction into integrated flux so that

$$\bar{F}(\Delta\alpha, \Delta\phi) = \sum_{\Delta\nu_i} I(\Delta\alpha, \Delta\phi, \Delta\nu_i) W(\Delta\nu_i) \delta\nu, \quad (19)$$

where  $\delta\nu$  is again the frequency channel bandwidth,  $\Delta\nu_i$  loops over each frequency interval, and  $W(\Delta\nu_i)$  is the selection function. In this paper, we choose

$$W(\Delta\nu) = 1, \text{ for } |\Delta\nu| < 3.5 \text{ MHz}, \quad (20)$$

which corresponds to  $|v| \lesssim 1000 \text{ km s}^{-1}$ . The choice of summing within the  $\pm 1000 \text{ km s}^{-1}$  range is justified in Section 4.3, and we note that the stacked image is only used for visual checks in the data analysis later.

As we will show in Section 4.2, the structure of the HI emission in the stacked image follows closely the shape of the beam. It is therefore also useful to check the polar average of the stacked image, so that

$$\begin{aligned} \bar{F}_{1\text{D}}(\Delta\theta_i) &= \sum_{\Delta\alpha, \Delta\phi} \bar{F}(\Delta\alpha, \Delta\phi) W(\sqrt{\Delta\alpha^2 + \Delta\phi^2}; \Delta\theta_i) \\ &/ \sum_{\Delta\alpha, \Delta\phi} W(\sqrt{\Delta\alpha^2 + \Delta\phi^2}; \Delta\theta_i), \end{aligned} \quad (21)$$

where  $W(\sqrt{\Delta\alpha^2 + \Delta\phi^2}; \Delta\theta_i)$  is a selection function that selects pixels that are inside the  $i$ th angular annulus bin. In this work, we choose seven annulus bins to be linearly spaced between 0 and 3.5 deg. We can also compute the polar average of the primary beam, by simply replacing the stacked image  $\bar{F}(\Delta\alpha, \Delta\phi)$  with the frequency-averaged primary beam  $B(l, m)$  in Equation (21).

Alternatively, the stacked cube can be summed along the angular plane into a spectrum so that

$$\bar{I}(\Delta\nu) = \sum_{i,j} I(\Delta\alpha_i, \Delta\phi_j, \Delta\nu) W(\Delta\alpha_i, \Delta\phi_j), \quad (22)$$

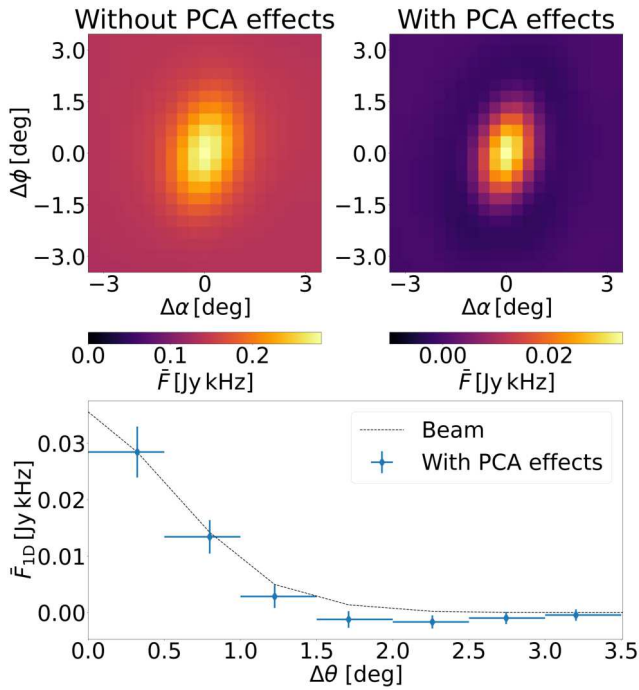
where  $i, j$  loops over each pixel along the angular plane, and  $W(\Delta\alpha_i, \Delta\phi_j)$  is the selection function. Since Equation (16) normalizes over the sum of weights, the stacked spectrum can be understood as the HI signal for the average galaxy. In this paper, for visualization, we consider the range ( $|\Delta\alpha| < 3^\circ$ ,  $|\Delta\phi| < 3^\circ$ ), and for the spectrum, we consider

$$W(\Delta\alpha_i, \Delta\phi_j) = 1, \text{ for } \sqrt{\Delta\alpha^2 + \Delta\phi^2} < 1.2^\circ, \quad (23)$$

based on the mock discussed later in Section 4.3. Furthermore, in the stacked spectrum, we choose the binning in frequency to be 3 times the frequency channel bandwidth.

It is worth pointing out that, at the angular scales of the MeerKAT beam  $\sim 1^\circ$  corresponding to  $\sim 30$  Mpc, it is not evident that a stacked signal can be measured. A simple way to see this is to imagine two extreme scenarios: one where the beam is uniform across the sky, and one where the beam is infinitely small. If the beam is uniform across the full sky, for any arbitrary position as the center of the stacking, the stacked cubelet will always simply return the average HI flux. If the beam is infinitely small, for stacking on the positions of HI galaxies, the stacked cubelet will give an excess signal at the center angular pixel and no signal anywhere else. In general, due to beam smoothing, the stacked cube will always contain an excess signal at the center of the cubelet from the source we are stacking on, and an extra signal from double counting the other sources near the target stacking source. In terms of a two-point correlation function, the effect of the beam can be understood as a mixture between the one-halo and two-halo terms. ‘‘Two-halo term’’ corresponds to the correlation between HI galaxies that are not in the same parent dark matter halo, whereas the one-halo term describes the inner-halo correlation between HI galaxies of the same halo (see A. Cooray & R. Sheth 2002 for a review and, e.g., Z. Chen et al. 2021 for HI halo models). A cubelet at a given separation to the detected galaxy, which would contain its HI signal and therefore correspond to the one-halo term, would also include signal from a different, close-by galaxy, which is a clustering signal. The  $\sim 30$  Mpc scale of the beam means that the mixture

<sup>22</sup> We express velocity in terms of  $v$  instead of  $\Delta\nu$ , since it is difficult to visually distinguish  $\Delta\nu$  and frequency offset  $\Delta\nu$ .

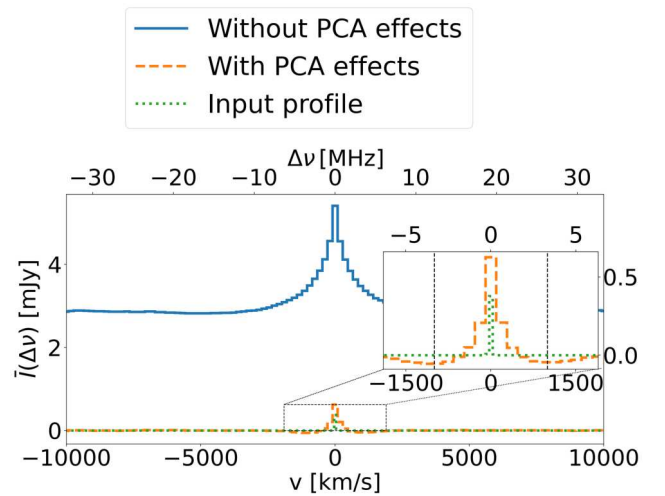


**Figure 6.** The angular stacking signal from H I only mocks (no foregrounds and thermal noise), averaged over 100 realizations. The stacked image is calculated according to Equation (19) and averaged into a 1D profile according to Equation (21). The upper-left panel shows the stacked image without any PCA cleaning effects, and the upper-right panel shows the case with PCA cleaning of 10 modes. The excess signal at the center corresponds to the primary beam. The apparent ellipticity and orientation of the excess signal around the center are due to projection, as explained in Figure 3. Note that the pixels away from the center region also have a nonzero emission in the left panel. In the H I only case, the first 10 eigenmodes contain much more H I signal, as there is no foreground to be removed, and therefore, the signal loss is more severe. The lower panel shows the 1D polar average of the stacked image for the PCA cleaned case (“With PCA effects”). The error bars are the standard deviations of the stacked profile among the realizations. For illustration, the primary beam profile matching the amplitude of the first  $\Delta\theta$  bin is plotted in black dashed line (“Beam”).

happens at small inner-halo scales and out to large physical scales, corresponding to the two-halo correlation.

We illustrate this effect of double counting, using the angular stacking signal following Equation (19). As shown in the left panel of Figure 6, the stacked H I flux shows a consistent positive signal at  $\sim 0.15$  Jy kHz level, regardless of how far away the angular position is from the center. This indicates a severe amount of double counting from the beam smoothing. Despite the double counting, the excess emission around the center of the image is still clearly visible, and its profile corresponds to the shape of the beam. This suggests that stacking the intensity maps onto the GAMA galaxy positions produces a detectable signal.

The double counting is further demonstrated in Figure 7. The H I signal spectrum (“Without PCA effects”) shows a plateau of emission-line signal, indicating the existence of double counting. Comparing the H I signal spectrum with the average input H I profile of the galaxy subsample, one can see that in addition to a signal plateau, the peak around the center frequency is also amplified with a greatly enlarged velocity width. If the excess signal is only a result of beam smoothing and summing over the angular plane, there is no correlation along the frequency direction. The widened velocity profile of the stacked spectrum suggests that the stacked cubelets are



**Figure 7.** The spectral stacking signal from H I only mocks, averaged over 100 realizations. The stacked spectrum is calculated according to Equation (22). The stacked H I signal without any PCA cleaning (“Without PCA effects”), with PCA cleaning of 10 modes (“With PCA effects”), and the average of the input H I profiles of the GAMA-like galaxies (“Input profile”) are shown. The zoomed-in panel shows the stacked spectrum after PCA for  $|\Delta\nu| < 1600$  km s $^{-1}$ , compared against the average of the input H I galaxies. The stacked spectrum is much larger than the input profile due to the beam smoothing, which causes double counting of H I sources. The black dashed lines show the  $|\Delta\nu| = 1000$  km s $^{-1}$  boundary within which the stacked angular image is produced. In the H I only case with no foreground and noise, the first 10 eigenmodes of PCA contain much more H I signal, as there is no foreground to be removed, and therefore, the signal loss is more severe compared to the full mock case shown later in Figure 9.

correlated at these frequency scales. This points to the fact that, the H I signal is correlated at  $\Delta\nu < 5$  MHz frequency intervals and  $\sim 30$  Mpc transverse scales. The  $\Delta\nu < 5$  MHz frequency scales correspond to a line-of-sight comoving distance of  $\lesssim 25$  Mpc. This is an effect of clustering of the H I signal. Similar features of clustering in the stacked signal have also been found in the context of CO line-intensity mapping (D. A. Dunne et al. 2025).

We emphasize that the effects of double counting and clustering are distinctive. The double counting is due to the high number density of H I emitters compared to the angular size of the beam. One can imagine a case where there is no clustering, and the number density of H I sources is extremely low within the size of the primary beam. In that case, along the line of sight of any source, there will be very little leakage from other sources, as they are sufficiently away. Hence, there will not be a positive plateau in the stacked signal. On the other hand, even if the average number density of H I sources is low, there will always be pairs of sources close to each other if there is clustering. Therefore, the widening of the stacked spectrum at small separations will be present as a result of the clustering.

#### 4.2. Effects of PCA

We have established that for the MeerKLASS  $L$ -band intensity maps, a stacking signal can be measured with the caveat of the stacked signal having a large contribution due to beam smoothing and source clustering. The additional contribution is then partially removed due to the procedure of foreground cleaning, which removes the mean of the signal and causes signal loss. Therefore, it is important to validate

that the excess signal in the central region of the stacked cube is robust against the PCA.

To perform the PCA cleaning, we first calculate the weighted average of the maps at each channel, and obtain the mean centered intensity mapping data matrix  $\mathbf{M}_{ip}$ , where  $i$  iterates over each frequency and  $p$  iterates over each map pixel. We can then compute the frequency–frequency covariance matrix so that

$$C_{ij} = \left( \sum_p w_{ip} w_{jp} \mathbf{M}_{ip} \mathbf{M}_{jp} \right) / \left( \sum_p w_{ip} w_{jp} \right). \quad (24)$$

Eigendecomposition of the frequency–frequency covariance matrix is then performed to find its eigenvalues and eigenvectors. The eigenvectors  $\mathbf{u}_i$  are then sorted based on the values of their corresponding eigenvalues from largest to smallest. Choosing an  $N_{\text{fg}}$  number of modes, we can define a mixing matrix  $\mathbf{A}$  so that

$$\mathbf{A} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{N_{\text{fg}}}], \quad (25)$$

from which we can then define a cleaning matrix  $\mathbf{R}$

$$\mathbf{R} = \mathbf{I} - \mathbf{A} \mathbf{A}^T, \quad (26)$$

where  $\mathbf{I}$  is the identity matrix, and  $^T$  denotes transpose of a matrix. The residual data matrix after cleaning is then

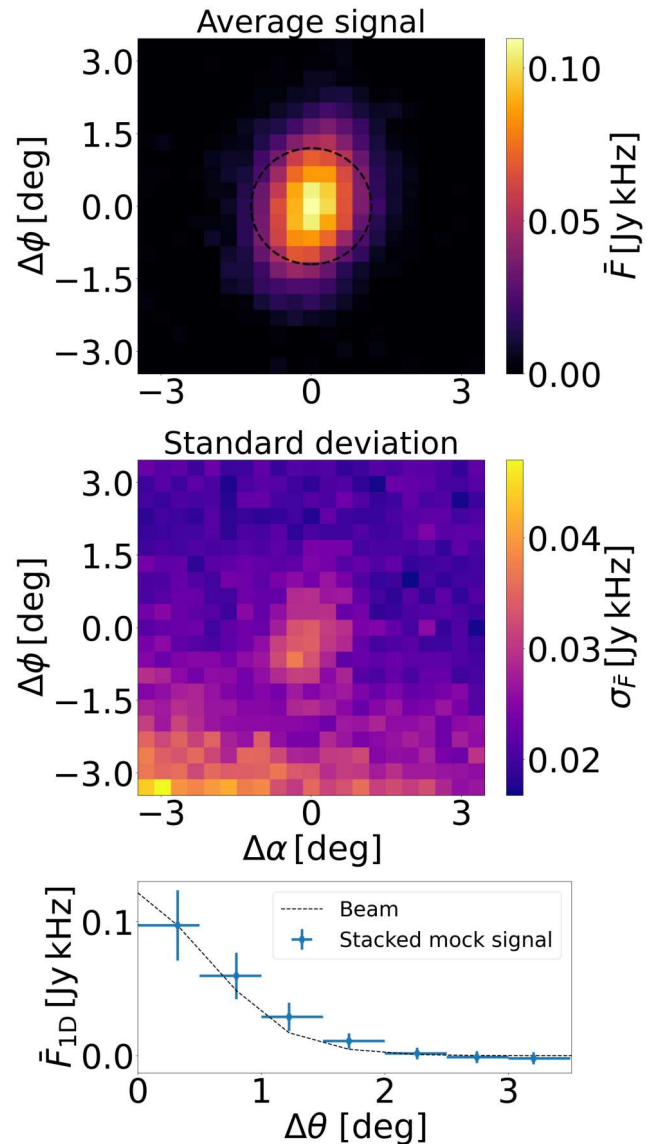
$$\mathbf{M}_{ip}^{\text{res}} = \sum_j \mathbf{R}_{ij} \mathbf{M}_{jp}. \quad (27)$$

The residual data  $\mathbf{M}^{\text{res}}$  can then be used to perform the stacking analysis.

The choice of the number of modes to be removed,  $N_{\text{fg}}$ , is subtle and depends on the level of contamination in the data (see, e.g., S. Cunnington et al. 2023a). Throughout this paper, we choose  $N_{\text{fg}} = 10$  as suggested in MK25. This choice is to be consistent with the level of foreground removal needed in the data, as discussed in MK25.

We first discuss the case where only H I signal is considered. We present the angular stacking image for H I only mocks after PCA cleaning in the right panel of Figure 6. As can be seen, the excess signal in the central region of the stacked image is still present after PCA cleaning. Significant signal loss can be found, as is evident in the decrease in the amplitude of the emission signal with respect to the case before PCA cleaning. The region of excess emission still follows closely the shape of the primary beam, suggesting that the excess is indeed caused by the source in the GAMA-like subsample (recall Figure 4). In the bottom panel of Figure 6, we show the polar average of the stacked image after PCA cleaning according to Equation (21). As we can see, within the primary beam of  $\sim 1^\circ$ , the excess emission closely follows the beam. The emission then decreases as the angular distance increases and reaches zero.

The removal of the extra emission away from the cubelet center results in the removal of the plateau in the stacked spectrum, as shown in the orange dashed line (“With PCA effects”) in Figure 7. For large  $\Delta\nu$ , the signal is consistent with zero. The amplitude of the central peak is much smaller compared to the spectrum before PCA, while still being much larger than the input. The PCA cleaning also results in a slightly negative amplitude at  $|\Delta\nu| \sim 1000 \text{ km s}^{-1}$ . The removal of relatively large line-of-sight modes, together with the subtraction of the mean, results in the negative amplitude



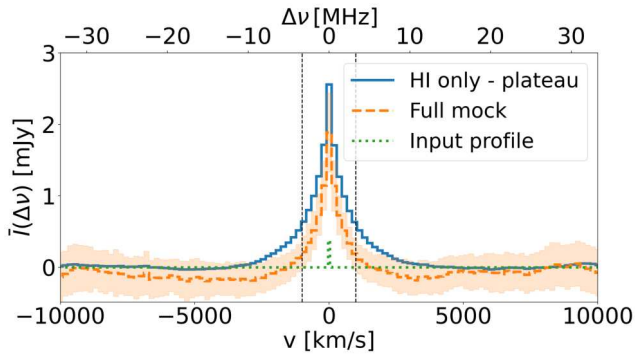
**Figure 8.** The angular stacking signal from full mocks (including foreground and noise simulation), averaged over 100 realizations. The top panel shows the average of the stacked images across the realizations, and the central panel shows the standard deviation of the realizations. The dashed circle in the top panel denotes the  $1.2^\circ$  boundary within which the signal is summed into the stacked spectrum. Note the differences in the color scale in the two panels. The lower panel shows the 1D polar average of the stacked image for the full mock (“Stacked mock signal”). The error bars are the standard deviations of the stacked profile among the realizations. For illustration, the primary beam profile matching the amplitude of the first  $\Delta\theta$  bin is plotted in black dashed line (“Beam”).

(see also Figure 18 of CHIME Collaboration et al. 2023). The complication of these effects requires forward modeling to describe.

