

# Fermilab

## 21 cm Power Spectrum Analysis of North Celestial Pole Observations with the Tianlai Dish Pathfinder Array

FERMILAB-PUB-26-0264-T

arXiv:2604.12200

DOI: 10.1088/1674-4527/ae6170

Fermilab Accepted Manuscript

This manuscript has been authored by Fermi Forward Discovery Group, LLC under Contract No. 89243024CSC000002 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

## 21 cm Power Spectrum Analysis of North Celestial Pole Observations with the Tianlai Dish Pathfinder Array

Guangzhi He<sup>1,2</sup>, Shifan Zuo<sup>1,2</sup>, Jixia Li<sup>1,2</sup>, Yichao Li<sup>3</sup>, Furen Deng<sup>1,2</sup>, Shijie Sun<sup>1,3</sup>, Reza Ansari<sup>4</sup>, Olivier Perdereau<sup>5</sup>, Peter Timbie<sup>6</sup>, Albert Stebbins<sup>7</sup>, Ayodeji Ibitoye<sup>8,9,10</sup>, Fengquan Wu<sup>1</sup>, Yougang Wang<sup>1,2,3</sup> and Xuelei Chen<sup>1,2</sup>

- <sup>1</sup> State Key Laboratory of Radio Astronomy and Technology, National Astronomical Observatories, CAS, A20 Datun Road, Chaoyang District, Beijing, 100101, P. R. China;  
*sfzuo@bao.ac.cn, wangyg@bao.ac.cn, xuelei@cosmology.bao.ac.cn*
- <sup>2</sup> School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, P. R. China
- <sup>3</sup> Key Laboratory of Cosmology and Astrophysics (Liaoning) & College of Sciences, Northeastern University, Shenyang 110819, China
- <sup>4</sup> Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191 Gif sur Yvette, France
- <sup>5</sup> IJCLab, CNRS/IN2P3, Université Paris-Saclay, 91405 Orsay, France
- <sup>6</sup> Department of Physics, University of Wisconsin Madison, 1150 University Ave, Madison WI 53703, USA
- <sup>7</sup> Fermi National Accelerator Laboratory, P.O. Box 500, Batavia IL 60510, USA
- <sup>8</sup> Department of Physics, Guangdong Technion—Israel Institute of Technology, Shantou, Guangdong 515063, People’s Republic of China
- <sup>9</sup> Centre for Space Research, North-West University, Potchefstroom 2520, South Africa
- <sup>10</sup> Department of Physics and Electronics, Adekunle Ajasin University, P. M. B. 001, Akungba-Akoko, Ondo State, Nigeria

Received 2026 month day; accepted 202x month day

**Abstract** The Tianlai Dish Pathfinder Array (TDPA) is a radio interferometer designed to test techniques for 21 cm intensity mapping in the post-reionization universe as a means of measuring large-scale cosmic structure. Using 9 nights of observations targeting the North Celestial Pole (NCP) field, totaling approximately 107 hours of integration time, we analyze data in the frequency range 700–800 MHz (corresponding to redshift  $z \sim 0.9$ ). We do the data format conversion, radio frequency interference (RFI) flagging, calibration, imaging and point source subtraction, and foreground removal via Singular Value Decomposition (SVD). The spherically averaged power spectrum  $\Delta^2(k)$  is obtained. This work successfully establishes and validates a comprehensive data analysis framework for the TDPA. We identify key improvements including sky model refinement, increased integration time, and pipeline optimization that will enable future detection of the 21 cm signal through auto-correlation and cross-correlation with optical galaxy surveys.

**Key words:** cosmology: large-scale structure of universe — instrumentation: interferometers — radio lines: general — methods: data analysis

## 1 INTRODUCTION

Understanding the distribution of matter in the large-scale structure of the Universe is fundamental to cosmology. It enables us to test and constrain the standard  $\Lambda$ CDM cosmological model, investigate the nature of dark energy, and probe the physics of the early Universe.

Neutral hydrogen (HI) at the 21 cm radio wavelength serves as an excellent tracer of matter distribution. In the post-reionization Universe at redshifts  $z \lesssim 6$ , the remaining HI is predominantly contained within dark matter halos (Villaescusa-Navarro et al. 2018), meaning that the 21 cm radiation emitted by HI acts as a tracer of the large-scale structure. However, at these redshifts, individual HI-rich galaxies are too faint to be detected with high signal-to-noise ratio by traditional galaxy surveys. Instead, HI intensity mapping (IM) measures the integrated 21 cm emission from unresolved sources across large sky areas, providing a highly efficient method to survey vast cosmic volumes and access redshifts ranging from the local Universe to the epoch of reionization (Bharadwaj et al. 2001; Battye et al. 2004; Wyithe et al. 2008; Chang et al. 2008; Li et al. 2023; Yang et al. 2025).

However, HI intensity mapping faces significant observational challenges. The technique records all radiation entering the telescope receiver, of which only a tiny fraction originates from HI. Foreground contamination from Galactic synchrotron and free-free emission, as well as strong extragalactic point sources, exceeds the cosmological HI signal by 3–5 orders of magnitude (Mozdzen et al. 2019; Spinelli et al. 2021; Irfan et al. 2022; Matshawule et al. 2021). Furthermore, instrumental systematics can distort these foreground signals in complex ways, making their removal particularly challenging (Shaw et al. 2015; Spinelli et al. 2022). RFI from terrestrial sources presents an additional obstacle to HI measurements (Harper & Dickinson 2018; Engelbrecht et al. 2025).

Nevertheless, several experiments have reported promising preliminary results (Masui et al. 2013; Anderson et al. 2018; Collaboration et al. 2025; Paul et al. 2023; Townsend et al. 2026), while other experiments are currently ongoing or under construction and are expected to deliver results in the near future (Li et al. 2023; Yang et al. 2025; Battye et al. 2012; Newburgh et al. 2016).

The Tianlai project is an experiment designed to address these challenges in 21 cm intensity mapping (Chen 2012). It consists of two pathfinder arrays: the Tianlai cylinder pathfinder array (Li et al. 2020) and the Tianlai dish pathfinder array (TDPA, Wu et al. 2021). The present work is devoted to the data analysis of the latter. The dish array is equipped with dual linear-polarization feeds and arranged in a compact, centrally condensed configuration (Zhang et al. 2016). The beam profile of the dish antenna has been measured with a drone Zhang et al. (2021), and the cross-couplings between the antennas was studied by simulation (Kwak et al. 2024). The overall performance of the dish array was studied in (Wu et al. 2021).

The Tianlai dishes are steerable, but to mitigate possible variations in ground pick up and other environmental factors, drift scan is the favored observation mode. The North Celestial Pole (NCP) is a point of special interest for drift scan observation, as it does not move, and allows accumulation of large integration time for deep observation. A survey of the NCP region is of particular interest (Perdereau et al. 2022), and complementary to the radio survey, the optical Tianlai-WIYN survey of the NCP region is on-going (Ansari et al. 2025).

In this work, we present results from 9 nights of observations with the Tianlai Dish Pathfinder Array (TDPA), totaling approximately 107 hours of integration time on the center of the NCP field. This is only a small part of the designed NCP survey in Perdereau et al. (2022), but it may serve as a trial. We describe our complete data processing pipeline and perform power spectrum estimation to assess the current performance and systematic error budget of the instrument. This paper is organized as follows: Section 2 describes the observations and data characteristics; Section 3 details the pre-processing, calibration, and imaging procedures; Section 4 presents the foreground-cleaning, temperature-conversion, and power-spectrum analysis; and Section 5 discusses the implications and future improvements.