#### 4.3. Expected Detection Significance

Based on the viability of detecting a stacked signal around the central region of the stacked cube, we then include foreground signal and thermal noise and perform the PCA subtraction to see if the excess emission can be detected given the depth of the MeerKLASS  $L$ -band survey.

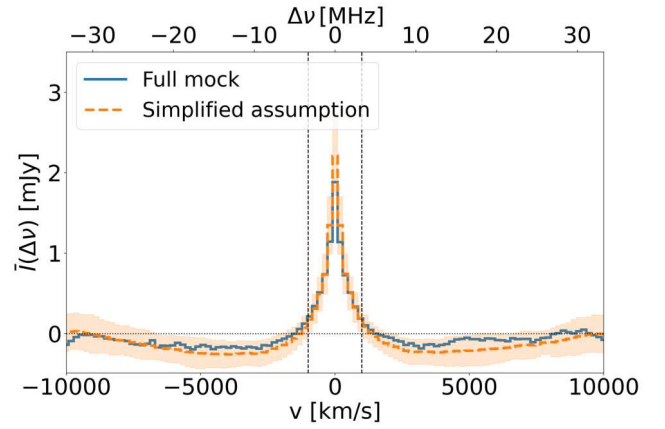
In the top panel of Figure 8, we show the angular stacked image averaged from 100 independent realizations. Comparing



**Figure 9.** The spectral stacking signal from full mocks, averaged over 100 realizations. The H I only signal without any PCA cleaning is shown in the blue solid line, with the plateau from double counting subtracted (“HI only—plateau”). The averaged spectrum of the full mocks with PCA cleaning of 10 modes is shown in the yellow dashed line. The shaded region corresponds to the standard deviation among the realizations. The average of the input H I profiles of the GAMA-like galaxies (“Input profile”) is also shown for reference.

the stacked image with the one in Figure 6, we can see that the excess signal remains, with the background becoming noisy due to the presence of thermal noise. Notably, the signal loss is much less severe. This is because in the H I only case, the first 10 eigenmodes contain much more H I signal, as there is no foreground to be removed. When foregrounds and thermal noise are included, the PCA first cleans the smooth foregrounds, and the same number of modes  $N_{\text{fg}} = 10$  removes much less H I signal. The mocks can also be used to calculate the signal variance by taking the standard deviation among the realizations, as we show in the bottom panel of Figure 8. Overall, the stacked signal is around a factor of 3 larger than the standard deviation, suggesting that the stacked image can indeed be detected. Note that, the bottom half, and especially the bottom-left part, of the stacked image has a higher level of variance. This can be traced back to the survey area shown in Figure 1. When stacking near the boundary of the GAMA region, the upper half of the cubelet is always sampled by the 21 cm intensity maps, whereas the region lower than the GAMA galaxies is not covered by the MeerKLASS survey area. The bottom half of the stacked image is therefore less sampled and has a higher noise level. The polar average of the stacked image, shown in the bottom panel of Figure 8, is consistent with the beam profile. We find that for the stacked image, the region within  $1.2^\circ$  of the center of the cube has a clear excess, which is the criterion we use for calculating the stacked spectrum in Equation (23).

We then proceed to calculate the stacked spectrum as shown in Figure 9. To compare against the signal without PCA, we show the H I only simulation without PCA cleaning, and subtract out the plateau from the double counting (“HI only—plateau”) for comparison. The averaged stacked spectrum (“Full mock”) shows similar amplitude compared to the input, with a visible amount of signal loss due to PCA cleaning of foregrounds. Again, the stacked spectrum contains a large component of beam smoothing and clustering; it is much larger than the input H I profile. Overall, it is expected that the central peak can be detected with high statistical significance  $>3\sigma$ . The peak of the stacked spectrum extends across  $|\Delta\nu| \lesssim 1000 \text{ km s}^{-1}$ , and we therefore choose  $1000 \text{ km s}^{-1}$  as the upper limit for the frequency channels in calculating the stacked image in Equation (20).



**Figure 10.** The stacked spectrum of the H I signal from simplified assumption (“Simplified assumption”), compared against the true H I signal (“Full mock”). The results shown are averaged across 100 realizations. The shaded region around the dashed line shows the standard deviation for the simplified assumption case among the realizations.

The overall detection significance of the stacked signal is harder to calculate than simply comparing the variance with the signal. This is because the signal is correlated between different angular positions and frequencies, as one pixel in the intensity map will be averaged into different voxels in the stacked cube. To quantify the overall significance, the full covariance matrix is needed, which will be discussed in Section 5.

#### 4.4. Biasing from Simplified Forward Modeling

As mentioned in Section 3, in the actual data analysis, we would not have information on the underlying distribution of the H I mass of the sources. The detection in our data, while statistically significant, as shown later in Section 6, is not enough to allow for a large number of parameters to be constrained. Therefore, as a simplification, we need to assume that all H I density resides within the GAMA galaxy subsample, and the galaxies have the same H I mass when using forward modeling to model the signal. This leads to a mismatch between the interpreted model and the underlying truth. We examine the impact of this biasing using the mocks with the following steps.

The alternative H I map, described in Section 3.1, is applied with the *original* PCA cleaning matrix from the full simulation. This is because in the data analysis, the PCA cleaning matrix is obtained based on the data itself to ensure the same level of signal loss in the data as well as forward modeling. The cleaned H I map is then passed to the stacking pipeline to generate the stacked cube. The resulting stacked signal is the “forward modeling” case, since the simplified assumption is applied to the simulation while keeping the overall H I density unchanged. For simplicity, we only show the stacked spectrum, since later in Section 5 we demonstrate that the stacked spectrum is more robust for covariance estimation.

The comparison between the original mock and the simplified assumption is shown in Figure 10. The simplified assumption closely follows the true H I signal, with the differences between the two smaller than the standard deviation across the realizations. The H I signal from the simplified assumption has a higher central peak with a slightly

narrower width, due to far fewer sources having much higher HI mass than the full mock. Integrating the spectrum between  $|v| < 2500 \text{ km s}^{-1}$  where the peak resides, we find that the differences in the integrated flux are smaller than 5%. We deduce that, due to the large physical scale of the beam, the clustering signature of the HI flux density in the simplified assumption is similar to the full mock. Even though there are far fewer sources resulting in a more extreme distribution of HI mass, after beam smoothing, the aggregate HI signal over a resolution element is similar.

In conclusion, the simplified assumption only slightly biases the modeled HI signal, and is sufficiently accurate for the purpose of this work. We note that, in this case, the inferred HI mass per source is not the actual mass, but should be interpreted as the total HI mass over the number of galaxies in the stacking subsample.

## 5. Covariance Estimation

### 5.1. Mock Covariance

The covariance of the stacking measurement consists of two uncorrelated components, which are the signal covariance and the noise covariance. Understanding the contribution of both components is necessary to fully quantify the detection significance and perform model inference. In this section, we first use the realizations to calculate the covariance, and then discuss the method for covariance estimation from one realization, which is then used later for the data analysis.

To calculate the covariance of the stacked cube, in each realization, we start from the PCA cleaning matrix in Equation (26), and apply the cleaning separately to the HI signal, foregrounds, and noise. We find that the residual foregrounds are negligible compared to the level of the HI signal. The residual HI signal and noise are then used to calculate the stacked cube, respectively, producing 100 realizations of stacked cubes for both components. The covariance can then be calculated from the realizations,

$$C_{ij}^{\text{mock}} = \sum_n \frac{(I_i^n - \bar{I}_i)(I_j^n - \bar{I}_j)}{N_r - 1}, \quad (28)$$

where  $i, j$  denote two voxels in the stacked cube,  $n$  iterates over the realizations,  $I_i^n = I^n(\Delta\alpha_i, \Delta\phi_i, \Delta\nu_i)$  is the stacked signal for one realization,  $\bar{I}$  is the average of the stacked cube across the realizations, and  $N_r$  is the number of mock realizations. For any covariance matrix  $C$ , we can also compute the correlation matrix,

$$\text{corr}_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}}\sqrt{C_{jj}}}, \quad (29)$$

which informs us on the correlation between different data points of a measurement.

The covariance obtained from Equation (28) is for the 3D stacked cube. In practice, we are only interested in the stacked image and the stacked spectrum. The covariance for the averaged image/spectrum can be calculated by simply substituting the 3D signal  $I^n$  with the averaged signal in Equation (28).

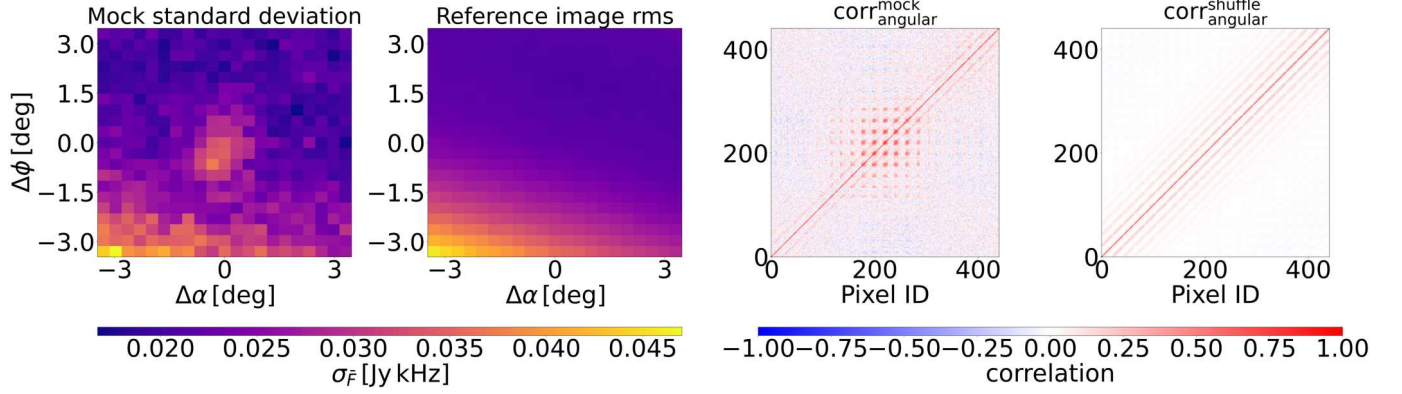
### 5.2. Covariance Estimation Using Random Shuffling

In reality, we only have access to the data without much information on the underlying model. The ‘‘true covariance’’ from Equation (28), built from specific choices of model parameters, does not necessarily reflect the observation. Moreover, the observation data contain multiplicative systematics that affect the covariance of the data, which we will discuss in detail later in Section 7. We therefore need a method of covariance estimation from the data itself, which we can then compare against the true covariance using the mocks. In this work, we explore using random shuffles of galaxy positions for covariance estimation. We generate realizations of random galaxy positions following the same procedure and clustering statistics of the galaxy catalog, but based on realizations of mock dark matter uncorrelated with the HI signal. Such a random shuffling of galaxy positions should have the same statistical significance as the true catalog, but produce a stacking signal consistent with zero. Therefore, it is common to use such galaxy shuffling as a null detection test in the data analysis, as shown in MK25. In this work, we refer to the stacked image/spectrum using a random shuffle as the ‘‘reference image/spectrum.’’ A covariance can be then calculated based on multiple realizations of the shuffling *in a single mock*,

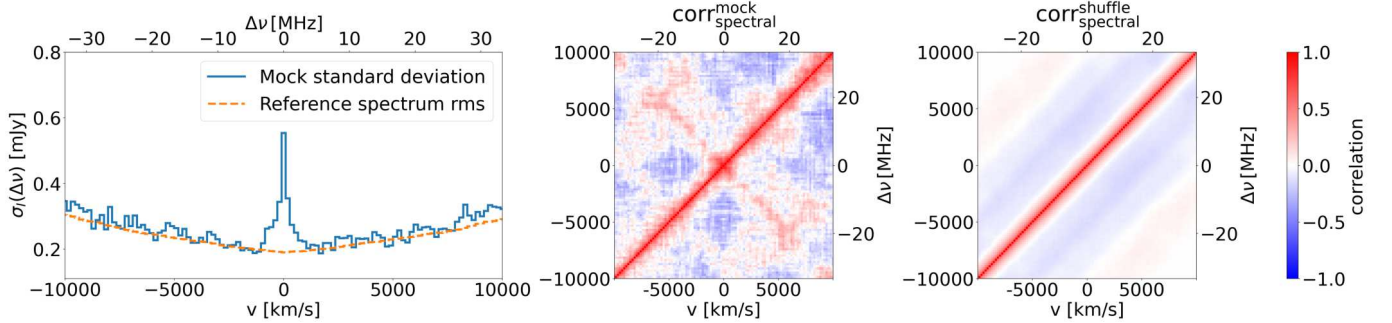
$$\hat{C}_{ij}^{\text{shuffle}} = \sum_n \frac{(I_i^{n,\text{shuffle}} - \bar{I}_i^{\text{shuffle}})(I_j^{n,\text{shuffle}} - \bar{I}_j^{\text{shuffle}})}{N_{\text{shuffle}} - 1}, \quad (30)$$

where  $i, j$  denote two voxels in the stacked cube,  $n$  iterates over the realizations of the shuffling of galaxy positions,  $I^{n,\text{shuffle}}$  is the stacked signal over the shuffled galaxy positions for one shuffle,  $\bar{I}^{\text{shuffle}}$  is the averaged stacked signal over all shuffles, and  $N_{\text{shuffle}}$  is the number of shuffling in total. In this work, we choose  $N_{\text{shuffle}} = 400$  and find that convergence has been reached. From Equation (30), it is also easy to see that the standard deviation of the reference stacked signal is simply the diagonal elements of  $\hat{C}^{\text{shuffle}}$ . Using the 400 realizations of shuffled galaxy positions, we can then estimate the covariance for each mock observation, and compute the average of the estimated covariance across all mock realizations. We denote the average of  $\hat{C}^{\text{shuffle}}$  across all mock realizations as  $\bar{C}^{\text{shuffle}}$ . For simplicity, we do not show the results for the entire 3D cube, but instead average the signal into the stacked image and the stacked spectrum for examination.

The comparison between the true mock covariance and the random shuffling estimate is shown in Figure 11. From the variance of the angular stacking shown in the left panels of Figure 11, we can see that the random shuffling captures accurately the overall noise level of the signal. In particular, the higher noise variance at the lower half of the image, which is due to less sampling as discussed in Section 4.3, is also reflected in the variance of the reference image. The random shuffling does not capture the HI signal variance at the center of the image though, which is expected, since the shuffling does not contain the excess HI signal at the center of the stacked cube. Apart from the lack of signal variance, it is also expected that the correlation of the signal due to the primary beam is also missing from the shuffling estimate. In the right panels of Figure 11, we can see that the true correlation matrix contains strong correlation between pixels near the center of



**Figure 11.** Left panel: the standard deviation of the angular stacked image across the mock realizations, which is also the square root of the diagonal elements of the mock covariance matrix  $C_{\text{angular}}^{\text{mock}}$ . Center-left panel: the reference image standard deviation among the random shuffles, averaged across all mock realizations, which is also the square root of the diagonal elements of the estimated covariance from shuffling  $C_{\text{angular}}^{\text{shuffle}}$ . Center-right panel: the correlation matrix from mock covariance. The indexing of the array iterates from left to right, and then bottom to top of the stacked image. Right panel: the correlation matrix from covariance estimation using shuffling.



**Figure 12.** Left panel: the blue solid line shows the standard deviation of the spectral stacked spectrum across the mock realizations (“Mock standard deviation”), which is also the square root of the diagonal elements of the mock covariance matrix  $C_{\text{spectral}}^{\text{mock}}$ . The orange dashed line shows the average of the standard deviation of the randomly shuffled reference spectrum across the realizations (“Reference spectrum rms”). Center-right panel: The correlation matrix from mock covariance. Right panel: the correlation matrix from covariance estimation using random shuffling.

the stacked image. This correlation is not being captured by the random shuffling.

We then examine the differences between the true mock covariance and the shuffling estimate in the stacked spectrum in Figure 12. The overall amplitude of the variance from the reference spectrum matches closely the true mock variance, as seen in the left panel. Similar to the stacked image, the shuffling estimate does not capture the variance of the HI signal near the center  $\Delta\nu \sim 0$ . Comparing the correlation between different velocities shown in the right panel of Figure 12, we can see that the correlation between nearby channels is largely consistent between the truth and the shuffling estimate. The largest inconsistency comes from the antidiagonal direction, i.e., correlation between positive and negative values of  $\Delta\nu$ . This inconsistency again stems from signal covariance, as the same HI sources are averaged into the spectrum multiple times at different  $\Delta\nu$  due to the double counting. The PCA cleaning, while removing the amplitude of the plateau as discussed in Section 4.2, does not fully remove the correlation induced by the plateau.

In conclusion, we find that the random shuffling can be used to estimate the overall amplitude of the variance of the stacked signal. Due to not including the excess HI signal at the center of the stacked cube, the shuffling estimate does not include the HI signal covariance nor its corresponding correlation near the center of the stacked image. In the stacked spectrum, the

correlation is largely consistent, but the HI signal variance around the center is still missing from the shuffling estimate. From now on, we focus on the stacked spectrum for covariance estimation.

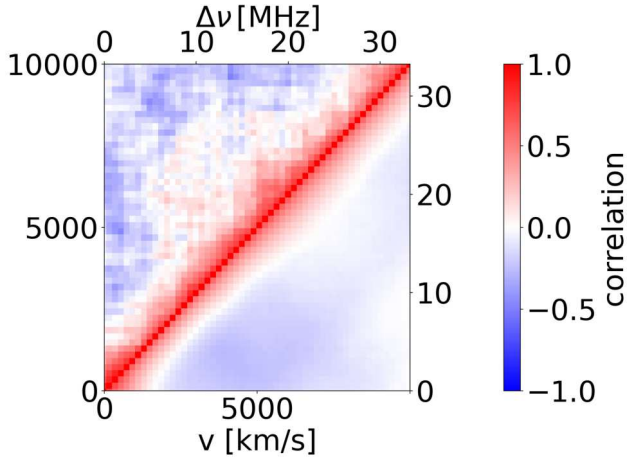
### 5.3. Mock-corrected Covariance Estimation

Based on the conclusions reached in Section 5.2, we can see that an accurate covariance estimation can be obtained in the stacked spectrum, where we have summed over the angular pixels. There are two problems that remain, which are the missing HI signal variance around  $\Delta\nu \sim 0$ , and the fact that the antidiagonal direction of the correlation matrix is not captured by the shuffling estimate.

We first discuss how to mitigate the problem of antidiagonal correlation. Note that, if only the upper-right quadrant of the correlation matrix is considered, i.e.,  $\Delta\nu \geq 0$ , then the antidiagonal direction is naturally excluded. This prompts the usage of symmetrized stacking (F. Sinigaglia et al. 2022a) so that

$$\bar{I}_{\text{sym}}(\Delta\nu) = (\bar{I}(\Delta\nu) + \bar{I}(-\Delta\nu))/2, \quad (31)$$

where for the averaged spectrum  $\bar{I}_{\text{sym}}$  at  $\Delta\nu$ , both the  $\Delta\nu$  and  $-\Delta\nu$  of the original stacked spectrum are included. Since the stacked HI signal is on average symmetric along  $\Delta\nu = 0$  as seen in Figure 9, we expect no loss of information from the



**Figure 13.** The comparison between the correlation matrix of the symmetrized stacked spectrum using the true mock covariance and the shuffling estimate. The upper triangle of the matrix shows values from the true mock covariance, and the lower triangle shows values from the shuffling.

symmetrization. In our case where the angular plane is collapsed into the spectrum, Equation (31) is equivalent to the  $180^\circ$  rotational symmetry along the zero-velocity axis described in F. Sinigaglia et al. (2022a). In our case, this does not increase the signal-to-noise ratio, as suggested in F. Sinigaglia et al. (2022a), since in our case, there is significant double counting, and there is no meaningful increase in the number of sampling at each  $\Delta\nu$  from symmetrization.

Using the symmetrized stacked spectrum, we recalculate the mock stacked signal in all realizations, the mock covariance, and the shuffling estimate. The resulting correlation matrix is shown in Figure 13. Since the spectrum has been symmetrized, we only need to consider the range  $\Delta\nu \geq 0$ . Comparing the results from Figures 13 and 12, one can see that the correlation matrix from the shuffling estimate becomes more consistent with the underlying truth when the spectrum is symmetrized. The correlation matrix is highly consistent for small values of  $|\Delta\nu|$ , where small inconsistencies still persist at intermediate values.

We now turn to the fact that we need to correct for the missing HI signal covariance. We construct a mock-informed correction of the covariance estimation so that

$$\hat{C}_{ij}^{\text{data}} = \mathbf{r}_i \mathbf{r}_j \hat{C}_{ij}^{\text{data,shuffle}}, \quad (32)$$

where  $\hat{C}^{\text{data,shuffle}}$  is the shuffling estimate described in Equation (30), using the actual data.  $\mathbf{r}$  is a vector of the ratio between the square root of the diagonal elements of mock covariance and the shuffling estimate in the mock so that

$$\mathbf{r}_i = \sqrt{C_{ii}^{\text{mock}} / \hat{C}_{ii}^{\text{mock,shuffle}}}. \quad (33)$$

It is easy to see that  $\hat{C}^{\text{data,shuffle}}$  and the final  $\hat{C}^{\text{data}}$  share the same correlation matrix. For the mock simulations, Equation (32) returns the true mock covariance when averaged over all realizations, since in this case, “data” in  $\hat{C}^{\text{data,shuffle}}$  simply stands for one realization of the mock.

In short, Equation (32) describes the covariance estimation from the stacked intensity maps in two steps: First, use the symmetrized stacking and randomly shuffled galaxy positions

to calculate an initial estimate  $\hat{C}_{ij}^{\text{data,shuffle}}$ . Second, use the mock realizations to obtain a correction of the amplitude from the shuffling to the true covariance. This method has several desired advantages. The amplitude of the covariance does not rely on the amplitude obtained in the mock, but instead depends on the data itself through  $\hat{C}^{\text{data,shuffle}}$ . The correlation matrix also follows the structure of the data. Even if there are systematics not considered in the mock, for example, the multiplicative systematics that we discuss in Section 7, the correlation matrix will follow the data affected by the systematics instead of following the mock, since the correlation matrix follows  $\hat{C}^{\text{data,shuffle}}$  (although the correlation will be slightly distorted; A simple analytical derivation is presented in Appendix). The downside of Equation (32) is that the fractional correction of missing HI covariance is based on the mock, where the amplitude of the HI signal and the noise are likely to be different from the actual data. We come back to the effect of covariance estimation later in Section 9.

## 6. Stacking Measurement

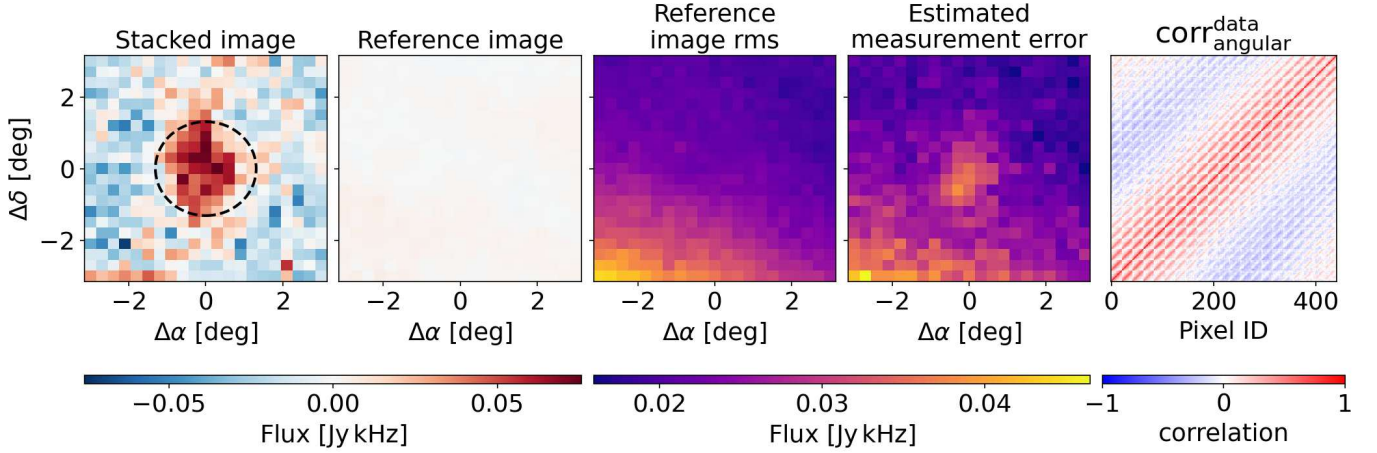
In this section, we present the measurement of the stacked signal using the MeerKLASS *L*-band deep-field intensity maps. The intensity maps are cleaned using PCA, removing 10 modes following MK25. We note the difference between the foreground cleaning in MK25 and this work is that we no longer perform the deconvolution (see MK25 and the discussion in I. P. Carucci et al. 2024). We will come back to the effect of deconvolution and its relation to the systematics in Section 7. We perform the stacked signal estimation following Equation (16), and then collapse the stacked cube into the stacked image and the stacked spectrum following Equations (19) and (22). The stacked spectrum is subsequently symmetrized. We then use random shuffling of galaxy positions to perform covariance estimation following Equation (32).

### 6.1. Stacked Image

In Figure 14, we present the stacked image of the MeerKLASS *L*-band deep-field data over the GAMA galaxies. The central area of the stacked image shows a clear excess of HI signal, corresponding to the structure of the primary beam blurred by the random thermal noise. The amplitude of the peak flux is  $\sim 0.7$  Jy kHz, slightly smaller than the expected signal level at  $\sim 1$  Jy kHz in our mock simulation shown in Section 4.3. This is a combination of the fact that our HI model in the mock simulation is not based on the  $z \sim 0.4$  redshift range, and that the data itself possess different foregrounds and systematics that change the PCA cleaning matrix with the same number of modes removed.

The reference image from random shuffling reveals an average that is consistent with zero. It suggests that there is no overall foreground residual in the data. Since foreground residuals are uncorrelated with the HI data, any foreground residual would have similar structures in the stacked image and in the reference image. The reference image, however, does not show any statistically significant structure of excess emission. The null test using the reference image suggests that the excess in the stacking signal indeed originates from the HI emission of the GAMA galaxies.

Using the random shuffling, we can estimate the covariance of the measurement. The resulting measurement errors and the correlation matrix are shown in Figure 14. The estimated



**Figure 14.** Left panel: the measured angular stacked image. The dashed circle shows the  $1.2^\circ$  boundary within which we use to calculate the stacked spectrum. Center-left panel: the average of the reference image across all random shuffles. Center panel: the reference image standard deviation among the random shuffles. Center-right panel: the measurement error at each pixel in the stacked image, which is also the square root of the diagonal elements of the estimated covariance matrix. Note that the color scale for the center and center-right panels are different from the color scale for the left- and center-left panels. Right panel: the estimated correlation matrix from mock covariance. The indexing of the array iterates from left to right, and then bottom to top of the stacked image.

measurement error is consistent with the mock variance we find in Figure 8. It also correctly reproduces the higher variance at the bottom half of the image as we see in the mock. The correlation matrix, on the other hand, is inconsistent with the mock, as can be seen by the comparison between the rightmost panels of Figures 14 and 11. In the mock simulation, the reference image is dominated by noise, which is not related to the primary beam, and therefore there is an inconsistency between the true mock covariance and the estimate as we have shown in Figure 11. In the data, however, the estimated covariance produces a strong correlation between the central pixels, seemingly suggesting a convolution between thermal noise and the primary beam, which is not possible. As we explain later in Section 7, this is due to the chromaticity of the beam affecting the PCA, which then is applied to the entire data vector, which affects the noise as well.

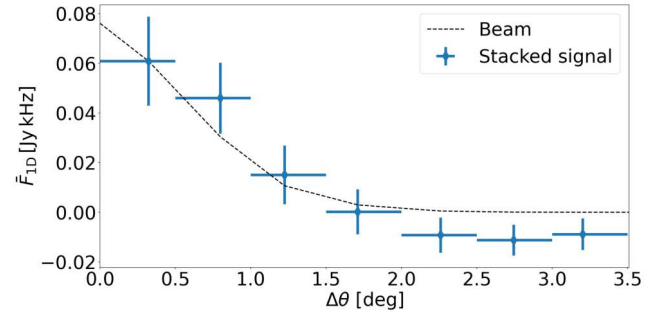
For reference, we calculate the 1D polar average of the stacked data image and show the results in Figure 15. The stacked emission follows the attenuation of the primary beam, with most of the detection significance lying within the  $1.2^\circ$  range. Therefore, we use the central  $(1.2 \text{ deg})^2$  area of the stacked image to calculate the detection significance. Given a data vector  $\mathbf{d}$  and its covariance matrix  $\mathbf{C}$ , the detection significance can be quantified as

$$\chi^2 = \mathbf{d}^\top \mathbf{C}^{-1} \mathbf{d}. \quad (34)$$

Using the central  $(1.2 \text{ deg})^2$  area and the estimated covariance, we find a detection significance of  $8.66\sigma$ . The high detection significance suggests that the HI signal is dominant compared to the noise level in the MeerKLASS  $L$ -band deep-field intensity maps.

## 6.2. Stacked Spectrum

In Figure 16, we present the stacked spectrum of the MeerKLASS  $L$ -band deep-field data over the GAMA galaxies. The unsymmetrized stacked spectrum in the upper panel shows clear detection of excess signal relative to the noisy background and the reference spectrum. The peak amplitude of the excess signal is at  $\sim 1$  mJy, which is around the expected level from the mock simulation as shown in

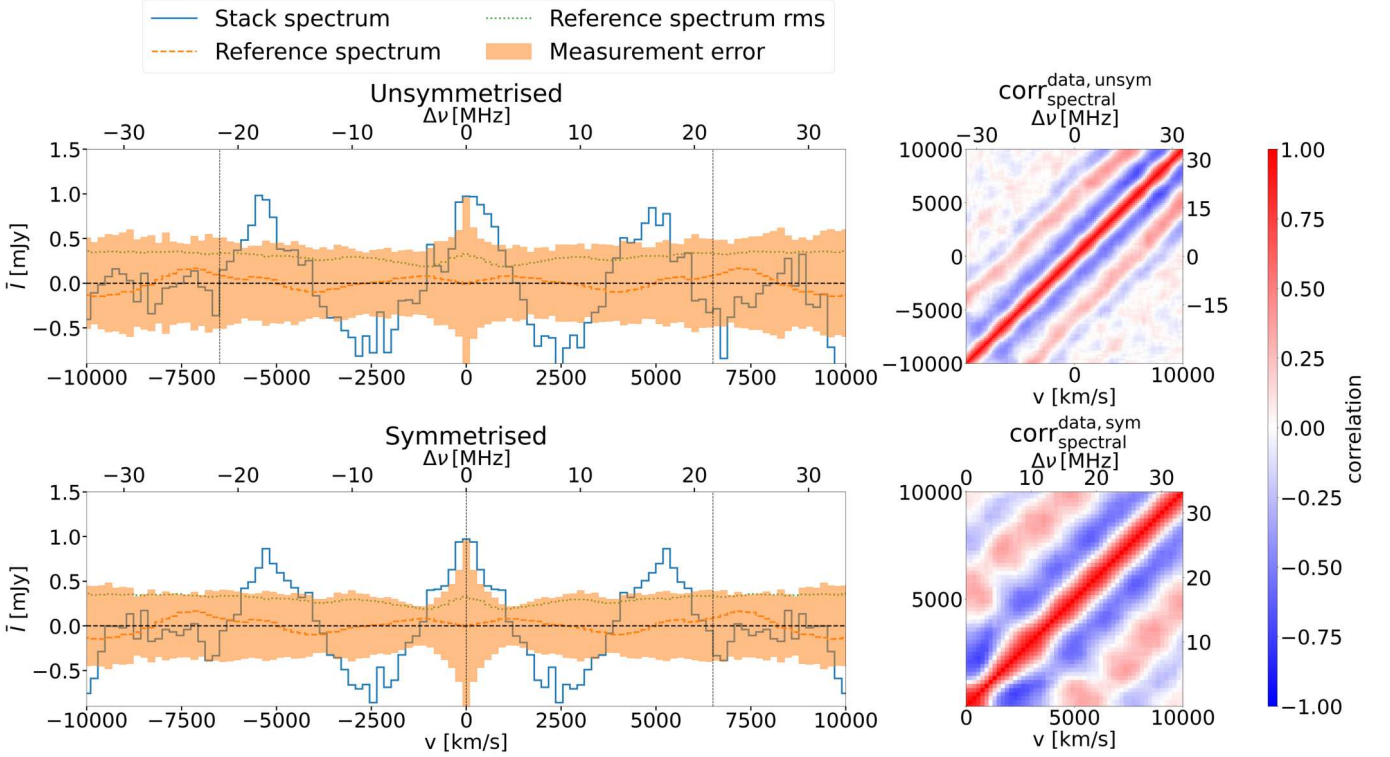


**Figure 15.** The 1D polar average of the stacked image (“Stacked signal”). The error bars are the standard deviations of the stacked profile among the realizations. For illustration, the primary beam profile matching the amplitude of the first  $\Delta\theta$  bin is plotted in black dashed line (“Beam”).

Figure 9. The width of the central peak is  $\sim 2000 \text{ km s}^{-1}$ , similar to the mock signal. However, the stacked spectrum also exhibits a clear structure of systematics, shown by the repeated peaks at  $\sim -5000 \text{ km s}^{-1}$  and  $\sim 5000 \text{ km s}^{-1}$  that are statistically significant against null detection. Correspondingly, at  $\sim \pm 2500 \text{ km s}^{-1}$  there are negative amplitude troughs in the spectrum, clearly reflecting an oscillating structure with an oscillating period of  $\sim 5000 \text{ km s}^{-1}$ , or equivalently  $\sim 15 \text{ MHz}$ .

The detection of the stacked spectrum indeed originates from the HI signal, as the reference spectrum is consistent with zero. The standard deviation of the stacked spectrum is at  $\sim 0.3\text{--}0.4 \text{ mJy}$ , also consistent with the noise level found in the mock simulation, as shown in Figure 12. Similar to the angular stacked image, we find that the estimated covariance matrix does not follow the structure seen in the mock simulation. Instead, the covariance also shows structures indicating the existence of a component of systematics.

As we have discussed in Section 5, in order to eliminate the need to evaluate correlations between positive and negative values of  $\Delta\nu$ , we need to use the symmetrized stacked spectrum, which we show in the lower panel of Figure 16. The symmetrized spectrum and the estimated covariance show similar oscillating structures as seen in the unsymmetrized spectrum.



**Figure 16.** Upper-left panel: the measurement of the stacked spectrum of the MeerKLASS  $L$ -band deep-field data over the GAMA galaxies (“Stacked spectrum”). The orange dashed line shows the average of the reference spectrum over the random shuffles (“Reference spectrum”). The green dotted line shows the standard deviation of the reference spectrum among the random shuffles (“Reference spectrum rms”). The shaded region shows the estimated measurement error (“Measurement error”), which is also the square root of the diagonal elements of the estimated covariance matrix. The vertical dashed lines show the  $|v| < 6500 \text{ km s}^{-1}$  boundary within which we use to calculate the detection significance. Upper-right panel: the estimated correlation matrix. Lower-left panel: the same as the upper-left panel, but the stacked spectrum is symmetrized. The vertical dashed lines show the  $0 < v < 6500 \text{ km s}^{-1}$  boundary within which we use to perform model inference. Lower-right panel: the estimated correlation matrix for the symmetrized spectrum.

Using the estimated covariance, we can quantify the detection significance. We restrict the stacked spectrum to  $|v| < 6500 \text{ km s}^{-1}$  for the unsymmetrized spectrum, as there is no signal outside this range, and the detection significance is found to be  $7.45\sigma$ . The significance is slightly lower than the one in the stacked image. It is worth noting that a large contribution to the significance is from the peaks and troughs outside the  $v \sim 0$  region from the oscillating systematics. For the symmetrized spectrum, we only need the  $v > 0$  region and apply the same  $v < 6500 \text{ km s}^{-1}$  cut. The detection significance is found to be  $5.29\sigma$ . The decrease is due to the averaging of the spectrum, which symmetrizes the spectrum. If we limit the velocity range to  $v < 1500 \text{ km s}^{-1}$  where only the central peak is included, the detection significance decreases to  $2.43\sigma$ . It suggests that a large part of the information contained in the stacked spectrum is on the systematics in the data.

The detection of the stacked spectrum indicates the depth of the MeerKLASS  $L$ -band deep-field data. The peak amplitude of the systematics at  $|v| \sim 5000 \text{ km s}^{-1}$  is similar to the HI signal at  $v \sim 0$ , suggesting that the HI signal and the effect of systematics are around the same order of magnitude. A similar conclusion can also be found in the auto-power spectrum of the data, as shown in Figure 14 of MK25. While the presence of the systematics is still large, it is no longer dominant in the observed signal. If we can understand the origin of the oscillating systematics, we can parameterize and model its effects, allowing for the inference of the systematics as well as the HI model.