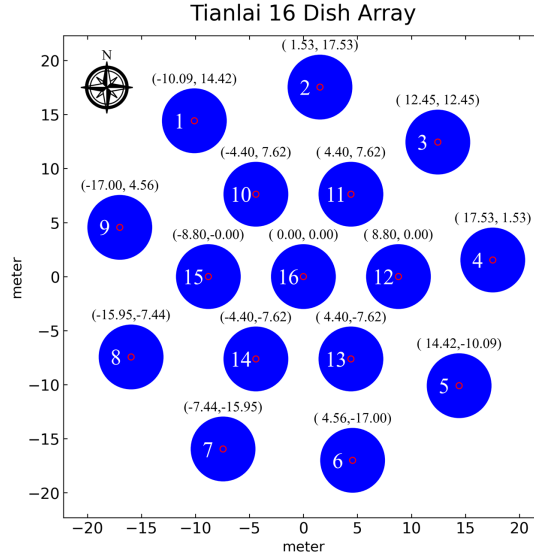


Fig. 1: Tianlai Dish Pathfinder Array layout with antenna number and ground coordinates with respect to a reference point (unit: meter).

## 2 OBSERVATIONS

The TDPA is located at a radio quiet site ( $44^{\circ}9'N$ ,  $91^{\circ}48'E$ ) in Hongliuxia, Balikun County, Xinjiang Autonomous Region, in northwest China. The array features 16 on-axis dishes, each 6 meters in diameter. As shown in Fig. 1, the antennas are arranged in a compact circular configuration, comprising one central dish surrounded by two concentric rings (Yu et al. 2022).

It operates as a transit interferometer in drift-scan mode, with the dishes pointed at a fixed direction in the meridian plane while the sky drifts through the primary beam. For a northern-hemisphere transit array, the North Celestial Pole (NCP) is the only field that can be monitored continuously, so long NCP drift scans maximize the signal-to-noise ratio for a given survey time (Wu et al. 2021). Consequently, the array has accumulated extensive drift-scan data on this field. Moreover, simulations of planned low-redshift surveys with the TDPA show that deep NCP observations achieve lower noise levels and are weakly affected by mode mixing compared to mid-latitude scans (Perdereau et al. 2022). In this work, we analyzed 9 nights of observations from the TDPA, amounting to approximately 107 hours of observation. The data were recorded over a frequency range of 700 to 800 MHz, with a spectral resolution of 0.244 MHz and a temporal resolution of 1 second. A summary of all observations used for this analysis is provided in Table 1.

## 3 DATA REDUCTION

Standard 21 cm intensity-mapping analyses typically proceed through pre-processing, calibration, imaging, foreground subtraction, and power-spectrum estimation (e.g., (Patil et al. 2017; Mertens et al. 2020)). Our pipeline for the TDPA data is summarized in Fig. 2 and comprises five stages: (1) Pre-processing, where format conversion and RFI flagging are performed; (2) Calibration, including self-calibration using sky models; (3) Imaging and point source subtraction; (4) Residual foreground subtraction; (5) Power-spectrum estimation.

Table 1: List of all observation nights analyzed in this work. The datasets for 0307 and 0309 correspond to two consecutive days, each lasting approximately 21 hours. Daytime data were removed early in the analysis, and only the analyzed data lengths are shown.

Data Set	Start time (UTC+0h)	Start time (UTC+8h)	Length (hours)
NP_20200303	2020-03-03 13:38:36	2020-03-03 21:38:36	10.0
NP_20200304	2020-03-04 13:38:56	2020-03-04 21:38:56	11.0
NP_20200306	2020-03-05 18:19:56	2020-03-06 02:19:56	6.0
NP_20200307	2020-03-07 14:59:01	2020-03-07 22:59:01	9.0
	2020-03-08 11:59:01	2020-03-08 19:59:01	12.0
NP_20200309	2020-03-09 13:21:07	2020-03-09 21:21:07	10.0
	2020-03-10 12:21:07	2020-03-10 20:21:07	11.0
NP_20200311	2020-03-11 14:09:48	2020-03-11 22:09:48	9.0
NP_20200312	2020-03-12 16:02:27	2020-03-13 00:02:27	8.0
NP_20200313	2020-03-13 14:02:08	2020-03-13 22:02:08	10.0
NP_20200314	2020-03-14 13:12:40	2020-03-14 21:12:40	11.0

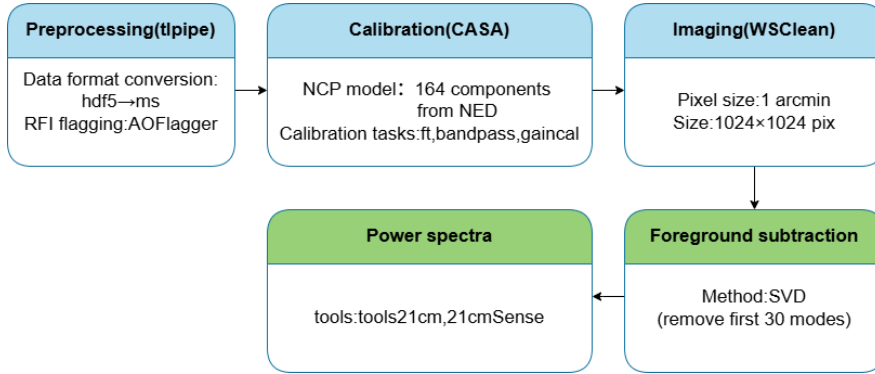


Fig. 2: Schematic diagram of the TDPA processing pipeline. Blue blocks indicate the data-processing stages (implemented mainly with `tlpipe` and standard radio-astronomy packages; Section 3), while green blocks indicate the analysis stages for foreground subtraction and power-spectrum estimation (implemented mainly with Python packages; Section 4).

### 3.1 Pre-processing

In this work we build our pipeline using the standard radio astronomical software like CASA (Team 2022) and WSClean (Offringa et al. 2014), which are well tested and easy to use in the calibration and imaging tasks for a small sky region such as the NCP, complementary to the home made data processing pipeline software called `tlpipe` (Zuo et al. 2021). The data is converted from the original HDF5 format to CASA Measurement Set files to be processed with these software tools. Fig. 3 shows the  $uv$  coverage constructed from the full dataset after data format conversion. Each sampled point represents the  $(u, v)$  coordinate of a baseline at a given time stamp, thereby illustrating the overall sampling of the interferometric measurements in the Fourier plane.

Subsequently, we performed RFI flagging using `AOFlagger` (Offringa et al. 2012). Fig. 4 displays the data quality flags for the full dataset, including the intervals when the calibration noise source (CNS), which is used to provide relative calibration in the absence of sufficiently bright point sources, was periodically injected. Edge channels are more susceptible to systematic effects, resulting in a higher percentage of flagged data. Meanwhile, Antenna 013 appears to exhibit errors across multiple observations, suggesting it may be malfunctioning or damaged, and thus all its data have been flagged.

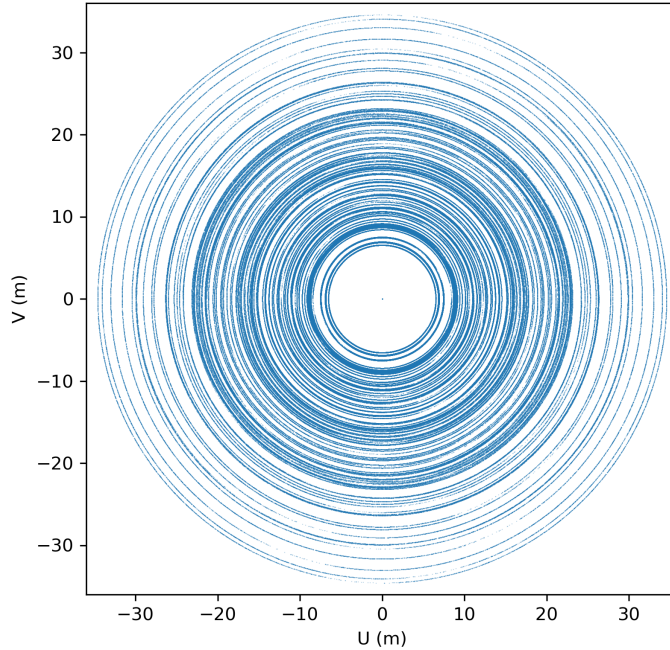


Fig. 3:  $uv$  coverage constructed from the full TDPA dataset after data format conversion from HDF5 to CASA Measurement Set. The plot shows the distribution of baseline sampling in the  $uv$  plane, where each sampled point represents the  $(u, v)$  coordinate of a baseline at a given time stamp.

### 3.2 Calibration

The calibration begins with a self-calibration (Cornwell & Wilkinson 1981; Pearson & Readhead 1984) and the main goal is to address instrumental and propagation effects. We first perform gain calibration on the visibilities with a sky model consisting of 164 bright point sources from the NCP sky model. The calibration is carried out by `ft`, `bandpass` and `gaincal` tasks in CASA, where `ft` inserts the source model into the visibility data, `bandpass` solves for the frequency-dependent complex bandpass response, and `gaincal` derives the time-dependent complex antenna gains. In Fig. 5, we present the amplitudes and phases of a representative baseline as a function of time and frequency, derived from the dataset recorded on March 4, 2020 (NP\_20200304). Prior to calibration, the raw visibilities typically exhibit significant temporal drifts and frequency-dependent ripples caused by the uncorrected instrumental response and environmental fluctuations. After applying the self-calibration, Fig. 5 demonstrates a marked improvement in stability: the phase fluctuations are largely flattened, and the amplitude variations across the bandpass are significantly suppressed.

### 3.3 Imaging

The calibrated visibilities are transformed into images and deconvolved using the multiscale CLEAN algorithm implemented in WSClean. The imaging is performed on the 700–800 MHz band of our dataset, and we use multi-frequency synthesis to combine the frequency channels into a single continuum map. To ensure that the synthesized beam is adequately sampled, we set the pixel size to 1 arcmin, providing sufficient resolution relative to the array’s longest baselines. The initial image dimensions are configured

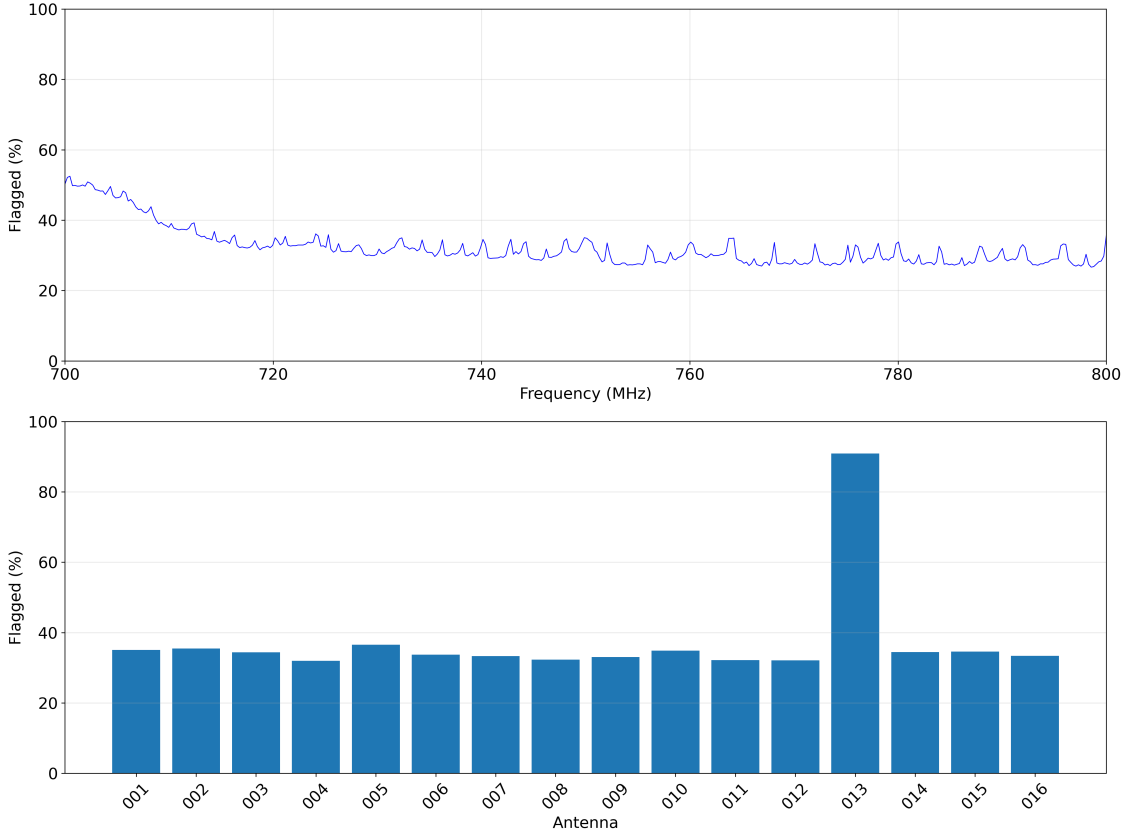


Fig. 4: Percentage of flagged data as a function of frequency channel and antenna. Edge channels show higher flagging rates due to stronger system effects. Antenna 013 exhibits consistently high flagging across multiple observations, indicating potential malfunction or damage, resulting in all its data being flagged.

to  $1024 \times 1024$  pixels, resulting in a wide field of view (FoV) of approximately  $17^\circ \times 17^\circ$ . However, retaining such a large FoV can introduce significant errors, as regions extending far beyond the main beam suffer from severe signal attenuation and low signal-to-noise ratios (SNR). In the 700–800 MHz band, the main beam’s full width at half maximum (FWHM) fluctuates between  $3.8^\circ$  and  $4.6^\circ$ , averaging around  $4.2^\circ$  (Wu et al. 2021)). To optimize the signal-to-noise ratio while minimizing the risk of excluding valid emission, we restrict our analysis to the central  $512 \times 512$  pixel region, corresponding to an  $8.5^\circ \times 8.5^\circ$  FoV, as illustrated in Fig. 6. This region is approximately twice the beam FWHM and therefore comfortably encompasses the reliable main beam response. Fig. 6 shows the imaging results at successive processing stages: the dirty image in panel (a), the deconvolved image generated with WSClean in panel (b), the residual map after source subtraction in panel (c), and the residual map after subsequent SVD cleaning in panel (d). All subsequent radio-source identification and categorization are performed within this optimized region. Specifically, the red circles in panel (b) mark the positions of the 164 bright point sources in the initial sky model used for calibration. To facilitate successive foreground removal, we further employed PyBDSF (Mohan & Rafferty 2015) for automated source extraction; the newly identified sources are indicated by cyan crosses. We evaluated the astrometric accuracy by cross-matching the PyBDSF catalog with the reference model. Our analysis reveals that the positional offsets for the majority of sources are contained within 10 arcmin. Furthermore, the distribution of these offsets

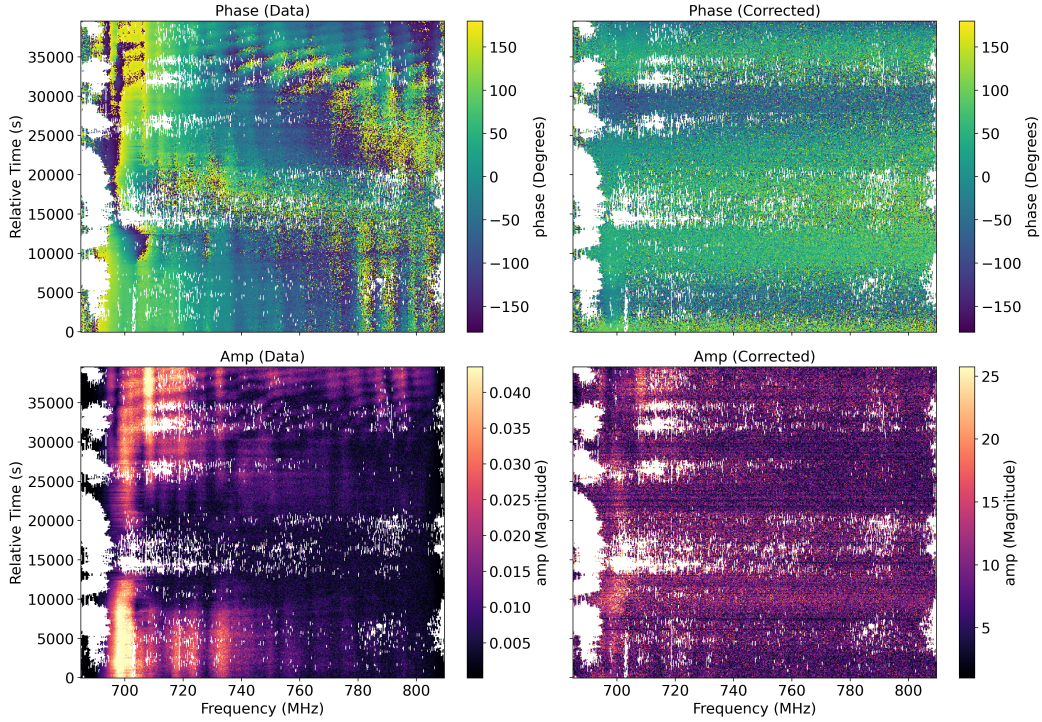


Fig. 5: Amplitude and phase of a representative baseline as a function of time and frequency, before (left) and after (right) self-calibration. The calibrated visibilities show significantly reduced phase fluctuations and amplitude variations across the bandpass, demonstrating improved stability after correction for instrumental and propagation effects.

appears random without significant systematic drift, thereby confirming the robustness of our global phase calibration.