## 7. Nature of Systematics

In this section, we demonstrate evidence that the oscillating systematics seen in the data are a convolutional effect caused by the diffraction of the secondary reflector of the MeerKAT telescope, which affects the chromaticity of the primary beam. We then investigate the parameterization and the modeling of the stacked HI signal.

### 7.1. Additive and Multiplicative Systematics

In general, for a data vector of summary statistics  $\mathbf{d}$ , the effect of systematics can be written as two components (e.g., N. Weaverdyck & D. Hutner 2021),

$$\mathbf{d} = \sum_i \mathbf{S}_M^i \mathbf{d}_i + \mathbf{S}_A, \quad (35)$$

where  $i = \{\text{HI}, \text{n}, \text{fg}\}$  represents different components of the signal including the HI, noise and foregrounds, respectively,  $\mathbf{d}_i$  is the underlying uncontaminated signal,  $\mathbf{S}_M^i$  is the multiplicative systematics matrix, and the additive systematics  $\mathbf{S}_A$  is added as an additional component.

The systematics have different origins for different tracers of the LSS. For example, in galaxy clustering surveys, the additive systematics can be induced by interlopers (e.g., A. R. Pullen et al. 2016), and the multiplicative systematics can be induced by source blending (e.g., P. Melchior et al. 2021). Intensity mapping surveys can be contaminated by a number of sources of systematics. Residual RFI contamination can be present in the data (B. N. Engelbrecht et al. 2025),

leading to additive systematics. Calibration errors due to insufficient modeling of the sky lead to multiplicative systematics as well as additive residual foreground leakage (N. Barry et al. 2016). It is therefore important to first determine the type of systematics that contributes the most to the MeerKLASS data.

We note that, since the foregrounds and RFI are not of cosmological origin, the additive systematics should not correlate with the positions of the GAMA galaxies. If additive systematics have significant contributions to the stacked signal, the oscillating features seen in Figure 16 should be present both in the signal and the reference spectrum. This is not the case, as the reference spectrum and the reference image are consistent with null detection. Therefore, the oscillating systematics must be a multiplicative component applied to the signal data vector.

### 7.2. Evidence of Convolutional Systematics

The nature of the multiplicative systematics can be split into two categories. Effects such as bandpass errors are multiplied to the data vector at each pixel without convolving the signal. On the other hand, effects such as beam chromaticity convolve the data vector in the angular plane and also create structures in the spectral direction. The distinction between the two effects can be seen in the covariance of the data,

$$\begin{aligned} \mathbf{C} &= \langle \mathbf{d} \mathbf{d}^T \rangle = \sum_i \mathbf{S}_M^i \langle \mathbf{d}_i \mathbf{d}_i^T \rangle (\mathbf{S}_M^i)^T \\ &= \sum_i \mathbf{S}_M^i \mathbf{C}_i (\mathbf{S}_M^i)^T, \end{aligned} \quad (36)$$

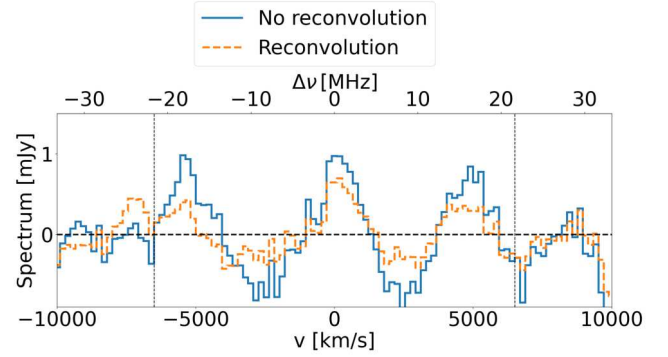
where we have omitted the additive systematics.

For a data component  $\mathbf{d}_i$  and a multiplicative systematics matrix  $\mathbf{S}_i$ , effects such as bandpass errors lead to a diagonal  $\mathbf{S}_i$  matrix, whereas convolutional effects lead to a nondiagonal matrix. It is easy to see that, when  $\mathbf{S}_i$  is diagonal,  $\mathbf{S}_M^i \mathbf{C}_i (\mathbf{S}_M^i)^T$  produces the same correlation matrix as the underlying covariance  $\mathbf{C}_i$ . As a result, we expect that the estimated correlation matrix from the data, shown in Figures 14 and 16, should be similar to the correlation matrix in the mock simulations. As discussed in Section 6, this is not the case, as the effects of the systematics can be clearly seen in the correlation matrices. Therefore, the systematics must be a convolutional effect on the data. This also aligns with the oscillation feature seen in the stacked spectrum, which requires a convolution of the oscillation feature with the H I emission-line peaks along the spectral direction.

The existence of convolutional systematics hints toward the connection between the oscillations in the stacked spectrum and the instrument beam. Beam chromaticity is known to be a limiting factor for intensity mapping surveys (e.g., A. Sampath et al. 2024), introducing contamination in the foreground removal procedure (S. D. Matshawule et al. 2021; M. Spinelli et al. 2022). To confirm the connection, we examine the stacked signal with reconvolved intensity maps. Instead of directly stacking on the data, we first perform a convolution of the intensity map at each frequency channel so that

$$I_{\text{reconv}}(l, m, \nu) = B_{\text{reconv}}(l, m, \nu) \otimes I(l, m, \nu), \quad (37)$$

where the reconvolution kernel,  $B_{\text{reconv}}(l, m, \nu)$ , deconvolves a frequency-dependent Gaussian beam and then convolves the



**Figure 17.** The comparison between the stacked spectra with and without the reconvolution. The blue solid line shows the stacked spectrum without the reconvolution (“No reconvolution”). The orange dashed line shows the spectrum with the reconvolution (“Reconvolution”). The vertical dashed lines show the  $|\nu| < 6500 \text{ km s}^{-1}$  boundary.

map to a common Gaussian kernel so that

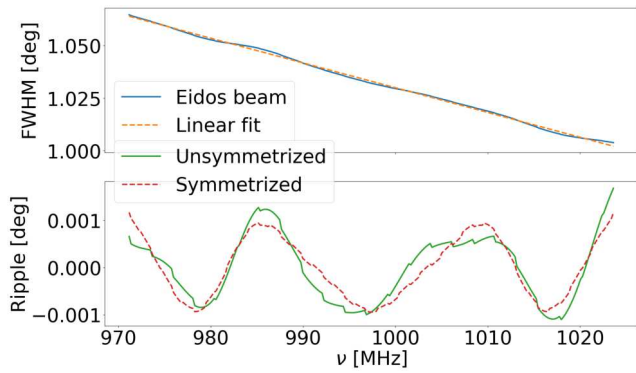
$$B_{\text{reconv}}(l, m, \nu) = \exp \left[ -\frac{1}{2} \frac{\theta^2}{\gamma \sigma_{\text{max}}^2 - \sigma(\nu)^2} \right], \quad (38)$$

where  $\sigma(\nu)$  is the beam size of the MeerKAT telescope assuming a Gaussian beam at each frequency,  $\sigma_{\text{max}}$  is the maximum beam size in the frequency sub-band, and  $\gamma$  is a scaling factor to scale down the final resolution. We follow MK25 to calculate the reconvolution kernel, and perform the PCA cleaning and stacking with the reconvolved maps.

In Figure 17, we show the comparison of the stacked spectra with and without the reconvolution. For simplicity, we only showcase the unsymmetrized spectrum. When reconvolved, the stacked signal exhibits an overall decrease in amplitude. This is expected, since smoothing the maps to a lower resolution attenuates fluctuations at small scales. More importantly, there is a visible decrease in the amplitude of the systematics relative to the central peak. The oscillation structure is also less localized, as seen in the  $\nu \sim -5000 \text{ km s}^{-1}$  region. The fact that resmoothing the maps affects the oscillation features of the systematics suggests the systematics are related to the chromaticity of the primary beam. Reconvolution partially eliminates the frequency-dependency of the map resolution, which is the incentive of performing the reconvolution in the power spectrum analysis. The stacking measurement suggests that while there is a small effect of mitigating the systematics from the reconvolution, the systematics are still significant and become harder to describe as the oscillations are less localized (see also Appendix B of S. D. Matshawule et al. 2021).

### 7.3. Beam Oscillations

In S. D. Matshawule et al. (2021), it is found that the contamination after foreground cleaning can be caused by the frequency ripple in the primary beam of the MeerKAT telescope. The primary beam size of MeerKAT oscillates in frequency, due to the diffractive interference between the secondary and the primary reflector of the dish (D. I. L. de Villiers 2013). Measurements of the beam in K. M. B. Asad et al. (2021) show that it leads to a small modulation in the supposedly smooth dependency on frequency. It is then natural



**Figure 18.** Upper panel: the blue solid line shows the FWHM of the MeerKAT primary beam across the frequency sub-band from the EIDOS model (“Eidos beam”). The orange dashed line shows a linear best fit of the FWHM (“Linear fit”). Lower panel: the green solid line shows the differences between the FWHM and its linear fit (“Unsymmetrized”), which represents the ripple in the primary beam. The red dashed line shows the symmetrized beam ripple (“Symmetrized”) according to Equation (42), which is used for forward modeling.

to speculate that the ripple in the beam leads to the oscillations in the stacked spectrum.

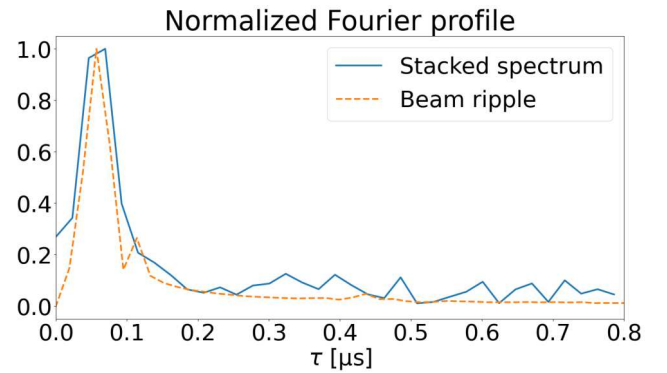
In Figure 18, we use the EIDOS beam model to calculate the ripple in the primary beam. Following the procedure in K. M. B. Asad et al. (2021), we first calculate the beam area using only the primary beam, which we choose to be where the beam is larger than 0.1. We then use the primary beam area to calculate the effective FWHM of the primary beam. The primary beam FWHM indeed exhibits a small ripple, as seen in the upper panel of Figure 18. We then fit the smooth component using a linear best fit, and subtract the smooth part out to obtain the ripple. The ripple shown in the lower panel exhibits an oscillating feature, similar to the one seen in the stacked spectrum.

To further illustrate the connection, we show that the characteristic frequency scale of the oscillation and the beam ripple is the same. We use the Fourier transform along the frequency direction, defined as

$$\tilde{f}(\tau) = \int d\nu \exp[-2\pi i\tau\nu]f(\nu), \quad (39)$$

where  $f(\nu)$  is a function of frequency, and  $\tau$  is the Fourier pair of  $\nu$ . We then perform the Fourier transform for the beam ripple shown in Figure 18 and the unsymmetrized stacked profile shown in Figure 16. The transformed property is then rescaled, so that the maximum amplitude in Fourier space is 1. The results are shown in Figure 19. The beam ripple and the stacked profile share the same peak position centered around  $\sim 0.05 \mu\text{s}$ , giving an oscillating frequency of  $\sim 20 \text{ MHz}$ . Both the position of the peak and the width of the peak overlap closely, providing strong evidence that the beam ripple and the oscillating systematics have the same origin. The Fourier transform of the stacked profile is similar to the line-of-sight power spectrum often used in intensity mapping studies. For example, in S. D. Matshawule et al. (2021), it is also found that insufficient cleaning of foregrounds in the presence of beam ripple can lead to a peak in the line-of-sight power spectrum.

In conclusion, the oscillating systematics originate from the chromatic ripple in the primary beam of the MeerKAT telescope. We note that, however, the interaction between the



**Figure 19.** The blue solid line shows the Fourier transformed profile of the unsymmetrized stacked spectrum (“Stacked spectrum”). The orange dashed line shows the Fourier profile of the beam ripple (“Beam ripple”). The profiles are normalized so that the maximum amplitude is 1.

beam ripple and the stacked signal is not direct. In Section 4, we use the EIDOS beam, which includes the beam ripple, and we find no systematics in the stacked signal in the mock. If the systematics in the data are simply caused by the beam convolution, the noise-dominated reference spectrum of the data should not have the systematics. Yet, in the shuffling covariance of the data, we observe the oscillating systematics when visualizing the correlation matrix in Figure 16. It is worth noting that the beam model affects the modeling of the sky signal for calibration, as described in Section 2. The calibration solution is then affected by the systematics, leading to chromatic calibration errors that are commonly seen in the case of imperfect sky model (e.g., N. Barry et al. 2016; I. Heywood et al. 2020). The calibration errors serve as a multiplicative effect on the sky signal, which then affects the PCA cleaning of the data. The PCA cleaning matrix, as described in Equation (26), is then applied to the entire data vector including the noise. In conclusion, the systematic oscillation in the stacked spectrum is introduced by the PCA, whose cleaning matrix is modulated by the MeerKAT beam ripple. An end-to-end study of the systematics from the calibration of the time-ordered data is beyond the scope of this work. Instead, we use the connection between the beam ripple and the systematics as a starting point for forward modeling the signal.

## 8. Model Fitting

In this section, we describe the model inference framework we use in this work to constrain the systematics and the HI signal.

### 8.1. Forward Modeling

As we have extensively discussed in Section 4.4, in the forward modeling, we use the simplified assumption and distribute the total HI mass inside the GAMA survey region to the GAMA galaxies. Therefore, we can simulate the HI signal with one free parameter  $\bar{M}_{\text{HI}}$  following the procedure described in Section 3.1. We briefly review the procedure below.

First, we generate the positions of galaxies using the lognormal simulation routine. The number density is set so that the expected number of galaxies within the GAMA region is equal to the catalog. Note that in each realization, the number of galaxies is not equal to the number of galaxies in the

GAMA catalog due to assigning Poisson random to the mock dark matter field. This is desired, as the modeling is supposed to reflect the Poisson fluctuations.

We then assign a uniform HI mass,  $\bar{M}_{\text{HI}}$ , to each galaxy and generate the HI profile. The HI signal is then convolved with the beam model to produce the HI map. We then apply a modeling of the oscillating systematics,  $f_{\text{sys}}(\nu)$ , to the map so that in each pixel the HI signal is convolved with the systematics,

$$I_{\text{sys}}(\alpha, \phi, \nu) = \sum_{\Delta\nu_i=\nu-\nu_0}^{\nu_1-\nu} f_{\text{sys}}(\Delta\nu_i + \nu_0) I_{\text{HI}}(\alpha, \phi, \nu - \Delta\nu_i), \quad (40)$$

where  $[\nu_0, \nu_1]$  are the lower and upper limits of the frequency sub-band,  $I_{\text{HI}}$  is the HI map before applying systematics, and  $\Delta\nu_i$  iterates over a step size of frequency channel bandwidth. The parameterization of  $f_{\text{sys}}(\nu)$  is discussed later in detail in Section 8.2.

The HI signal with systematics is then cleaned by applying the PCA cleaning matrix *calculated from data*. This is to ensure that the model has the same level of signal loss as the data. As we have shown in Section 4.2, for the HI-only mock and the full mock, the level of signal loss from PCA is very different due to the change in the removed modes. Since we do not have prior information on the exact level of the oscillating systematics, and the fact that the foreground model and the data have a mismatch, the PCA cleaning matrix from the mock simulation is different from the data, leading to different signal loss properties. Therefore, applying the PCA cleaning matrix from data is important to keep the signal loss consistent.

The HI map and the GAMA-like catalog are then used to perform stacking, using the same weighting as the data. The stacked signal is then used for model fitting.

### 8.2. Parameterizing the Beam Oscillation

We first describe the parameterization of the beam oscillation, which we use to describe the systematics. As shown in Figure 19, the main feature of the beam oscillation is the peak structure in Fourier space. The systematics in Fourier space can then be parameterized to reflect the peak structure, so that

$$|\tilde{f}_{\text{sys}}(\tau)|^2 = \exp\left[-\frac{\left(\tau - \frac{1}{\nu_{\text{sys}}}\right)^2}{2\sigma_\tau^2}\right], \quad (41)$$

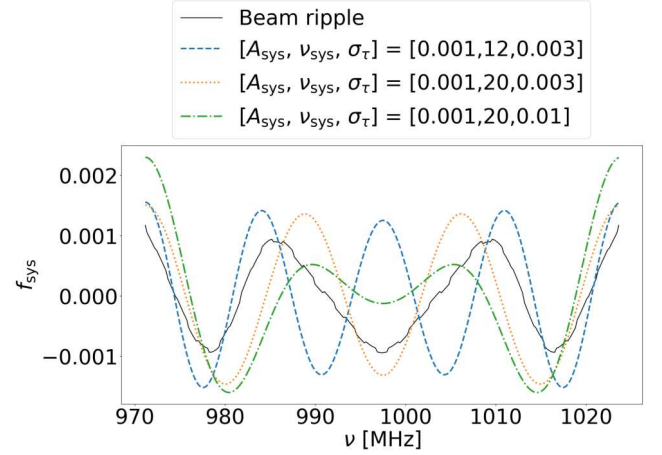
where  $1/\nu_{\text{sys}}$  is the position of the peak corresponding to an oscillating frequency of  $\nu_{\text{sys}}$ , and  $\sigma_\tau$  is the width of the peak.

After specifying the parameters  $\nu_{\text{sys}}$  and  $\sigma_\tau$ , we then perform an inverse Fourier transform of  $\tilde{f}_{\text{sys}}(\tau)$ .  $\tilde{f}_{\text{sys}}(\tau)$  is assumed to be real-valued, which leads to a symmetric function in real space so that

$$f(\nu_0 + \Delta\nu) = f(\nu_1 - \Delta\nu), \quad (42)$$

where  $\nu_0$  and  $\nu_1$  are the lower and upper limits of the frequency sub-band, respectively. Subsequently, the beam ripple discussed in Section 7.3 can be symmetrized in the same way.

The function in real space is then mean-subtracted, and then rescaled so that the standard deviation of the function in the frequency sub-band is 1. We then multiply another free factor,



**Figure 20.** An illustration of the parameterization of the systematics  $f_{\text{sys}}$ . The black solid line shows the beam ripple. The other three lines show three different sets of parameters for the systematics listed in the figure.

$A_{\text{sys}}$ , so that

$$\text{std}(f_{\text{sys}}) = A_{\text{sys}}, \quad (43)$$

to denote the amplitude of the systematics.

In short, we use three parameters,  $\{A_{\text{sys}}, \nu_{\text{sys}}, \sigma_\tau\}$  to model the systematics to describe the amplitude, characteristic frequency, and the shape of the systematics, respectively. An illustration of the parameterization is shown in Figure 20.

In the fitting, to examine the impact of the parameterization, we adopt three different scenarios. First, we only vary  $A_{\text{sys}}$ , and keep the shape of the  $\tilde{f}_{\text{sys}}(\tau)$  fixed to the beam oscillation. Second, we fix  $\sigma_\tau$  to be very small and vary  $A_{\text{sys}}$  and  $\nu_{\text{sys}}$ , so that the oscillating frequency varies while  $f_{\text{sys}}$  has the shape of a sine function. Finally, we allow all three parameters to vary to explore the full parameter space.

We emphasize that, the forward modeling routine used in this work is simplified. For example, we fix cosmological parameters and other HI-related parameters and only keep  $\bar{M}_{\text{HI}}$  as a free parameter. The assumption that the systematics  $f_{\text{sys}}$  directly convolve with the HI signal along the frequency direction is an effective approach to model a much more complex phenomenon that interacts with the data at the level of the time-ordered data. Given that our detection significance in the symmetrized spectrum is  $\sim 5.5\sigma$ , we are limited by the number of parameters that we can vary to achieve sensible constraints. A more detailed study is left for the future MeerKLASS data analysis, where the depth of both the HI intensity maps and the galaxy catalog will be larger.

### 8.3. Importance Nested Sampling

The forward modeled HI signal produces a stacked spectrum,  $\bar{I}_{\text{model}}$  to be used to fit against the data. Using the estimated covariance  $\mathbf{C}$  and the measured symmetrized spectrum  $\bar{I}_{\text{data}}$ , relative changes in the likelihood  $\mathcal{L}$  can be calculated as

$$\Delta \log \mathcal{L} = -\frac{1}{2}(\mathbf{I}_{\text{model}} - \mathbf{I}_{\text{data}})\mathbf{C}^{-1}(\mathbf{I}_{\text{model}} - \mathbf{I}_{\text{data}})^{\text{T}}, \quad (44)$$

where  $\log$  stands for natural logarithm,  $\mathbf{I}_{\text{model}}$  is the model vector for the model spectrum at each  $\Delta\nu$ , and  $\mathbf{I}_{\text{data}}$  is the data vector.

**Table 1**  
Priors We Use for the Parameters in This Work

$\log_{10}[\bar{M}_{\text{HI}}/M_{\odot}]$	$A_{\text{sys}}$	$\nu_{\text{sys}}$ (MHz)	$\sigma_{\tau}$ ( $\mu\text{s}$ )
[7.0, 14.0]	[0.0, 1.0]	[10, 32]	[0.003, 0.04]

**Note.** All priors are flat priors.

The likelihood is then used to perform Bayesian inference using importance nested sampling (INS; e.g., F. Feroz et al. 2019). INS explores the prior volume of the parameter space, identifies the region with the highest likelihood, and the sample with the posterior for Bayesian inference. In practice, the algorithm starts with a number of points randomly distributed in the prior volume, finding the live points with the highest likelihood. Based on the positions of the live points, it estimates a boundary around each point to sample within, and identifies a new iteration of the random points with higher minimum likelihood. Iterating this process will result in convergence so that the boundary no longer changes for each point. We use NAUTILUS<sup>23</sup> (J. U. Lange 2023) for performing the sampling.

The sampling depends on the priors we set for the parameters. Given our lack of understanding of the model parameters, we adopt wide flat priors, which we list in Table 1. The prior for  $\log_{10}[\bar{M}_{\text{HI}}/M_{\odot}] \in [7.0, 14.0]$  translates to an effective HI density of  $\Omega_{\text{HI}} \in [1.5 \times 10^{-8}, 0.155]$ . Measurements of HI density at various redshifts give the HI density to be  $\sim 5 \times 10^{-4}$  (see, e.g., Figure 14 of W. Hu et al. 2019 and references therein), and our prior is significantly looser than the current constraints from observations. The amplitude of the systematics is sampled from  $A_{\text{sys}} \in [0, 1]$ , since values outside this range give unphysical negative values. The oscillating frequency is sampled from  $\nu_{\text{sys}} \in [10, 32]$  MHz. Higher values of the oscillation frequency will not be captured in our  $\Delta\nu \lesssim 20$  MHz range for the stacked spectrum. Lower values, on the other hand, lead to rapid oscillations that are unphysical given the physical distance between the primary and secondary reflector of the MeerKAT telescope. Finally, the width of the peak in the Fourier space is sampled from  $\sigma_{\tau} \in [0.003, 0.04]$   $\mu\text{s}$ . Values smaller than  $0.003 \mu\text{s}$  are below the resolution of the frequency sub-band we use. Values larger than  $0.04 \mu\text{s}$  will result in a wide peak, so that  $f_{\text{sys}}$  in frequency space is almost completely flat, which is unphysical.

For each fitting, 2000 live points are used for sampling. A weighted sample of parameter values is then returned for Bayesian inference, which we discuss in the next section.

## 9. Results from Forward Modeling

In this section, we present the main results of this paper, obtained from the Bayesian inference of the stacked spectrum. As discussed in Section 8.2, we consider three different parameterizations of the systematics, varying the full shape of the oscillations, only the amplitude and the frequency, and only the amplitude. In each case, we also treat the HI mass of the galaxies  $\bar{M}_{\text{HI}}$  as a free parameter for sampling. From now on, for simplicity, we will denote each case using their respective number of free parameters, for example the two-

parameter case refers to varying  $[\bar{M}_{\text{HI}}, A_{\text{sys}}]$  and fixing the shape of beam oscillation. Later in Section 9.2, we discuss the fitting results with the full shape of the systematics but using a shuffling covariance instead of the corrected covariance, which we denote as the shuffling covariance case.

### 9.1. Parameter Constraints

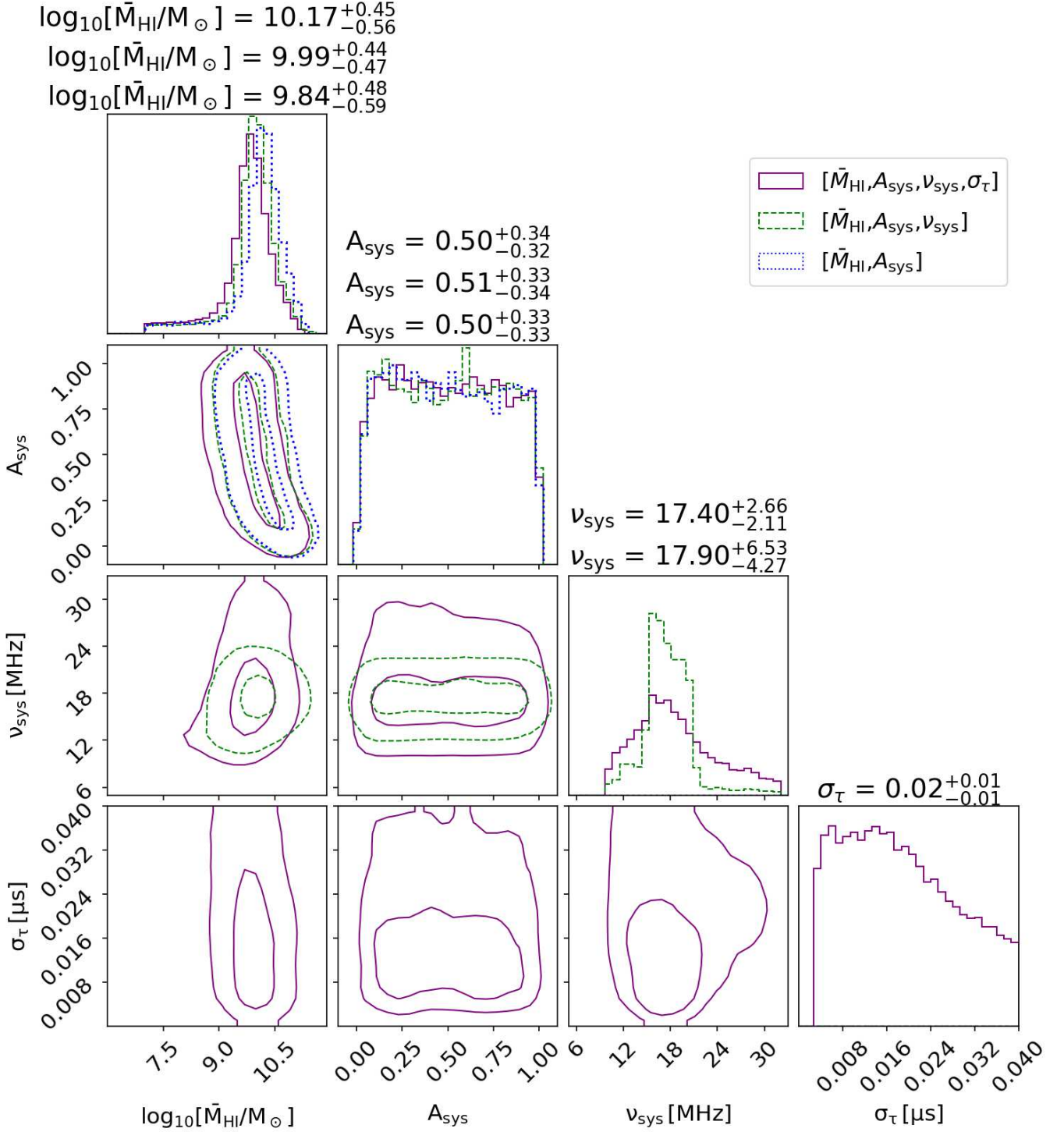
Using the posterior obtained from INS, we compute the  $1\sigma$  confidence interval, i.e., the 16th, 50th, and 84th percentiles, of each parameter and show the results in Table 2. The posterior is also used to visualize the 2D posterior of the parameters, as shown in Figure 21. For all cases of model fitting, we obtain consistent constraints on  $\bar{M}_{\text{HI}}$  with the results from all three scenarios, with their differences smaller than the  $1\sigma$  confidence interval. As the number of free parameters increases for describing the systematics, the amplitude of the HI emission,  $\bar{M}_{\text{HI}}$ , decreases. The consistent shifts of the amplitude of HI suggests that the modeling of the systematics may impact the inference of the HI signal.

The amplitude of the systematics,  $A_{\text{sys}}$ , is not constrained, as the 68% interval of the posterior extends to almost the entire prior volume. While we have not achieved a constraint on the amplitude of the systematics, we note that  $A_{\text{sys}}$  is likely higher than expected. For example, for the four-parameter case, the 5th percentile of the posterior gives  $A_{\text{sys}} > 6.71\%$ . The beam oscillation, on the other hand, is  $\sim 0.1\%$ , as shown in Figure 18. This confirms our previous discussion in Section 7.3, that the way that systematics are coupled to the data is not a trivial convolution of the primary beam and worth further investigation in future work.

Comparing the two-parameter and the three-parameter case, we can see that varying the oscillation frequency does not lead to larger measurement error but in fact a smaller 68% interval for  $\bar{M}_{\text{HI}}$ . This suggests that, while the beam oscillation does match the systematics in the stacked spectrum well, the exact oscillation frequency may be different from the measurement from the EIDOS beam. The improvement in the fitting can also be seen in the small increase in Bayesian evidence  $\log\mathcal{Z}$ . The slight mismatch is expected, since the frequency of the beam ripple of the telescope has a dependency on elevation, and the effective frequency of the oscillation may be different from the measurements using one night of tracking observation in K. M. B. Asad et al. (2021).

When we vary the full shape of the systematics function in the four-parameter case, it can be seen that the constraints on the oscillation frequency  $\nu_{\text{sys}}$  degrade significantly. When  $\sigma_{\tau}$  is fixed so that the systematics follow the shape of a sine function, the constraint on  $\nu_{\text{sys}}$  gives  $\nu_{\text{sys}} = 17.40_{-2.11}^{+2.66}$  MHz. In the four-parameter case, however, the measurement error increases by a factor of  $\sim 2$ , which gives  $\nu_{\text{sys}} = 17.90_{-4.27}^{+6.53}$  MHz.  $\sigma_{\tau}$  is not well constrained, as the 68% interval occupies a large part of the prior volume. The increase in the errors is likely due to the fact that the stacked spectrum is measured in relatively low-frequency resolution, and therefore cannot be used to describe the shape of the oscillations in detail. There is also a decrease in Bayesian evidence  $\log\mathcal{Z}$ . This suggests that a small value of  $\sigma_{\tau}$ , which makes the oscillations following a sine function, is good enough for modeling the stacked signal. Nevertheless, we adopt the more conservative estimation in the four-parameter case as our final results. The measured oscillation frequency  $\nu_{\text{sys}} \approx 18$  MHz is consistent with the  $\sim 20$  MHz beam ripple discussed in the literature.

<sup>23</sup> <https://nautilus-sampler.readthedocs.io/>



**Figure 21.** The posterior distribution of the parameter fitting results of this work. The histogram plots show the posterior distribution of each model parameter when marginalizing over all other parameters. The contour plots show the marginalized 2D posterior distribution of each parameter pair. The outer contour denotes the  $2\sigma$  confidence region, whereas the inner contour denotes the  $1\sigma$  confidence region. Three scenarios are shown, with the shape of systematics fixed (“ $[\bar{M}_{\text{HI}}, A_{\text{sys}}]$ ”), varying the amplitude and the oscillation frequency (“ $[\bar{M}_{\text{HI}}, A_{\text{sys}}, \nu_{\text{sys}}]$ ”), and varying the full shape of the systematics (“ $[\bar{M}_{\text{HI}}, A_{\text{sys}}, \nu_{\text{sys}}, \sigma_{\tau}]$ ”). The title of each histogram plot shows the median and the 68% interval. From top to bottom, the results shown are for the two-parameter, three-parameter, and four-parameter cases, respectively.

We now further examine the constraints of the model parameters in terms of the degeneracy between parameters, as shown in Figure 21. For the HI amplitude  $\bar{M}_{\text{HI}}$ , we can see that the distribution of the 1D posterior is well constrained, with an extended tail at the lower end of the distribution. We discuss

the implications of the  $\bar{M}_{\text{HI}}$  posterior in more detail later in Section 9.3. The amplitude of the systematics,  $A_{\text{sys}}$ , is indeed not constrained, as the posterior simply extends throughout the flat prior. When the three-parameter model is considered, the oscillation frequency  $\nu_{\text{sys}}$  is well constrained. In the four-

**Table 2**  
68% Intervals of the Model Parameters from Fitting the Stacked Spectrum of the MeerKLASS  $L$ -band Deep-field Data onto the GAMA Galaxies

	$\log_{10}[\bar{M}_{\text{HI}}/M_{\odot}]$	$A_{\text{sys}}$	$\nu_{\text{sys}}$ (MHz)	$\sigma_{\tau}$ ( $\mu\text{s}$ )	$\log \mathcal{Z}$
$[\bar{M}_{\text{HI}}, A_{\text{sys}}, \nu_{\text{sys}}, \sigma_{\tau}]$	$9.84^{+0.48}_{-0.59}(10.24)$	$0.50^{+0.33}_{-0.33}(0.63)$	$17.90^{+6.53}_{-4.27}(16.17)$	$0.02^{+0.01}_{-0.01}(0.004)$	-14.08
$[\bar{M}_{\text{HI}}, A_{\text{sys}}, \nu_{\text{sys}}]$	$9.99^{+0.44}_{-0.47}(10.77)$	$0.51^{+0.33}_{-0.34}(0.10)$	$17.40^{+2.66}_{-2.11}(19.49)$	...	-13.85
$[\bar{M}_{\text{HI}}, A_{\text{sys}}]$	$10.17^{+0.45}_{-0.56}(10.17)$	$0.50^{+0.34}_{-0.32}(0.87)$	...	...	-13.92
Shuffling Cov.	$9.93^{+0.57}_{-1.05}(10.92)$	$0.43^{+0.37}_{-0.31}(0.10)$	$17.84^{+6.41}_{-4.11}(19.30)$	$0.02^{+0.01}_{-0.01}(0.003)$	-21.31

**Note.** Three different parametrizations of the systematics are considered, which vary the full shape of the oscillations (“ $[\bar{M}_{\text{HI}}, A_{\text{sys}}, \nu_{\text{sys}}, \sigma_{\tau}]$ ”), only the amplitude and the frequency (“ $[\bar{M}_{\text{HI}}, A_{\text{sys}}, \nu_{\text{sys}}]$ ”), and only the frequency (“ $[\bar{M}_{\text{HI}}, A_{\text{sys}}]$ ”). For reference, the maximum a posteriori estimations of each parameter are listed in the brackets. The last column lists the Bayesian evidence  $\log \mathcal{Z}$  of each fitting. The last row shows the results using the shuffling estimate of the covariance (“Shuffling Cov.”) instead of the corrected covariance estimate for the full shape case.

parameter case, on the other hand, the tails of the posterior reach the physically driven prior. This is caused by the posterior of  $\sigma_{\tau}$  not being constrained.

Finally, we comment on the fact that the estimated  $\bar{M}_{\text{HI}}$  is much lower than expected. Note that, as we have discussed in Section 4.4, the values of  $\bar{M}_{\text{HI}}$  should not be interpreted as the average HI mass of the GAMA galaxies, but as the total HI mass in the GAMA survey region distributed among the GAMA sample. For  $\Omega_{\text{HI}} \sim 0.5 \times 10^{-3}$ , we expect  $\bar{M}_{\text{HI}} \sim 10^{11} M_{\odot}$ , and our estimation is an order of magnitude lower than expected. This suggests that there may be issues in the model fitting, and while there is a tentative measurement of the HI density, the estimation is likely to be biased. For the rest of this section, we examine issues of covariance estimation and parameter degeneracy that contribute to the underestimation.

### 9.2. Impact of Covariance Estimation

In Section 5, we discussed in detail how we obtain the covariance estimation using the random shuffling of galaxy positions. The covariance estimation, while corrected for signal covariance, leads to a distortion compared to the true covariance, which may impact the inference of the systematics, as shown in Appendix. While a more accurate covariance estimation is beyond the scope of this work, we can use the shuffling covariance without the correction to perform the sampling and compare the results to understand the impact of covariance estimation.

In the bottom row of Table 2, we show the four-parameter case with shuffling covariance. Note that, without the signal covariance correction, the measurement error of the stacked spectrum is lower, as seen in Figure 16. However, the resulting measurement error for  $\bar{M}_{\text{HI}}$  becomes larger, and the Bayesian evidence  $\log \mathcal{Z}$  decreases significantly compared to the corrected covariance. This suggests that indeed a correction to the shuffling covariance is needed. The constraints on the oscillation frequency, on the other hand, are robust against the choice of systematics. This is expected, as the primary feature of the stacked spectrum is the oscillation, so that the information in the stacked spectrum mostly goes to constraining  $\nu_{\text{sys}}$ . Furthermore, the measured maximum a posteriori (MAP) values of the systematics are lower when the shuffling covariance is used, while the HI amplitude is larger.