Gain-corrected sky-model visibilities are computed after calibration by applying the gain solutions to the predicted sky-model visibilities and subsequently subtracting them from the observed visibilities. Panel (c) of Fig. 6 shows the residual image of the NCP field after source subtraction. While most sources are removed successfully, residual emission remains around the brightest sources, with flux densities ranging from  $-1.2$  to  $+0.2$  Jy/beam.

#### 4 FOREGROUND SUBTRACTION AND POWER SPECTRUM ESTIMATION

Although the calibration and point-source subtraction described in Section 3 can effectively remove the discrete sky components, the residual visibilities are still dominated by diffuse foreground emission. At the frequencies relevant to this work, the leading contaminants are Galactic synchrotron radiation and free-free emission, together with the integrated contribution from unresolved extragalactic sources. These foregrounds are typically brighter than the expected cosmological 21 cm signal by about 4–5 orders of magnitude (Santos et al. 2005; Furlanetto et al. 2006).

A crucial distinction, however, lies in their spectral behavior: astrophysical foregrounds are generally smooth functions of frequency, while the 21 cm signal varies rapidly along the line-of-sight direction due to the underlying density and ionization fluctuations. This contrast motivates a broad class of foreground-mitigation techniques that aim to identify and remove spectrally smooth components, while preserving the spectrally fluctuating cosmological signal (Liu & Shaw 2020).

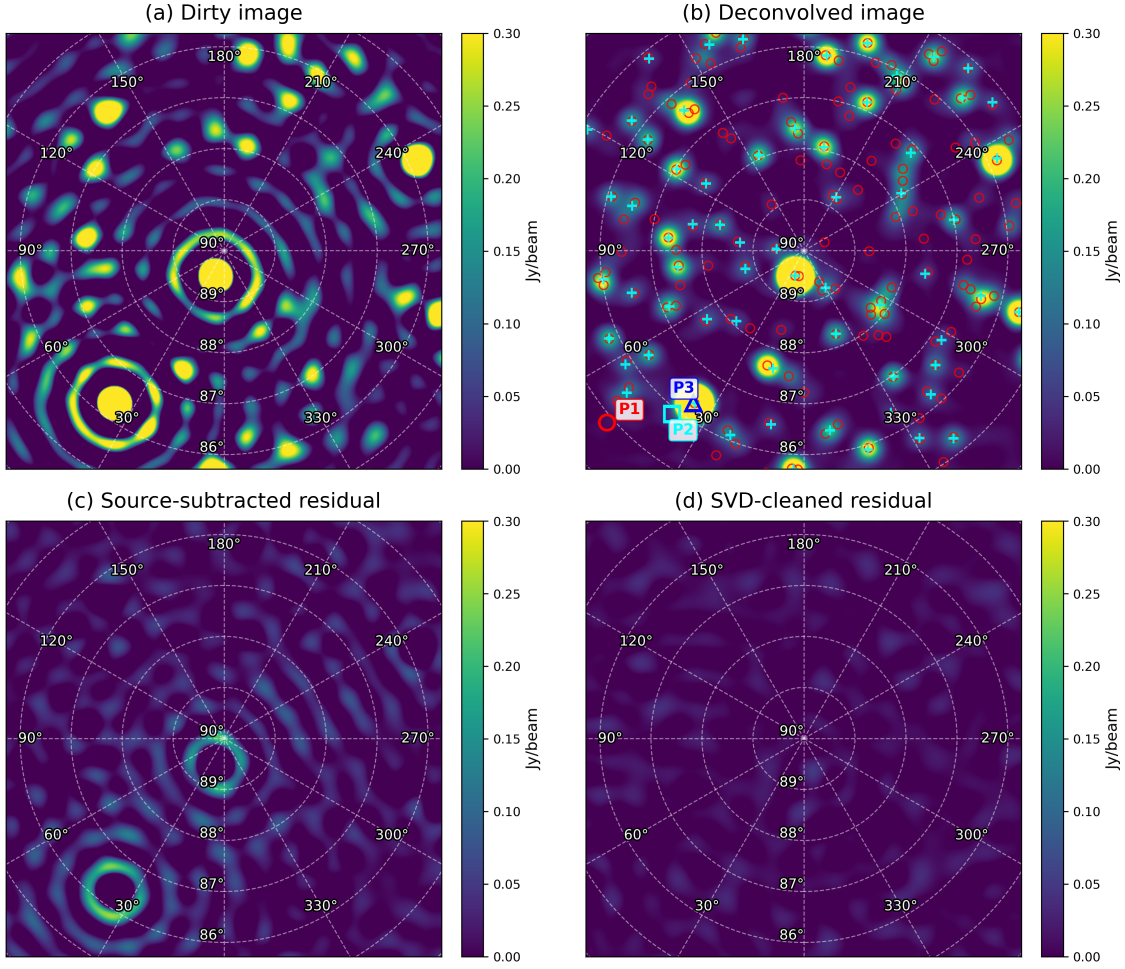


Fig. 6: Stokes I continuum images of the NCP field over 700–800 MHz. **(a)** Dirty image. **(b)** Deconvolved image generated using WSClean. Red circles denote the 164 primary calibrators, while cyan crosses mark the sources extracted by PyBDSF. The three selected pixels, P1, P2, and P3, are highlighted, and their frequency spectra are presented in the subsequent panels. **(c)** Residual map after source subtraction. **(d)** Residual map after source subtraction and SVD-based foreground removal, with the first 30 SVD modes removed.

Among the various approaches, blind component-separation methods are widely adopted in 21 cm analyses. Their main appeal is that they do not rely on an accurate parametric model of the foreground emission, nor do they require a detailed description of how the instrument and calibration residuals imprint frequency-dependent systematics onto the data. Given that our knowledge of low-frequency sky emission and its coupling to instrumental chromaticity remains incomplete, and that the intensity mapping/EoR analysis pipeline is still actively evolving, it is both practical and robust to prioritize such blind methods (Wolz et al. 2017; Anderson et al. 2018).

In this section, we first apply a Singular Value Decomposition (SVD) (Shlens 2014) based technique to identify and remove the dominant foreground modes in the residual image. We then convert the cleaned image into brightness temperature units, and finally estimate the 21 cm power spectrum from the foreground-subtracted data.

#### 4.1 SVD foreground cleaning on the map cube

We apply an SVD-based blind foreground cleaning directly to the image-domain data cube. The calibrated and imaged cube has an angular size of  $N_{\text{RA}} \times N_{\text{Dec}}$  pixels for each frequency channel, where  $N_{\text{RA}}$  and  $N_{\text{Dec}}$  denote the numbers of pixels along the right ascension (RA) and declination (Dec) directions, respectively. In our case,  $N_{\text{RA}} = N_{\text{Dec}} = 512$ , i.e. each map has  $512 \times 512$  angular pixels, and  $N_\nu = 412$ , corresponding to 700–800 MHz.