To further illustrate the effect of covariance estimation, we show the 2D posterior distribution of the model parameters for the shuffling covariance compared against the corrected covariance in Figure 22. Notably, the posterior for  $A_{\text{sys}}$  changes significantly, leading to a peak at  $A_{\text{sys}} \sim 0.15$ . As the amplitude

of the systematics decreases, we note that there is a turn in the 2D posterior distribution of  $A_{\text{sys}} - \bar{M}_{\text{HI}}$ , leading to a higher estimation of  $\bar{M}_{\text{HI}}$ . Similarly, a peak around small values of  $\sigma_{\tau}$  also appears when shuffling covariance is used.

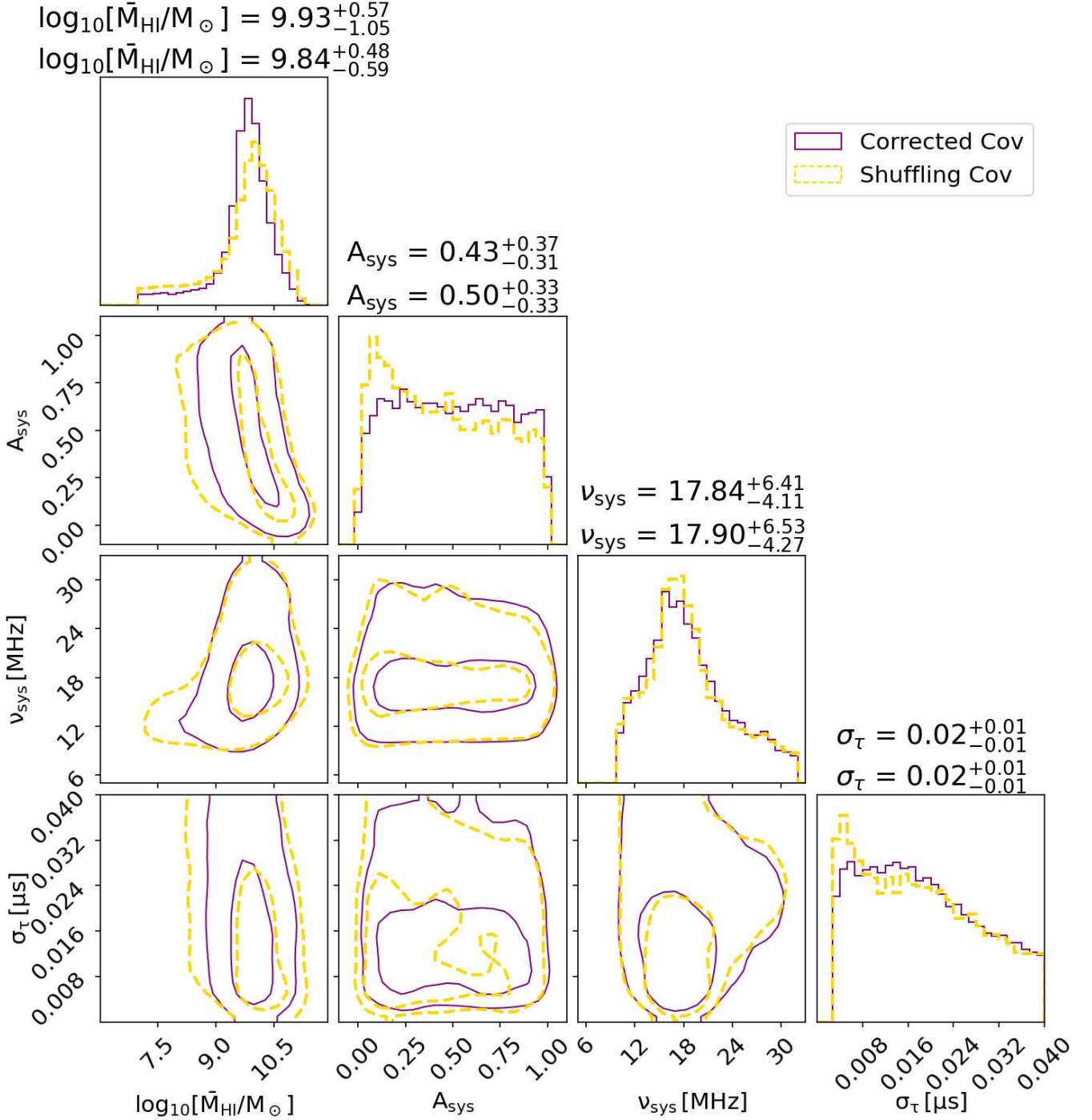
The comparison between the two choices of covariance shows that the constraints on systematics are affected by the covariance estimation. When the amplitude parameter reaches small values  $A_{\text{sys}} \lesssim 0.1$ , there is a stronger anticorrelation between the systematics amplitude and the HI mass. We note that this leads to a larger estimation of  $\bar{M}_{\text{HI}}$ , which suggests that covariance estimation impacts the underestimation of the HI density.

### 9.3. Projection Effects

We further explore the degeneracy between  $\bar{M}_{\text{HI}}$  and  $A_{\text{sys}}$ . Since we are interested in the underestimation of  $\Omega_{\text{HI}}$ , we convert the posterior of  $\bar{M}_{\text{HI}}$  to  $\Omega_{\text{HI}}$  and show the results in Figure 23. We include the contour of the  $3\sigma$  region to fully visualize the posterior. As shown, the degeneracy direction between the two parameters has a turn, illustrated by the black dotted line in Figure 23. When the systematics amplitude is large with  $A_{\text{sys}} \gtrsim 0.2$ , an increase in  $\Omega_{\text{HI}}$  leads to a sharp decrease in  $A_{\text{sys}}$ . As  $A_{\text{sys}}$  further decreases, the HI density increases significantly. In the ideal case of a stacking experiment, we expect that no systematics are present so that  $\Omega_{\text{HI}} \sim 0.5 \times 10^{-3}$ ,  $A_{\text{sys}} = 0$ , which we denote as the red star. The  $\Omega_{\text{HI}} \sim 5 \times 10^{-4}$  value lies slightly outside the  $3\sigma$  contour of the posterior, suggesting that there is likely a modeling imperfection that leads to the underestimation of the HI density.

Furthermore, the degeneracy between the two parameters leads to strong posterior projection effects (e.g., A. Gómez-Valent 2022) in the model inference. “Posterior projection effect” refers to the issue that the marginalized 1D distribution of the posterior may be skewed due to the complicated degeneracy between the model parameters. We first demonstrate that the 1D posterior distribution is indeed skewed. In the brackets of the reported values of Table 2, we show the MAP estimation<sup>24</sup> of the parameters. If the 1D posterior of the parameters follow ideal Gaussian distributions, it is expected that the MAP estimation and the median of the posterior should agree well with each other. In all cases, except the two-parameter model, we find the systematic shift of HI mass to higher values for the MAP estimation when comparing the median of the posterior. In particular, there is a  $>1\sigma$

<sup>24</sup> Since we adopt flat priors for all model parameters, in our case, the MAP estimation is simply the maximum of the 1D posterior distribution.

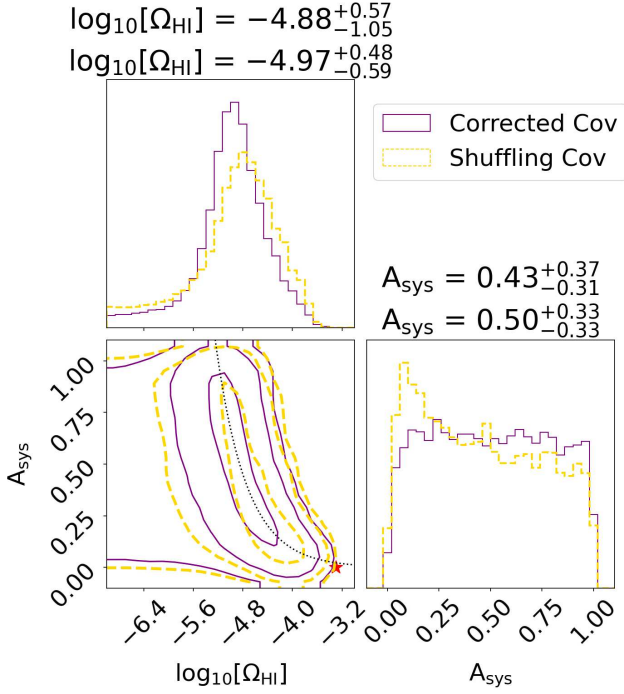


**Figure 22.** The same as Figure 21, but with cases of corrected covariance and shuffling covariance for the four-parameter model. From top to bottom, the titles shown are for the shuffling covariance and the corrected covariance, respectively.

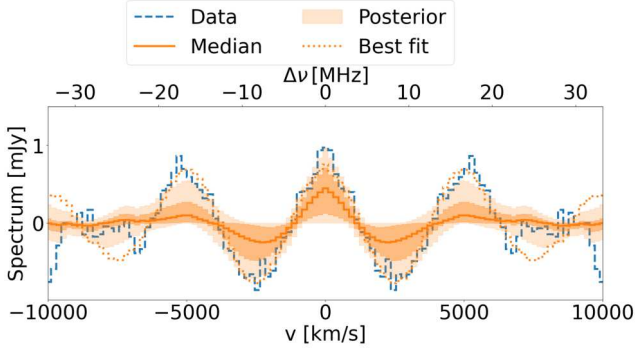
discrepancy between the MAP and the median of  $\bar{M}_{\text{HI}}$  for the three-parameter case, as well as in the shuffling covariance case. It suggests that the skewed distribution of the marginalized posterior contributes to the underestimation of the HI density. The skewing is consistent with the degeneracy direction between  $A_{\text{sys}}$  and  $\bar{M}_{\text{HI}}$  and, therefore, is due to projection effects caused by parameter degeneracy.

The projection effects can be better understood in the posterior of the fitted spectrum, which we show in Figure 24. Due to the relatively low signal-to-noise ratio, the posterior of

the model spectrum favors a small overall amplitude. The median of the posterior has a large deviation from the best-fit model. Around the secondary peaks of systematic oscillations, the best-fit model can be outside the 95% confidence interval. Recall that in Figure 23, the expected level of  $\Omega_{\text{HI}}$  is also at the boundary of the  $3\sigma$  confidence region. Large HI signal will produce a better fit in the central area  $\nu \sim 0$ , which is weighted less compared to the secondary peaks of the systematics due to the signal covariance correction. This leads to a stronger degeneracy between  $\Omega_{\text{HI}}$  and  $A_{\text{sys}}$ , since a lower HI amplitude



**Figure 23.** The 2D posterior distribution between the inferred H I density  $\Omega_{\text{HI}}$  and the amplitude of the systematics, for the cases with corrected covariance and shuffling covariance of the four-parameter model. The outer, middle, and inner contours denote the  $3\sigma$ ,  $2\sigma$ , and  $1\sigma$  regions. From top to bottom, the titles shown are for the shuffling covariance and the corrected covariance, respectively. The dotted line in the 2D posterior shows an illustrative direction of parameter degeneracy. The red star denotes  $\Omega_{\text{HI}} = 0.5 \times 10^{-3}$ ,  $A_{\text{sys}} = 0.0$ .



**Figure 24.** The posterior distribution of the fitted spectrum. The orange solid line denotes the median values of the model spectrum (“Median”). The dark shaded region shows the 68% interval of the fitted spectrum, and the light shaded region shows the 95th percentile. The orange dotted line shows the model spectrum of the highest log-likelihood (“Best fit”). The stacked spectrum is shown as the blue solid line for reference (“Data”).

can always be compensated by a higher level of systematics to fit peaks and troughs of the oscillating systematics. Therefore, when shuffling covariance is considered, higher values of  $\bar{M}_{\text{HI}}$  are preferred, as the central region is weighted more by the covariance. Since the secondary peaks are the convolution of systematics and the H I signal, in the ideal case, we should expect the amplitude of the variance will also follow the oscillations. However, as we do not have any prior information on the underlying H I signal, we resort to using the current covariance. Note that due to the signal-to-noise ratio of the measurement, we do not expect the posterior of  $A_{\text{sys}}$  to be well constrained. Therefore, while a better covariance estimation

gives a more accurate correlation matrix of the measured stacked spectrum, we do not expect the parameter space of  $A_{\text{sys}}-\Omega_{\text{HI}}$  to be tightened in the posterior.

Regardless of the covariance estimation, the constraints on the oscillation frequency  $\nu_{\text{sys}}$  remain robust and provide strong evidence that the systematics originate from the beam ripple of the instrument.

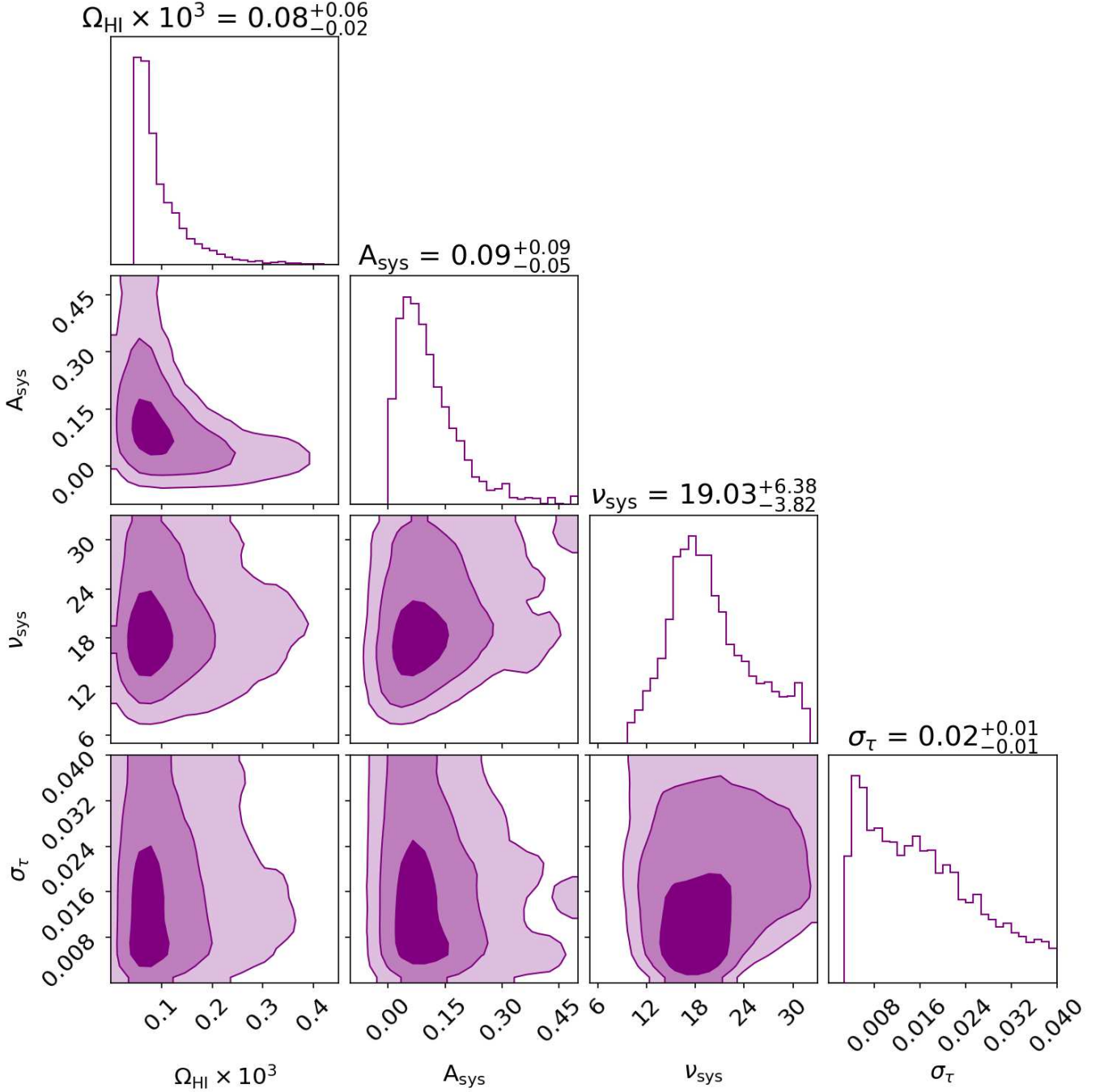
## 10. Discussion

In this paper, we showed that the emission-line stacking of the H I intensity maps can be used to examine the systematics in the data and infer the H I density of the survey volume. Using forward modeling of the stacked spectrum, Bayesian inference can be used to constrain the systematics as well as the H I signal. The findings of this paper in the context of stacking have implications on the cosmological analysis of the H I clustering signal, which we discuss below.

In the power spectrum analysis, it is common to fully utilize the 3D information on the clustering in  $k$ -space to maximize the information content, for example, by using the multipole clustering wedges (e.g., J. N. Grieb et al. 2017). For the stacked signal cube, the measured signal is the H I density at a given separation to the galaxy position, which is similar to the two-point cross-correlation function. It is natural to conclude that the optimal summary statistics is to model a cylindrical signal by averaging the stacked cube into  $\Delta\theta-\Delta\nu$  space (D. A. Dunne et al. 2025). However, as we explore in Section 5, the covariance of such a signal will be difficult to model, due to the complicated correlation between the angular pixels. If the primary purpose of stacking is to constrain the systematics and the overall H I amplitude, the stacked spectrum is sufficient and is easier to model.

In Section 9, we explore in detail how the estimation of the H I density is biased by the systematics due to the degeneracy between the two. From this conclusion, we can envision that if there is no additional mitigation of systematics, a lower bound of the amplitude of H I clustering will not be robust. The parameter space where the H I signal is extremely low and the signal is mainly driven by the systematics is within the 95% interval of the posterior. We note that the lack of constraining power on the lower limit can be resolved by simply having deeper observations. As shown in Figure 24, the median of the posterior from the full four-parameter fitting scenario is much lower than the measured stacked spectrum due to the relatively low signal-to-noise ratio, and a higher detection significance would naturally exclude the parameter space of extremely low H I density. It is expected that current and forthcoming observations by MeerKLASS will achieve this noise requirement. However, the biasing from the posterior projection effects will persist, and improvements in the data analysis to lower the systematics are needed to resolve this issue.

Due to the strong degeneracy, we can put mildly tighter priors on  $\Omega_{\text{HI}}$  and obtain a constraint on  $A_{\text{sys}}$ . We present the model fitting results for assuming a prior of  $\Omega_{\text{HI}} \in [5 \times 10^{-4}, 0.05]$  in Figure 25. In this case, a constraint on the systematics amplitude is achieved at  $A_{\text{sys}} = 0.09^{+0.09}_{-0.05}$ . However, as we have discussed in Section 7.3, the constraint cannot be directly converted to a measurement of the beam ripple. Future work should be focused on understanding the propagation of the beam ripple to the map-level systematics, and using the measurements of the beam ripple to put informative priors on  $A_{\text{sys}}$ , which will enable accurate measurements of  $\Omega_{\text{HI}}$ .



**Figure 25.** The same as the four-parameter case in Figure 21, but with an additional prior of  $\Omega_{\text{HI}} \in [5 \times 10^{-4}, 0.05]$ . The posterior of  $\bar{M}_{\text{HI}}$  has been converted to  $\Omega_{\text{HI}}$ . The outer, middle, and inner contours denote the  $3\sigma$ ,  $2\sigma$ , and  $1\sigma$  regions.

We note that, throughout this paper, the cosmological model is fixed. The HI signal and the covariance estimation are cosmology-dependent, and varying the HI galaxy clustering statistics is not discussed, as we do not have observation-driven priors on the HI power spectrum. For future data analysis of MeerKLASS survey and SKA-Mid, it is expected that the detection of HI auto-power will allow for a more thorough analysis of the impact of cosmology on the HI stacking. Moreover, if the modeling of the signal is good enough to estimate a covariance among the different summary statistics, we can jointly model the HI stacking, the cross-power spectrum, and the HI power spectrum to consistently

marginalize over the clustering model parameters. It can also maximize the science output, as additional information on the HI density helps break the degeneracy between the amplitude of the matter power and the average brightness temperature.

We have shown that the covariance estimation in this work can be further improved. If a better constraint of the amplitude of the oscillating systematics can be achieved in the future, we can simulate mock observations with the systematics to build the covariance. Furthermore, the systematics can be expressed as an operator in the data vector, and the data covariance can be modeled analytically in the quadratic estimator formalism together with the PCA cleaning matrix (see, e.g.,

N. S. Kern & A. Liu 2021; Z. Chen et al. 2023b). The quadratic estimator can in turn be used to reconstruct the systematics and clean the data, as suggested in H. Wang et al. (2022).

For the mock simulation, we have ignored the redshift evolution of the stacking subsample, by assuming a constant redshift kernel and randomly assigning H I mass from a fixed HIMF. While it is sufficient for our work, given the narrow sub-band and the signal-to-noise ratio, future work should consider the redshift distribution of the sources, as well as the impact of the magnitude limit on the H I mass distribution of the selected targets.

In this work, we have observed that the systematics modulate the covariance of the data for both the noise and the H I signal. Therefore, in power spectrum analysis, the impact of systematics needs to be considered for modeling the covariance as well. In MK25, transfer function realizations are used to calculate the covariance of the power spectrum. Since the mock realizations are injected into the data and then PCA cleaned, we expect that the systematics are included in the covariance estimation. However, similar to the shuffling covariance in this work, the systematic effects may be distorted in covariance estimation. The impact of covariance estimation with the existence of systematics needs to be carefully studied in the future for the power spectrum analysis.

The properties of the systematics obtained from this work have implications for the power spectrum analysis. For instance, we find that the secondary peaks sourced by systematics extend up to  $\sim 15\text{--}20$  MHz, corresponding to  $75\text{--}100$  Mpc; hence, it is expected that the measured power spectrum at the corresponding Fourier modes  $k_{\parallel} \sim 0.08 \text{ Mpc}^{-1}$  may be affected by these systematics. In the ongoing data analysis (MeerKLASS collaboration 2025, in preparation), we adopt several techniques that can mitigate the impact of systematics on the power spectrum estimation. First, we adopt a multiscale ‘‘mPCA’’ technique where the large and small scales of the intensity maps are cleaned separately, as outlined in I. P. Carucci et al. (2024). Second, we perform internal cross correlation, i.e., cross-correlating the H I intensity maps obtained from different observation blocks and/or different dishes. In Section 7.3, we discussed how the beam ripple may cause calibration errors that lead to the systematics in the data. The calibration solutions are time- and dish-dependent, and the chromatic errors may be reduced through cross-correlating data sets taken from different blocks and dishes. Finally, in the averaging of the 1D power spectrum, we exclude the small  $k_{\parallel}$  regions that contain the peak.

This work sets the path to the measurement of one-point statistics from line-intensity maps. One-point statistics are estimators of the probability distribution function of the measured brightness temperature (P. C. Breysse et al. 2017; J. L. Bernal 2024) and are very sensitive to the line luminosity function. Correctly combined with the power spectrum (H. T. Ihle et al. 2019; G. Sato-Polito & J. L. Bernal 2022), they can break the degeneracies between astrophysical and cosmological parameters, boosting the constraining power of line-intensity mapping (P. C. Breysse 2022; V. I. Sabla et al. 2024). However, these measurements are very prone to being contaminated by observational systematics: compared with stacking, they do not involve the average around resolved sources and actually account for the full temperature distribution, rather than the mean at two points. Nonetheless,

they have in common that they directly use the line-intensity maps. Although there are proposals to deal with the foreground contamination in one-point statistics using one-point cross correlations (P. C. Breysse et al. 2023; D. T. Chung et al. 2023) or conditional statistics (P. C. Breysse et al. 2019), the characterization of the systematic errors at map level that can be obtained from analysis like the one presented in this work will be key for a correct measurement of one-point statistics from the observations.

## 11. Conclusion

In this paper, we present a comprehensive data analysis and validation pipeline for performing emission-line stacking of MeerKLASS *L*-band deep-field survey intensity maps onto the positions of GAMA spectroscopic galaxies.

The MeerKLASS *L*-band intensity maps at  $z \sim 0.4$  were observed in single-dish mode using the MeerKAT telescope, with a primary beam of  $\sim 1$  deg. The comoving transverse scale of  $\sim 30$  Mpc of the beam is larger than the typical scales of dark matter halos, complicating the stacked signal of H I emission. Meanwhile, the overlapping GAMA G23 field has a number density of galaxies that is much lower than the expected number density of H I galaxies, making it highly incomplete for the stacking purposes. We build a mock pipeline to generate the population of H I galaxies and simulate the stacking of the intensity maps onto a GAMA-like galaxy sample, and find the following:

1. Due to the large physical scale of the primary beam, the stacked cubelet of each source contains contributions from emissions from other sources, as the H I emission is extended along the line of sight by the velocity dispersion and extended along the angular plane by the primary beam. As a result, the stacked cube contains extra H I emission for all voxels within, leading to severe double counting. At the same time, the central region of the stacked cube has an excess H I signal against the background, which is the desired H I emission from the target sources.
2. The extra emission from double counting is removed by the PCA cleaning procedure used to clean the foregrounds. When collapsed along the frequency direction into an angular image, the stacked signal after PCA cleaning shows no extra emission away from the center pixel. The excess H I signal in the central area follows the shape of the primary beam.
3. The stacked signal can also be averaged along the angular plane into a stacked spectrum. Comparing the input H I emission-line profile with the output stacked spectrum, we find that the amplitude of the stacked signal is massively amplified by the extra emission from double counting and the clustering of H I sources. The amplified signal then suffers signal loss from PCA cleaning. The differences in the final stacked spectrum compared to the input require forward modeling.
4. Given the depth of the MeerKLASS *L*-band deep-field survey, we find that the stacked measurement is feasible, and the H I signal can be detected with high statistical significance.

The viability of the stacking detection and the requirement for forward modeling call for detailed study into the modeling

of the signal and its covariance. We run 100 independent realizations of the mock observation and find that:

1. The HI signal in the stacked spectrum can be forward modeled under a simplified assumption, by assuming that the GAMA galaxy sample contains all of the HI mass in the survey volume and the HI mass is evenly distributed among the galaxies. Using this assumption to rerun the stacking simulation while keeping the overall HI density identical, we find that the differences between the stacked spectra of the original simulation and the simplified forward modeling are within the variance among different realizations.
2. The mock covariance can be calculated using the realizations. We find that, for the stacked image, the pixels in the central region are highly correlated due to the signal covariance, as the HI signal is smoothed by the primary beam. The covariance of the stacked spectrum also shows correlation between different velocity channels, due to the double counting as the same intensity map pixel that will be sampled multiple times into different channels.
3. Due to the complicated correlation of the stacked signal and the systematics in the data, a method of covariance estimation based on the data is needed. We test covariance estimation using random shuffling of galaxy positions. Each random shuffle produces a reference stacked cube that is consistent with null detection on average, and the covariance of the reference signal can be used to estimate the mock covariance. We find that the shuffling estimate correctly reproduces the variance away from the central region of the stacked cube where the noise variance dominates, but fails to take into account the HI signal variance near the center.
4. We find that the resulting correlation matrix in the stacked image does not match the true correlation in the mock, as the shuffling does not contain an excess signal convolved with the primary beam. For the stacked spectrum, the correlation matrices match relatively well, as long as the stacked spectrum is symmetrized so that only positive values of  $\Delta\nu$  are considered. We can construct a corrected covariance estimate by using the correlation matrix from the shuffling and rescaling the variance based on the mock.

We then use the MeerKLASS intensity mapping data to perform the stacking analysis and find that:

1. A stacking signal is detected in both the angular stacked image and the stacked spectrum. In the angular stacked image, we find that there is an excess signal in the central region, consistent with the primary beam. Using the corrected covariance estimate, we find the detection significance to be  $8.66\sigma$ .
2. The stacked spectrum shows a clear excess signal peak around  $\Delta\nu \sim 0$ , while also having a clear systematics component with oscillating features with a period of  $\sim 20$  MHz. The detection significance is found to be  $7.45\sigma$  for the unsymmetrized stacked spectrum and  $5.29\sigma$  for the symmetrized spectrum.

The stacking measurement reveals a clear feature of oscillation in the spectral direction. We investigate in detail the origin of the systematics and conclude the following:

1. The systematics most likely originates from the chromaticity of the beam. The diffractive interference between the primary and secondary reflector of the MeerKAT telescope modulates the beam, causing a ripple of the beam size across frequencies.
2. In the stacked spectrum, the systematics are of the same order of magnitude as the HI signal.
3. The systematics is not an additive component of the data vector. Using the random shuffling as a null test, we do not find any feature of the systematics in the reference stacked signal.
4. The systematics not only modulate the data vector of the stacked spectrum, they also change the data covariance. The covariance estimation shows that in the reference image and the reference spectrum, the data vector is correlated in a way that is not seen in the mock simulation. Therefore, the systematics must have been convolved with the map data, which then contributes to higher-order statistics in the covariance.
5. By comparing the structure of the beam ripple and the stacked spectrum in Fourier space, we find that the peaks of the two overlap with each other, which shows that the systematics and the beam ripple share the same oscillating frequency.
6. The beam ripple induces systematics into the map data in a convoluted way that requires further investigation in the calibration and map-making pipeline. If the chromaticity is simply due to the beam convolving with the sky signal, we should not expect the noise covariance to be affected by the systematics. However, the shuffling covariance using the noise-dominated reference signal shows a clear imprint of systematics. We conclude that the systematic effects happen at stages of data processing prior to the foreground cleaning. The structure of the systematics is then introduced to the eigenmodes of the frequency–frequency covariance, which then affects the PCA cleaning. The PCA cleaning matrix is then operated on the entire data vector including the noise.
7. The systematics can be modeled effectively by convolving an error function with the map data along the frequency direction. The oscillating feature of the systematics can be parameterized in Fourier space, with varying amplitude, oscillating frequency, and the shape of the oscillation.

Including the systematics in the forward modeling, we perform Bayesian inference on the stacked spectrum using the importance nested sampling technique. We impose wide flat priors and use the posterior of the parameter fitting to conclude the following:

1. The model fitting routine gives constraints on the average HI mass of the GAMA galaxies under the simplified assumption, which can be converted to an effective constraint on the HI density in the survey volume. The amplitude of the systematics is not well constrained, with the posterior occupying the wide prior volume.
2. By varying the number of free parameters in the systematics modeling, we find that the constraints on the HI density are consistent under different modeling complexity. The Bayesian evidence for using the shape of the beam ripple and fitting just the amplitude of the systematics is consistent with the case of varying the

oscillation frequency. It further supports the fact that the beam ripple describes the systematics well.

3. The fitting gives a constraint on the oscillating frequency of the systematics that is consistent with the beam ripple reported in previous literature. If the full shape of the systematics is varied instead of just the frequency, the constraining power degrades while the posterior of the oscillation frequency is consistent, suggesting the robustness of the constraints. The conservative case gives an estimation of the frequency to be  $\nu_{\text{sys}} = 17.90^{+6.53}_{-4.27}$  MHz.
4. The estimation of the HI density is found to be lower than expected based on the measurements at similar redshifts. We find that the covariance estimation impacts the estimation of the systematics amplitude. Due to the strong degeneracy between the systematics amplitude and the HI density, the deviation of the estimated covariance and the true covariance may contribute to the underestimation.
5. In the parameter space of the systematics amplitude and the HI density, the orientation of the degeneracy changes with the systematics amplitude, leading to strong posterior projection effects. In the posterior of the fitted stacked spectrum, we find that the best-fit spectrum is near the boundary of the 95% confidence interval of the posterior at the secondary peaks of the systematics. Due to the limited signal-to-noise ratio, there is a lack of constraining power in the systematics amplitude, and therefore, the constraints on the HI density are not robust.

Our findings provide strong incentive to include the stacking analysis in the cosmological analysis of the HI intensity mapping data in cross correlation with optical galaxies. The stacking measurement is a powerful tool to validate the detection of the HI signal and examine the quality of the data for residual systematics. The stacked spectrum can be modeled to infer the HI density, providing a unique window for measuring the evolution of cosmic HI across different redshifts using future HI intensity mapping data. As the data quality and depth improve for future MeerKLASS survey and SKAO, we expect that the stacking analysis can help disentangle the HI density and the matter clustering amplitude, while also allowing for the modeling of residual systematics as nuisance parameters in the cosmological analysis. It will serve as a robust summary statistic to maximize the constraining power of the data, and our work provides the first analysis of its kind as a starting point to build robust inference methods toward future SKAO.

### Acknowledgments

J.W. acknowledges support from the National SKA Program of China (No. 2020SKA0110100). Z.C. and A.P. are funded by a UKRI Future Leaders Fellowship [grant MR/X005399/1; PI: Alkistis Pourtsidou]. S.Cu. acknowledges support from the UKRI Stephen Hawking Fellowship (grant reference EP/U536751/1) and was also supported by a UKRI Future Leaders Fellowship grant [MR/V026437/1]. J.L.B. acknowledges funding from the Ramón y Cajal grant RYC2021-033191-I, financed by MCIN/AEI/10.13039/501100011033 and by the European Union “NextGenerationEU”/PRTR, as well as the project UC-LIME (PID2022-140670NA-I00), financed by MCIN/AEI/10.13039/

501100011033/FEDER, UE. I.P.C. is supported by the European Union within the Next Generation EU program [PNRR-4-2-1.2 project No. SOE\_0000136, RadioGaGa]. J.F. acknowledges support of Fundação para a Ciência e a Tecnologia through the Investigador FCT Contract No. 2020.02633.CEECIND/CP1631/CT0002, and the research grants UIDB/04434/2020 and UIDP/04434/2020. M.G.S. acknowledges support from the South African Radio Astronomy Observatory and National Research Foundation (grant No. 84156). The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation. We acknowledge the use of the Ilifu cloud computing facility, through the Inter-University Institute for Data Intensive Astronomy (IDIA).

We thank Ludwig Schwarzd, Mattieu de Villiers, and Dirk de Villiers for discussions on the primary beam of the MeerKAT telescope.

*Facility:* MeerKAT.

*Software:* NUMPY (C. R. Harris et al. 2020), SCIPY (P. Virtanen et al. 2020), ASTROPY (Astropy Collaboration et al. 2022), MATPLOTLIB (J. D. Hunter 2007).

### Appendix

#### Impact of the Covariance Estimation on the Systematics

In this appendix, we briefly discuss the limits of the covariance estimation routine presented in Section 5. Specifically, we aim to examine the impact of the mock-corrected covariance in Equation (32).

The stacked signal can be expressed as a data vector  $\mathbf{d}$ . In the mock, the true stacked signal can be written as a combination of the HI signal and noise,

$$\mathbf{d}_{\text{mock}} = \mathbf{d}_{\text{mock}}^{\text{HI}} + \mathbf{d}_{\text{mock}}^{\text{n}}, \quad (\text{A1})$$

assuming that the foreground has been sufficiently removed. The true mock covariance can then be expressed as a combination of the HI covariance and the noise covariance,

$$\langle \mathbf{d}_{\text{mock}} \mathbf{d}_{\text{mock}}^{\text{T}} \rangle = \mathbf{C}_{\text{HI}}^{\text{mock}} + \mathbf{C}_{\text{n}}^{\text{mock}}, \quad (\text{A2})$$

where  $\langle \rangle$  denotes the assemble average.

On the other hand, the shuffled data vector  $\mathbf{d}_{\text{shuffle}}$ , on average, contains only the noise component. Assuming that the shuffling, on average, reflects the sampling of the pixels of the real galaxy catalog, then the covariance of the shuffled data vector is simply

$$\mathbf{C}^{\text{mock,shuffle}} = \langle \mathbf{d}_{\text{mock,shuffle}} \mathbf{d}_{\text{mock,shuffle}}^{\text{T}} \rangle = \mathbf{C}_{\text{n}}^{\text{mock}}. \quad (\text{A3})$$

The discrepancy between the two covariances is in the amplitude as well as the correlation, as we have discussed in Section 5.2. In Equation (32), we defined a correction, which can be written as

$$\mathbf{R} = \text{diag}[\mathbf{r}], \quad (\text{A4})$$

$$\mathbf{r}_i = \sqrt{\mathbf{C}_{ii}^{\text{mock}} / \mathbf{C}_{ii}^{\text{mock,shuffle}}}, \quad (\text{A5})$$

$$\hat{\mathbf{C}}^{\text{mock}} = \langle \mathbf{R} \mathbf{d}_{\text{mock,shuffle}} \mathbf{d}_{\text{mock,shuffle}}^{\text{T}} \mathbf{R}^{\text{T}} \rangle = \mathbf{R} \mathbf{C}_{\text{n}}^{\text{mock}} \mathbf{R}^{\text{T}}, \quad (\text{A6})$$

where  $\text{diag}[\mathbf{r}]$  denotes a diagonal matrix with  $\mathbf{r}$  as its diagonal elements. It is straightforward to see that the diagonal elements of  $\hat{\mathbf{C}}^{\text{mock}}$  are equal to those of the true covariance  $\mathbf{C}^{\text{mock}}$ , whereas the correlation follows  $\mathbf{C}^{\text{mock,shuffle}}$  and therefore

$C_n^{\text{mock}}$ . This introduces a slight underestimation of correlation at intermediate intervals of  $\Delta\nu$ , as can be seen in Figure 13.