For each retained frequency channel, the 2D sky map is flattened into a one-dimensional vector, and the full cube is packed into a matrix

$$\mathbf{M} \in \mathbb{R}^{N_\nu \times N_\theta}, \quad N_\nu = 412, \quad N_\theta = N_{\text{RA}} N_{\text{Dec}} = 512 \times 512. \quad (1)$$

We compute the singular value decomposition

$$\mathbf{M} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T, \quad (2)$$

where  $\mathbf{U} \in \mathbb{R}^{N_\nu \times N_\nu}$  and  $\mathbf{V} \in \mathbb{R}^{N_\theta \times N_\theta}$  contain the left and right singular vectors and  $\mathbf{\Sigma}$  is diagonal with singular values sorted in descending order. The leading modes capture the largest-variance spectral structures, which are dominated by spectrally smooth foreground emission (and any residual smooth instrumental systematics), while the 21 cm signal and thermal noise are distributed across a much larger number of modes.

Foreground cleaning is implemented by removing the first  $N_{\text{fg}}$  modes, i.e. setting the largest  $N_{\text{fg}}$  singular values to zero and reconstructing the residual:

$$\Sigma'_{ii} = \begin{cases} 0, & i \leq N_{\text{fg}}, \\ \Sigma_{ii}, & i > N_{\text{fg}}, \end{cases} \quad \mathbf{M}_{\text{clean}} = \mathbf{U} \mathbf{\Sigma}' \mathbf{V}^T. \quad (3)$$

The cleaned matrix  $\mathbf{M}_{\text{clean}}$  is then reshaped back into a three-dimensional cube for the subsequent power spectrum estimation.

In our analysis, we find that small values of  $N_{\text{fg}}$  leave substantial residual foreground leakage, whereas excessively large values of  $N_{\text{fg}}$  remove an increasing fraction of the cosmological signal and consequently enlarge the final uncertainties (Cunnington et al. 2026). We therefore select  $N_{\text{fg}} = 30$  as a conservative compromise. This choice is guided primarily by the singular-value spectrum shown in Fig. 7. The left panel exhibits a clear transition around  $N_{\text{fg}} \sim 30$ , from a steep, foreground-dominated regime to a more slowly varying tail, suggesting that the leading smooth foreground modes are largely captured within the first  $\sim 30$  components. This interpretation is further examined using the Q–Q plot, the residual map, and the representative residual spectra discussed below.

The right panel of Fig. 7 shows a quantile–quantile (Q–Q) plot (Wilk & Gnanadesikan 1968), which provides a simple diagnostic of how closely the residual voxel distribution follows a Gaussian form. To construct this plot, we collect all finite voxel values in the SVD-cleaned residual cube, flatten them into a one-dimensional sample  $\{x_j\}_{j=1}^N$ , and sort them as  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$ . We then compute the sample mean  $\hat{\mu}$  and standard deviation  $\hat{\sigma}$ , and compare the ordered residuals  $x_{(i)}$  with the corresponding theoretical Gaussian quantiles

$$q_i = \hat{\mu} + \hat{\sigma} \Phi^{-1}(p_i), \quad p_i = \frac{i - 0.5}{N},$$

where  $\Phi^{-1}$  is the inverse cumulative distribution function of the standard normal distribution. If the residuals are Gaussian-distributed, the points  $(q_i, x_{(i)})$  should lie close to the one-to-one line  $y = x$ . Thus, the approximate linearity seen in the central part indicates that the bulk of the cleaned residuals is close to Gaussian, while the remaining deviations in the tails suggest residual non-Gaussian contamination, such as unsubtracted foreground structures, calibration artifacts, or imaging residuals. This behavior is consistent with the removal of the dominant foreground structures. As illustrated in panel (d) of Fig. 6, the residual map after SVD cleaning shows that the residuals associated with the brightest sources, which are still prominent in panel (c), have been effectively eliminated.

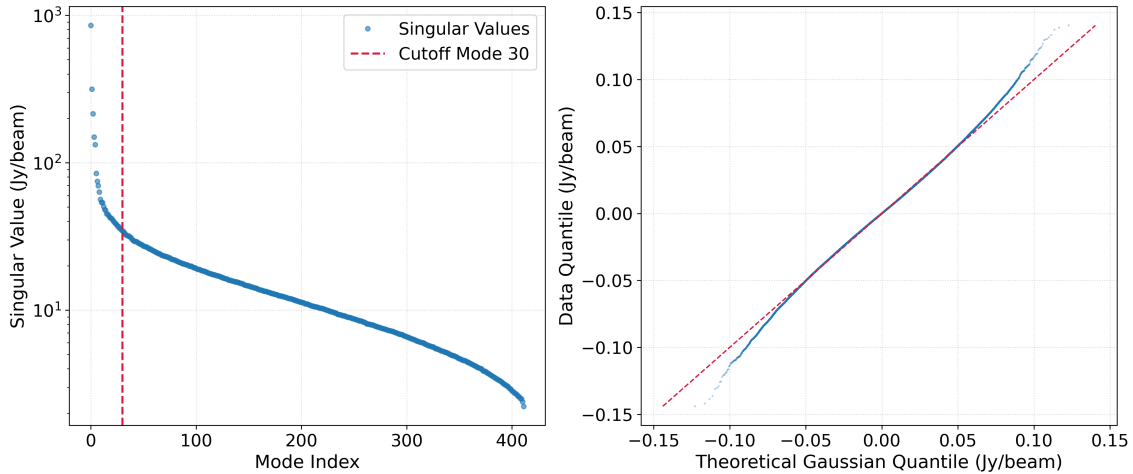


Fig. 7: Left: singular-value spectrum of the SVD foreground cleaning. The singular values decrease steeply at low mode number and begin to flatten around  $N_{\text{fg}} = 30$ , indicating a transition from dominant foreground modes to a more slowly varying tail. Right: quantile–quantile (Q–Q) plot of the residual voxel values after SVD cleaning with  $N_{\text{fg}} = 30$ , compared with a Gaussian reference distribution. The central part is approximately linear, indicating that the bulk of the residuals is close to Gaussian, while mild deviations remain in the tails.

This conclusion is also supported by the spectra extracted from the three representative pixels, P1, P2, and P3, marked in panel (b) of Fig. 6 and shown in Fig. 8. These three pixels correspond respectively to a source-free region, the edge of a source, and the center of a bright source. Before foreground removal, their spectra exhibit prominent frequency-dependent structures. For the pixels located at the source center and source edge, these features mainly reflect the spectral behavior of the bright point source itself, together with source-associated extended structures introduced by the synthesized beam and imaging response. In contrast, for the source-free pixel, the observed spectral structure is more likely due to a combination of diffuse foreground contamination, unresolved faint sources, and residual calibration/imaging systematics. After SVD cleaning with  $N_{\text{fg}} = 30$ , the spectral structures in all three cases are strongly suppressed, and the residual spectra become flatter and largely noise-like around zero, indicating effective suppression of the dominant foreground- and source-related contaminants.

## 4.2 Conversion to Brightness Temperature

The WSClean imager outputs the deconvolved Stokes- $I$  map cube in units of Jy/beam. For intensity mapping and power-spectrum estimation, it is more convenient to work in terms of brightness temperature, since the cosmological 21 cm signal is naturally described as fluctuations in  $T_{\text{b}}$  (K or mK). We therefore convert the map cube from Jy/beam to Kelvin using the Rayleigh–Jeans relation.

A pixel value in Jy/beam represents the flux density integrated over the clean beam. To convert this quantity to specific intensity  $I_{\nu}$  ( $\text{Jy sr}^{-1}$ ), we divide by the beam solid angle  $\Omega_{\text{beam}}$ . Approximating the clean beam as an elliptical Gaussian with major and minor full-width at half-maximum (FWHM)  $\theta_{\text{maj}}$  and  $\theta_{\text{min}}$ , the beam solid angle is

$$\Omega_{\text{beam}}(\nu) = \frac{\pi}{4 \ln 2} \theta_{\text{maj}}(\nu) \theta_{\text{min}}(\nu), \quad (4)$$

where  $\theta_{\text{maj}}$  and  $\theta_{\text{min}}$  are in radians. The corresponding specific intensity is then

$$I_{\nu} = \frac{S_{\nu}}{\Omega_{\text{beam}}(\nu)} \quad [\text{Jy sr}^{-1}], \quad (5)$$

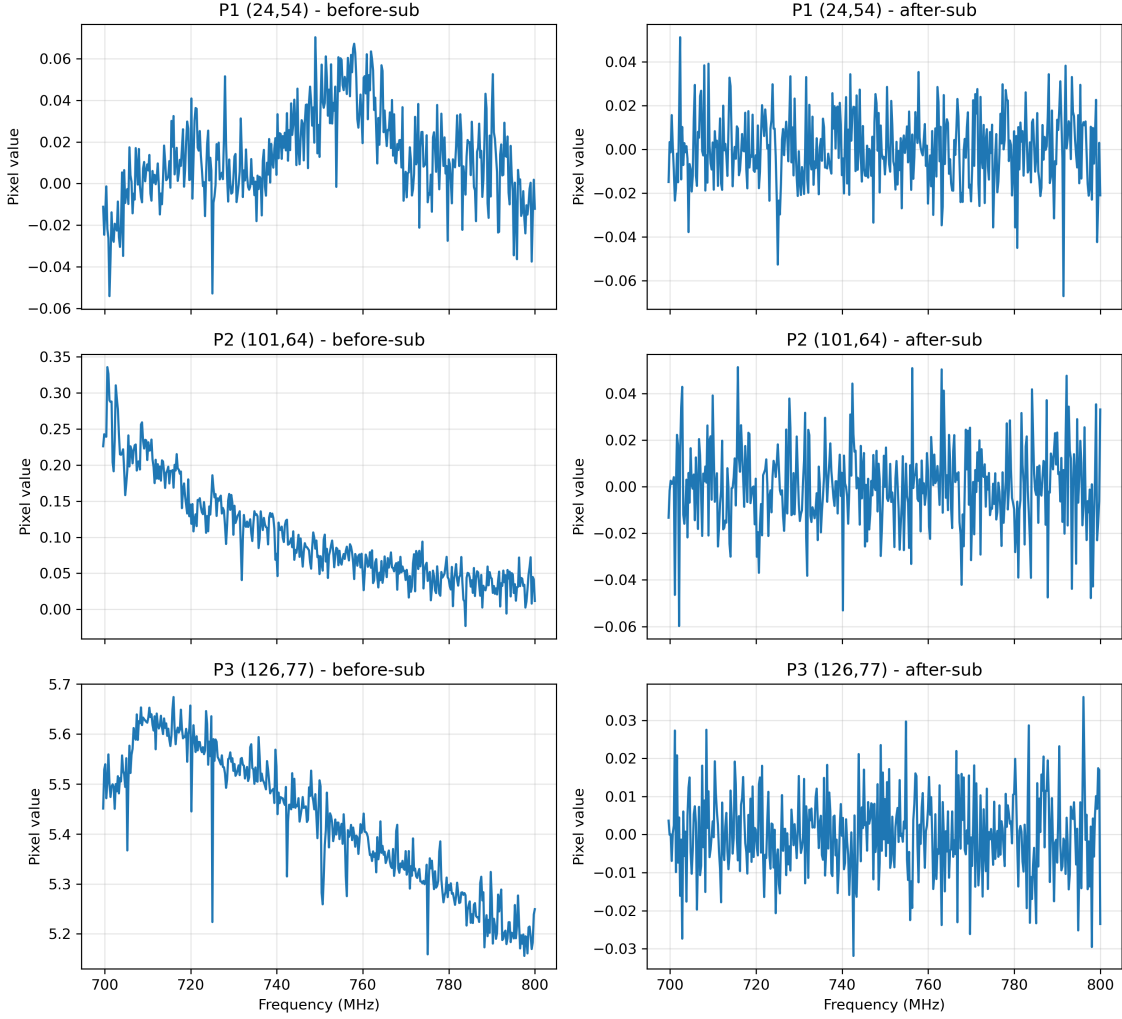


Fig. 8: Spectra extracted from the three representative pixels marked in the panel (b) of Fig. 6: P1 corresponds to a source-free region, P2 to the edge of a source, and P3 to the center of a bright source. For each pixel, the left column shows the spectrum before foreground removal, while the right column shows the corresponding residual spectrum after SVD cleaning with  $N_{\text{fg}} = 30$ . The strong suppression of the smooth spectral structures demonstrates the effectiveness of the foreground subtraction.

with  $S_\nu$  the map value in Jy/beam.

Using the Rayleigh–Jeans approximation, the brightness temperature is related to specific intensity by

$$T_b(\nu) = \frac{c^2}{2k_B\nu^2} I_\nu, \quad (6)$$

where  $c$  is the speed of light,  $k_B$  is the Boltzmann constant, and  $\nu$  is the observing frequency. Converting Jy to SI units via  $1 \text{ Jy} = 10^{-26} \text{ W m}^{-2} \text{ Hz}^{-1}$ , the full conversion applied to each frequency channel can be written as

$$T_b(\nu) = \left( \frac{S_\nu}{\Omega_{\text{beam}}(\nu)} \right) \times 10^{-26} \frac{c^2}{2k_B\nu^2}. \quad (7)$$

After this step, the map cube is expressed in Kelvin, providing the appropriate physical units for subsequent analysis and power spectrum estimation.

### 4.3 Power Spectrum Estimation

We estimate the 21 cm power spectrum from the brightness-temperature cube in Fourier space. The three-dimensional power spectrum is defined as a function of the wavevector  $\mathbf{k}$  by

$$P(\mathbf{k}) \equiv \frac{V_{\text{vox}}^2}{V_{\text{box}}} |\tilde{T}(\mathbf{k})|^2, \quad (8)$$

where  $V_{\text{box}} = L_x L_y L_z$  is the total comoving volume of the cube, and  $V_{\text{vox}} = V_{\text{box}}/N_{\text{vox}}$  is the comoving volume of a single voxel (Giri et al. 2020). Here  $N_{\text{vox}} = N_x N_y N_z$  is the total number of voxels, with  $N_x$  and  $N_y$  the transverse pixel numbers and  $N_z$  the number of frequency channels used in the analysis. The Fourier transform  $\tilde{T}(\mathbf{k})$  is computed from the discretised field  $T(\mathbf{r}_n)$  on the grid as

$$\tilde{T}(\mathbf{k}) \equiv \sum_{n=1}^{N_{\text{vox}}} T(\mathbf{r}_n) e^{-i\mathbf{k}\cdot\mathbf{r}_n}, \quad (9)$$

where  $\mathbf{r}_n$  denotes the comoving coordinate of voxel  $n$ . With the box dimensions expressed in  $\text{Mpc}/h$ , Eq. (8) yields  $P(\mathbf{k})$  in units of  $\text{K}^2(\text{Mpc}/h)^3$ .

The Fourier-space coordinates are constructed from the discrete mode indices and the box dimensions as

$$k_x = \frac{2\pi n_x}{L_x}, \quad k_y = \frac{2\pi n_y}{L_y}, \quad k_{\parallel} \equiv |k_z| = \left| \frac{2\pi n_z}{L_z} \right|, \quad (10)$$

and we define the cylindrical decomposition

$$k_{\perp} \equiv \sqrt{k_x^2 + k_y^2}, \quad k_{\parallel} \equiv |k_z|. \quad (11)$$

We then bin the three-dimensional power into a two-dimensional grid in the  $(k_{\perp}, k_{\parallel})$  plane to form the cylindrical power spectrum  $P(k_{\perp}, k_{\parallel})$ . Fig. 9 presents the cylindrically averaged power spectra of Stokes  $I$  at different steps of the processing. The comparison shows that the excess power is significantly suppressed after source subtraction, while the subsequent SVD foreground cleaning further reduces both the residual foreground contamination and the remaining excess power.