We then apply the correction to the data. The data vector can be written as a multiplicative systematic operator  $S$  on the underlying HI and noise data<sup>25</sup>

$$\mathbf{d}_{\text{data}} = S(\mathbf{d}_{\text{data}}^{\text{HI}} + \mathbf{d}_{\text{data}}^{\text{n}}). \quad (\text{A7})$$

The covariance of the data is then

$$C^{\text{data}} = \langle \mathbf{d}_{\text{data}} \mathbf{d}_{\text{data}}^{\text{T}} \rangle = S(C_{\text{HI}}^{\text{data}} + C_n^{\text{data}})S^{\text{T}}. \quad (\text{A8})$$

The reference stacked signal  $\mathbf{d}_{\text{data,shuffle}}$ , on the other hand, contains only the systematics and the noise. The estimated covariance matrix can then be written as

$$\hat{C}^{\text{data}} = \langle \mathbf{R} \mathbf{d}_{\text{data,shuffle}} \mathbf{d}_{\text{data,shuffle}}^{\text{T}} \mathbf{R}^{\text{T}} \rangle = \mathbf{R} S C_n^{\text{data}} S^{\text{T}} \mathbf{R}^{\text{T}}. \quad (\text{A9})$$

We can rewrite  $\hat{C}^{\text{data}}$  so that

$$\hat{C}^{\text{data}} = \mathbf{R} S \mathbf{R}^{-1} \mathbf{R} C_n^{\text{data}} \mathbf{R}^{\text{T}} (\mathbf{R}^{\text{T}})^{-1} S^{\text{T}} \mathbf{R}^{\text{T}} = \tilde{S}_R \mathbf{R} C_n^{\text{data}} \mathbf{R}^{\text{T}} \tilde{S}_R^{\text{T}}, \quad (\text{A10})$$

where we have defined a new matrix  $\tilde{S}_R = \mathbf{R} S \mathbf{R}^{-1}$ . Assuming that the mock correctly reflects the amplitude of the HI and noise signal in the data,  $\mathbf{R} C_n^{\text{data}} \mathbf{R}^{\text{T}} \approx C_{\text{HI}}^{\text{data}} + C_n^{\text{data}}$ , which is the target data covariance without the systematics. Comparing Equation (A8) with Equation (A10), we can see that the covariance estimate indeed includes the systematics. However, the systematics are distorted by the correction matrix  $\mathbf{R}$ .

Note that  $\mathbf{R}$  is diagonal, so that

$$(\tilde{S}_R)_{ij} = \frac{r_i}{r_j} S_{ij}. \quad (\text{A11})$$

The values of  $r$  are consistent with 1 at large values of  $|\Delta\nu|$  and larger than 1 at small values, as we have shown in Figure 12. As a result,  $\tilde{S}_R$  has a mismatch with the true  $S$ , therefore biasing the covariance and the subsequent inference.

As we have no prior knowledge on the amplitude of the systematics, we resort to using the shuffling for covariance estimation as a way of including the systematics blindly in the data. Alternatively, if no correction is made, the underlying data covariance is distorted, which biases the covariance and the inference differently. By quantifying the differences between the two covariances, we can examine the effect of the imperfect covariance estimation, which we discuss in Section 9.2.

## ORCID iDs

Zhaoting Chen (陈兆庭) <https://orcid.org/0000-0002-4965-8239>

Steven Cunnington <https://orcid.org/0000-0001-6594-107X>

Alkistis Pourtsidou <https://orcid.org/0000-0001-9110-5550>

Laura Wolz <https://orcid.org/0000-0003-3334-3037>

Marta Spinelli <https://orcid.org/0000-0003-0148-3254>

José Luis Bernal <https://orcid.org/0000-0002-0961-4653>

Matilde Barberi-Squarotti <https://orcid.org/0009-0007-8964-5807>

Stefano Camera <https://orcid.org/0000-0003-3399-3574>

Isabella P. Carucci <https://orcid.org/0000-0001-5287-0065>

José Fonseca <https://orcid.org/0000-0003-0549-1614>

Keith Grainge <https://orcid.org/0000-0002-6780-1406>

Melis O. Irfan <https://orcid.org/0000-0003-2021-7357>

Mario G. Santos <https://orcid.org/0000-0003-3892-3073>

Jingying Wang (王婧颖) <https://orcid.org/0000-0002-5598-2668>

## References

- Abdalla, E., Ferreira, E. G. M., Landim, R. G., et al. 2022, *A&A*, 664, A14
- Alonso, D., Ferreira, P. G., & Santos, M. G. 2014, *MNRAS*, 444, 3183
- Anderson, C. J., Luciw, N. J., Li, Y. C., et al. 2018, *MNRAS*, 476, 3382
- Asad, K. M. B., Girard, J. N., de Villiers, M., et al. 2021, *MNRAS*, 502, 2970
- Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, *ApJ*, 935, 167
- Baldry, I. K., Robotham, A. S. G., Hill, D. T., et al. 2010, *MNRAS*, 404, 86
- Barry, N., Hazelton, B., Sullivan, I., Morales, M. F., & Pober, J. C. 2016, *MNRAS*, 461, 3135
- Battye, R. A., Davies, R. D., & Weller, J. 2004, *MNRAS*, 355, 1339
- Bera, A., Kanekar, N., Chengalur, J. N., & Bagla, J. S. 2022, *ApJL*, 940, L10
- Bernal, J. L. 2024, *PhRvD*, 109, 043517
- Bernal, J. L., & Kovetz, E. D. 2022, *A&ARv*, 30, 5
- Beutler, F., Blake, C., Colless, M., et al. 2011, *MNRAS*, 416, 3017
- Bharadwaj, S., Nath, B. B., & Sethi, S. K. 2001, *JApA*, 22, 21
- Bianchetti, A., Sinigaglia, F., Rodighiero, G., et al. 2025, *ApJ*, 982, 82
- Bosman, S. E. I., Davies, F. B., Becker, G. D., et al. 2022, *MNRAS*, 514, 55
- Breyse, P. C. 2022, arXiv:2209.01223
- Breyse, P. C., Anderson, C. J., & Berger, P. 2019, *PhRvL*, 123, 231105
- Breyse, P. C., Chung, D. T., & Ihle, H. T. 2023, *MNRAS*, 525, 1824
- Breyse, P. C., Kovetz, E. D., Behroozi, P. S., Dai, L., & Kamionkowski, M. 2017, *MNRAS*, 467, 2996
- Carucci, I. P., Bernal, J. L., Cunnington, S., et al. 2024, arXiv:2412.06750
- Carucci, I. P., Irfan, M. O., & Bobin, J. 2020, *MNRAS*, 499, 304
- Chang, T.-C., Pen, U.-L., Peterson, J. B., & McDonald, P. 2008, *PhRvL*, 100, 091303
- Chen, Z., Chapman, E., Wolz, L., & Mazumder, A. 2023a, *MNRAS*, 524, 3724
- Chen, Z., Wolz, L., & Battye, R. 2023b, *MNRAS*, 518, 2971
- Chen, Z., Wolz, L., Spinelli, M., & Murray, S. G. 2021, *MNRAS*, 502, 5259
- CHIME Collaboration, Amiri, M., Bandura, K., et al. 2022, *ApJS*, 261, 29
- CHIME Collaboration, Amiri, M., Bandura, K., et al. 2023, *ApJ*, 947, 16
- Chung, D. T., Bangari, I., Breyse, P. C., et al. 2023, *MNRAS*, 520, 5305
- Coles, P., & Jones, B. 1991, *MNRAS*, 248, 1
- Cooray, A., & Sheth, R. 2002, *PhR*, 372, 1
- Crichton, D., Aich, M., Amara, A., et al. 2022, *JATIS*, 8, 011019
- Cunnington, S., Irfan, M. O., Carucci, I. P., Pourtsidou, A., & Bobin, J. 2021, *MNRAS*, 504, 208
- Cunnington, S., Li, Y., Santos, M. G., et al. 2023a, *MNRAS*, 518, 6262
- Cunnington, S., Wolz, L., Bull, P., et al. 2023b, *MNRAS*, 523, 2453
- de Villiers, D. I. L. 2013, *ITAP*, 61, 2457
- Driver, S. P., Bellstedt, S., Robotham, A. S. G., et al. 2022, *MNRAS*, 513, 439
- Driver, S. P., Norberg, P., Baldry, I. K., et al. 2009, *A&G*, 50, 5.12
- Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, *MNRAS*, 413, 971
- Dunne, D. A., Cleary, K. A., Breyse, P. C., et al. 2024, *ApJ*, 965, 7
- Dunne, D. A., Cleary, K. A., Breyse, P. C., et al. 2025, arXiv:2503.21743
- Engelbrecht, B. N., Santos, M. G., Fonseca, J., et al. 2025, *MNRAS*, 536, 1035
- Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2019, *OJAp*, 2, 10
- Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, *PhR*, 433, 181
- Gogate, A. 2022, PhD thesis, Univ. Groningen doi:10.33612/diss.206445119
- Gómez-Valent, A. 2022, *PhRvD*, 106, 063506
- Grieb, J. N., Sánchez, A. G., Salazar-Albornoz, S., et al. 2017, *MNRAS*, 467, 2085
- Guo, H., Jones, M. G., Wang, J., & Lin, L. 2021, *ApJ*, 918, 53
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, 585, 357
- Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1982, *A&AS*, 47, 1
- Hellwig, H., Vessot, R. F. C., Levine, M. W., et al. 1970, *ITIM*, 19, 200
- Heywood, I., Lenc, E., Serra, P., et al. 2020, *MNRAS*, 494, 5018
- Heywood, I., Ponomareva, A. A., Maddox, N., et al. 2024, *MNRAS*, 534, 76
- Hilbert, S., Hartlap, J., & Schneider, P. 2011, *A&A*, 536, A85
- Hu, W., Hoppmann, L., Staveley-Smith, L., et al. 2019, *MNRAS*, 489, 1619

<sup>25</sup> Note that, in reality, the systematics operators on the HI data and the noise should be different, with the noise only having systematic effects through the PCA cleaning matrix. Here, for simplicity, we write them as one matrix  $S$ . Note that the derivation for the distortion of covariance is not affected by this simplification.

- Hunter, J. D. 2007, *CSE*, **9**, 90
- Ihle, H. T., Chung, D., Stein, G., et al. 2019, *ApJ*, **871**, 75
- Irfan, M. O., Bull, P., Santos, M. G., et al. 2022, *MNRAS*, **509**, 4923
- Jarvis, M., Taylor, R., Agudo, I., et al. 2016, in Proc. of Science 277, MeerKAT Science: On the Pathway to the SKA, ed. R. Taylor et al. (Trieste: SISSA), 006
- Jonas, J. L., Baart, E. E., & Nicolson, G. D. 1998, *MNRAS*, **297**, 977
- Jones, M. G., Haynes, M. P., Giovanelli, R., & Moorman, C. 2018, *MNRAS*, **477**, 2
- Kaiser, N. 1987, *MNRAS*, **227**, 1
- Kern, N. S., & Liu, A. 2021, *MNRAS*, **501**, 1463
- Lange, J. U. 2023, *MNRAS*, **525**, 3181
- Lewis, A., & Challinor, A. 2007, *PhRvD*, **76**, 083005
- Lewis, A., & Challinor, A. 2011, CAMB: Code for Anisotropies in the Microwave Background, Astrophysics Source Code Library, ascl:1102.026
- Li, Y., Wang, Y., Deng, F., et al. 2023, *ApJ*, **954**, 139
- Liske, J., Baldry, I. K., Driver, S. P., et al. 2015, *MNRAS*, **452**, 2087
- Lujan Niemeyer, M., Bowman, W. P., Ciardullo, R., et al. 2022b, *ApJL*, **934**, L26
- Lujan Niemeyer, M., Komatsu, E., Byrohl, C., et al. 2022a, *ApJ*, **929**, 90
- Madau, P., Meiksin, A., & Rees, M. J. 1997, *ApJ*, **475**, 429
- Maddox, N., Frank, B. S., Ponomareva, A. A., et al. 2021, *A&A*, **646**, A35
- Masui, K. W., Switzer, E. R., Banavar, N., et al. 2013, *ApJL*, **763**, L20
- Matshawule, S. D., Spinelli, M., Santos, M. G., & Ngobese, S. 2021, *MNRAS*, **506**, 5075
- MeerKLASS Collaboration, Barberi-Squarotti, M., Bernal, J. L., et al. 2025, *MNRAS*, **537**, 3632
- Melchior, P., Joseph, R., Sanchez, J., MacCrann, N., & Gruen, D. 2021, *NatRP*, **3**, 712
- Meyer, M., Robotham, A., Obreschkow, D., et al. 2017, *PASA*, **34**, 52
- Murray, S. G. 2018, *JOSS*, **3**, 850
- Pal, S., Elahi, K. M. A., Bharadwaj, S., et al. 2022, *MNRAS*, **516**, 2851
- Pan, H., Jarvis, M. J., Allison, J. R., et al. 2020, *MNRAS*, **491**, 1227
- Pan, H., Jarvis, M. J., Santos, M. G., et al. 2023, *MNRAS*, **525**, 256
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, **641**, A6
- Ponomareva, A. A., Jarvis, M. J., Pan, H., et al. 2023, *MNRAS*, **522**, 5308
- Ponomareva, A. A., Muladzi, W., Maddox, N., et al. 2021, *MNRAS*, **508**, 1195
- Pullen, A. R., Hirata, C. M., Doré, O., & Raccanelli, A. 2016, *PASJ*, **68**, 12
- Rajohnson, S. H. A., Frank, B. S., Ponomareva, A. A., et al. 2022, *MNRAS*, **512**, 2697
- Ranchod, S., Deane, R. P., Ponomareva, A. A., et al. 2021, *MNRAS*, **506**, 2753
- 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
- Renard, P., Spinoso, D., Montero-Camacho, P., et al. 2024, *MNRAS*, **535**, 826
- Sabla, V. I., Bernal, J. L., Sato-Polito, G., & Kamionkowski, M. 2024, *PhRvD*, **110**, 023507
- Sampath, A., Crichton, D., Moodley, K., et al. 2024, arXiv:2412.09527
- Santos, M., Bull, P., Camera, S., et al. 2016, in Proc. of Science 277, MeerKAT Science: On the Pathway to the SKA, ed. R. Taylor et al. (Trieste: SISSA), 032
- Sato-Polito, G., & Bernal, J. L. 2022, *PhRvD*, **106**, 103534
- Schechter, P. 1976, *ApJ*, **203**, 297
- Sinigaglia, F., Elson, E., Rodighiero, G., & Vaccari, M. 2022a, *MNRAS*, **514**, 4205
- Sinigaglia, F., Rodighiero, G., Elson, E., et al. 2022b, *ApJL*, **935**, L13
- Sinigaglia, F., Rodighiero, G., Elson, E., et al. 2024, *MNRAS*, **529**, 4192
- SKAO Cosmology Science Working Group, Bacon, D. J., Battye, R. A., et al. 2020, *PASA*, **37**, e007
- Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, *MNRAS*, **341**, 1311
- Spina, B., Bosman, S. E. I., Davies, F. B., Gaikwad, P., & Zhu, Y. 2024, *A&A*, **688**, L26
- Spinelli, M., Carucci, I. P., Cunnington, S., et al. 2022, *MNRAS*, **509**, 2048
- Switzer, E. R., Masui, K. W., Bandura, K., et al. 2013, *MNRAS*, **434**, L46
- Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, *ApJ*, **761**, 152
- Thorne, B., Dunkley, J., Alonso, D., & Naess, S. 2017, *MNRAS*, **469**, 2821
- Tramonte, D., & Ma, Y.-Z. 2020, *MNRAS*, **498**, 5916
- Tramonte, D., Ma, Y.-Z., Li, Y.-C., & Staveley-Smith, L. 2019, *MNRAS*, **489**, 385
- Tudorache, M. N., Jarvis, M. J., Heywood, I., et al. 2022, *MNRAS*, **513**, 2168
- Tudorache, M. N., Jarvis, M. J., Ponomareva, A. A., et al. 2024, arXiv:2411.14940
- Tully, R. B., & Fisher, J. R. 1977, *A&A*, **54**, 661
- Vanderlinde, K., Liu, A., Gaensler, B., et al. 2019, The Canadian Hydrogen Observatory and Radio-transient Detector (CHORD), Canadian Long Range Plan for Astronomy and Astrophysics White Papers W028
- Villaescusa-Navarro, F., Genel, S., Castorina, E., et al. 2018, *ApJ*, **866**, 135
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Wang, H., Mena-Parra, J., Chen, T., & Masui, K. 2022, *PhRvD*, **106**, 043534
- Wang, J., Santos, M. G., Bull, P., et al. 2021, *MNRAS*, **505**, 3698
- Wang, J., Li, Y., Pan, H., et al. 2025, arXiv:2501.11872
- Weaverdyck, N., & Huterer, D. 2021, *MNRAS*, **503**, 5061
- Westmeier, T., Jurek, R., Obreschkow, D., Koribalski, B. S., & Staveley-Smith, L. 2014, *MNRAS*, **438**, 1176
- Wilensky, M. J., Irfan, M. O., & Bull, P. 2025, *MNRAS*, **539**, 3122
- Wolz, L., Pourtsidou, A., Masui, K. W., et al. 2022, *MNRAS*, **510**, 3495
- Wyithe, J. S. B., Loeb, A., & Geil, P. M. 2008, *MNRAS*, **383**, 1195
- Xi, H., Staveley-Smith, L., For, B.-Q., et al. 2021, *MNRAS*, **501**, 4550
- Zheng, H., Tegmark, M., Dillon, J. S., et al. 2017, *MNRAS*, **464**, 3486
- Zhang, T., Li, X., Dalal, R., et al. 2023, *MNRAS*, **525**, 2441
- Zhu, Y., Becker, G. D., Bosman, S. E. I., et al. 2022, *ApJ*, **932**, 76
- Zhu, Y., Becker, G. D., Bosman, S. E. I., et al. 2024, *MNRAS*, **533**, L49
- Zuo, S., Li, J., Li, Y., et al. 2021, *A&C*, **34**, 100439