Foreground contamination in interferometric 21 cm observations is expected to occupy a wedge-shaped region in the cylindrical  $(k_{\perp}, k_{\parallel})$  power spectrum, due to instrumental chromaticity that mixes spectrally smooth foreground emission into higher  $k_{\parallel}$  modes. The wedge boundary can be expressed as (Liu et al. 2014; Datta et al. 2010)

$$k_{\parallel} = \frac{H_0 E(z) D_M(z)}{c(1+z)} \sin \theta_0 k_{\perp}, \quad (12)$$

where  $D_M(z)$  is the transverse comoving distance,  $E(z) \equiv H(z)/H_0$  is the dimensionless Hubble expansion function, and  $\theta_0$  denotes the maximum angular distance from the phase centre over which foreground emission contributes significantly to the contamination.

In this work, the analysis is restricted to the central  $8.5^\circ \times 8.5^\circ$  region of the image cube. Taking half of this field as the characteristic angular extent, we adopt  $\theta_0 \simeq 4.2^\circ$ , which gives a FoV-limited wedge boundary of  $k_{\parallel} \simeq 0.042 k_{\perp}$  at  $z \simeq 0.90$ . For comparison, in the most extreme case where foreground leakage, including contributions from the side lobes, extends to the entire sky, one has  $\sin \theta_0 = 1$  (i.e.  $\theta_0 = 90^\circ$ ), corresponding to the conventional horizon limit  $k_{\parallel} \simeq 0.56 k_{\perp}$ .

We overplot the horizon limit ( $\theta_0 = 90^\circ$ ) in the cylindrical power-spectrum figures to indicate the foreground wedge region. For the spherically averaged (1D) power spectrum, we adopt a conservative

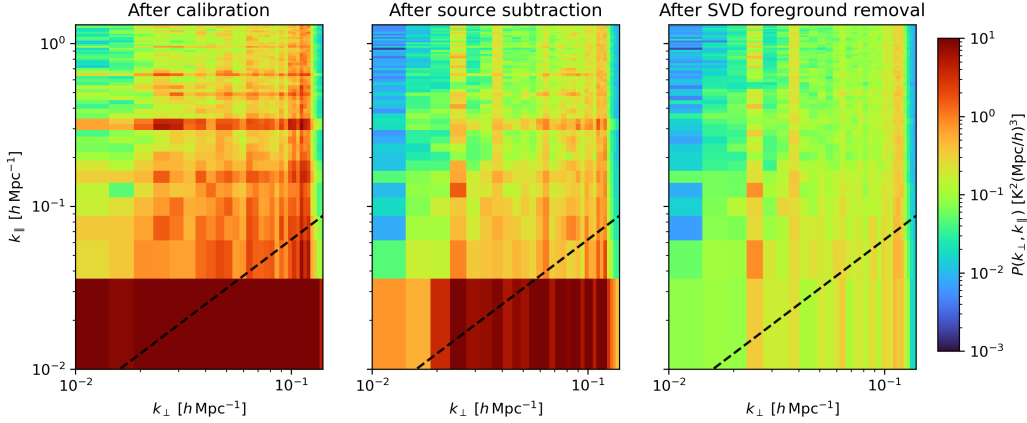


Fig. 9: Cylindrical power spectrum  $P(k_{\perp}, k_{\parallel})$  at different processing stages. *Left*: After calibration; *Middle*: After point source subtraction; *Right*: After SVD-based foreground cleaning with  $N_{\text{fg}} = 30$  modes removed. The diagonal dashed line indicates the horizon limit ( $k_{\parallel} = 0.56k_{\perp}$  at  $z \approx 0.9$ ), below which foreground contamination dominates the wedge-shaped region.

strategy and compute  $P(k)$  using only those  $(k_{\perp}, k_{\parallel})$  pixels that lie above the horizon-limit boundary, thereby minimising residual foreground leakage into the final 1D estimate.

Assuming an isotropic signal, we can average  $P(k)$  in  $k$ -bins to create the spherically averaged dimensionless power spectrum defined as:

$$\Delta^2(k) = \frac{k^3}{2\pi^2} \langle P(k) \rangle_k. \quad (13)$$

We estimate the noise power spectrum through  $\Delta_N^2(k)$  as (Pober et al. 2013; Ansari et al. 2012):

$$\Delta_N^2(k) \approx \frac{k^3}{2\pi^2} X^2 Y \frac{\Omega_{\text{beam}}}{2t} T_{\text{sys}}^2, \quad (14)$$

where  $X^2 Y$  is a scalar factor that translates observed units to cosmological distances.  $X$  converts angles on the sky to transverse distances, and  $Y$  converts bandwidth to line-of-sight distance. Specifically, we have:

$$X = D_A(1+z) \equiv \int_0^z \frac{c dz}{H(z)}, \quad Y = \frac{c(1+z)^2}{\nu_{21} H(z)}. \quad (15)$$

where  $\Omega_{\text{beam}}$  is the solid angle of the primary beam of a single element in steradians,  $T_{\text{sys}}$  is the system temperature, and  $t$  is the total integration time. We estimate the system temperature from our own Cas A transit observation. Specifically, we use a representative cross-correlation baseline ( $1 \times 6$ , Fig. 1) and derive the signal-to-noise ratio from the transit profile after averaging five central frequency channels. Using the flux density of Cas A at 708 MHz,  $S_{\text{CasA}} = 4085$  Jy, we derive an antenna temperature of  $T_A = 29.98$  K. The system temperature is then calculated as

$$T_{\text{sys}} = T_A \frac{\sqrt{2\Delta\nu t_{\text{int}}}}{\text{SNR}}, \quad (16)$$

where  $\Delta\nu = 5 \times 0.244$  MHz = 1.22 MHz,  $t_{\text{int}} = 1$  s and  $\text{SNR} = 475.95$ . This gives a representative value of  $T_{\text{sys}} = 98.39$  K, which is adopted in the following noise power spectrum estimate.

Fig. 10 presents the spherically averaged dimensionless power spectrum alongside the estimated noise power spectrum. We observe that the measured power exceeds the expected thermal noise level by approximately 3–4 orders of magnitude. Although the SVD cleaning effectively removes the visible foreground wedge from the two-dimensional spectrum, the persistently high power in the one-dimensional spectrum indicates significant foreground leakage into the clean window.

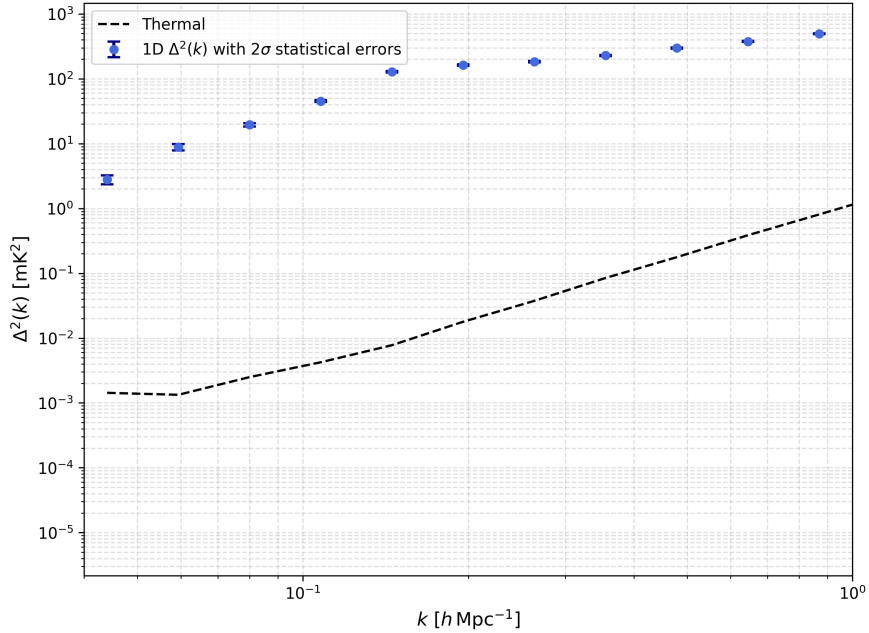


Fig. 10: Spherically averaged dimensionless power spectrum  $\Delta^2(k)$  after SVD foreground cleaning (blue points with error bars), compared with the theoretical thermal noise power spectrum (dashed line). The error bars denote the  $2\sigma$  statistical uncertainty estimated from the standard deviation of the Fourier-mode powers within each  $k$  bin. The measured power exceeds the thermal noise expectation by approximately  $10^3$ – $10^4$  times, indicating significant residual contamination from instrumental systematics, calibration errors, and incomplete foreground removal.

## 5 DISCUSSION AND CONCLUSIONS

In this study, we have presented a comprehensive data analysis pipeline for the Tianlai Dish Pathfinder Array, applying it to nine nights ( $\sim 107$  hours) of observations targeting the North Celestial Pole (NCP) field. Our analysis covers the frequency range of 700–800 MHz, corresponding to the post-reionization redshift of  $z \sim 0.9$ . By employing standard calibration procedures followed by an SVD-based blind foreground subtraction, we successfully obtained deconvolved images and estimated the resulting 21 cm power spectrum.

Our results demonstrate that the SVD technique, with the removal of  $N_{\text{fg}} = 30$  modes, effectively eliminates the visually prominent foreground wedge in the cylindrical 2D power spectrum  $P(k_{\perp}, k_{\parallel})$ . However, as shown in Fig. 10, the spherically averaged 1D power spectrum  $\Delta^2(k)$  reveals a power level approximately  $10^3$ – $10^4$  times higher than the theoretical thermal noise limit.

Understanding the origin of this excess power is crucial for the future detection of both the 21 cm auto-correlation and cross-correlation signals. This residual contamination is likely attributable to several instrumental and modeling non-idealities:

1. **Sky model incompleteness:** Our current calibration relies on a model of only 164 bright sources in the NCP region. This implies that a significant number of fainter point sources remain unsubtracted. Furthermore, bright sources such as Cassiopeia A and Cygnus A can inject power through the antenna sidelobes, contributing to the observed excess.
2. **Calibration artifacts:** Imperfect calibration, for example arising from an incomplete or inaccurate sky model, is known to introduce spurious power into the spectrum (Patil et al. 2016; Ewall-Wice et al. 2017; Barry et al. 2019).

3. **Instrumental and environmental effects:** Instrumental imperfections, such as bandpass ripples and antenna sidelobe chromaticity, couple smooth foregrounds into high- $k_{\parallel}$  modes. Additionally, low-level radio frequency interference that escapes detection by AOFlagger can further elevate the noise floor (Wilensky et al. 2019).

Based on this analysis, we identify several key areas for future improvement: (i) *Refining the sky model:* Including fainter compact sources both inside and outside the primary beam will enhance the precision of both calibration and foreground subtraction; (ii) *Increasing integration time:* While this work utilizes 107 hours of data, the TDPA has accumulated a much larger dataset. Incorporating the full available data will significantly improve the signal-to-noise ratio; (iii) *Pipeline optimization:* We will further optimize AOFlagger settings and CASA calibration parameters (e.g., solution intervals) to achieve superior RFI mitigation and gain stability. (iv) *Polarization diagnostics:* In future work, we will also form and analyze the Stokes Q and U map cubes in parallel with Stokes I. Since polarized foreground emission, after Faraday rotation and instrumental polarization leakage, can introduce frequency-dependent contamination into Stokes I, this will help assess whether residual polarized foregrounds contribute to the excess power that remains in the measured 1D power spectrum (Jelić et al. 2010; Moore et al. 2013).

Although the current excess noise level prevents direct extraction of the faint cosmological HI signal, this analysis has successfully established and validated a complete data processing pipeline for the TDPA. In the future, by implementing the improvements outlined above and combining our observations with cross-correlation analysis against optical galaxy surveys, we expect to overcome the current systematic error bottleneck. This progress will not only open the scientific discovery window for the Tianlai Array in HI intensity mapping, but will also provide valuable practical experience for future larger-scale radio interferometric arrays designed to probe cosmic evolution.

**Acknowledgements** We thank the referee for comments and suggestions that improved the paper. This work is supported by the National SKA Program of China (Nos. 2022SKA0110100 and 2022SKA0110101), the National Natural Science Foundation of China (NSFC) International (Regional) Cooperation and Exchange Project (No. 12361141814), the NSFC (No. 12303004,12203061,12473094,12273070.),the NSFC innovation group grant 12421003, the Specialized Research Fund for State Key Laboratory of Radio Astronomy and Technology, and the National Astronomical Observatories, Chinese Academy of Science (No. E5ZB0901). Y.L. acknowledges the support of the National Natural Science Foundation of China (Nos. 12533001,12473091). This work is also supported by science research grants from the China Manned Space Project with grant Nos. CMS-CSST-2021-B01 and CMS- CSST-2021-A01.

## References

- Anderson, C. J., Luciw, N. J., Li, Y.-C., et al. 2018, *Monthly Notices of the Royal Astronomical Society*, 476, 3382 2, 8
- Ansari, R., et al. 2012, *Astronomy and Astrophysics*, 540, A129 13
- Ansari, R., Marques, G. A., Marriner, J. P., et al. 2025, arXiv e-prints, arXiv:2512.23899 2
- Barry, N., Wilensky, M., Trott, C. M., et al. 2019, *Astrophysical Journal*, 884, 1 14
- Battye, R. A., Brown, M. L., Browne, I. W. A., et al. 2012, arXiv:1209.1041 2
- Battye, R. A., Davies, R. D., & Weller, J. 2004, *Monthly Notices of the Royal Astronomical Society*, 355, 1339 2
- Bharadwaj, S., Nath, B. B., & Sethi, S. K. 2001, *Journal of Astrophysics and Astronomy*, 22, 21 2
- Chang, T.-C., Pen, U.-L., Peterson, J. B., et al. 2008, *Physical Review Letters*, 100, 091303 2
- Chen, X. 2012, *International Journal of Modern Physics Conference Series*, 12, 256 2
- Collaboration, C., Amiri, M., Bandura, K., et al. 2025, arXiv:2511.19620 2
- Cornwell, T. J., & Wilkinson, P. N. 1981, *Monthly Notices of the Royal Astronomical Society*, 196, 1067 5
- Cunnington, S., Barberi-Squarotti, M., Bernal, J. L., et al. 2026, *Ap&SS*, 371, 16 9
- Datta, A., Bowman, J. D., & Carilli, C. L. 2010, *Astrophysical Journal*, 724, 526 12

- Engelbrecht, B. N., Santos, M. G., Fonseca, J., et al. 2025, *Monthly Notices of the Royal Astronomical Society*, 536, 1035 2
- Ewall-Wice, A., Dillon, J. S., Liu, A., et al. 2017, *Monthly Notices of the Royal Astronomical Society*, 470, 1849 14
- Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, *Physics Reports*, 433, 181 7
- Giri, S., Mellema, G., & Jensen, H. 2020, *Journal of Open Source Software*, 5, 2363 12
- Harper, S. E., & Dickinson, C. 2018, *Monthly Notices of the Royal Astronomical Society*, 479, 2024 2
- Irfan, M. O., Bull, P., Santos, M. G., et al. 2022, *Monthly Notices of the Royal Astronomical Society*, 509, 4923 2
- Jelić, V., Zaroubi, S., Labropoulos, P., et al. 2010, *MNRAS*, 409, 1647 15
- Kwak, J., Podczerwinski, J., Timbie, P., et al. 2024, *Journal of Astronomical Instrumentation*, 13, 2450002 2
- Li, J., Zuo, S., Wu, F., et al. 2020, *Science China Physics, Mechanics, and Astronomy*, 63, 129862 2
- Li, Y., Wang, Y., Deng, F., et al. 2023, *ApJ*, 954, 139 2
- Liu, A., Parsons, A. R., & Trott, C. M. 2014, *Physical Review D*, 90, 023018 12
- Liu, A., & Shaw, J. R. 2020, *Publications of the Astronomical Society of the Pacific*, 132, 062001 7
- Masui, K. W., Switzer, E. R., Banavar, N., et al. 2013, *Astrophysical Journal Letters*, 763, L20 2
- Matshawule, S. D., Spinelli, M., Santos, M. G., et al. 2021, *Monthly Notices of the Royal Astronomical Society*, 506, 5075 2
- Mertens, F. G., et al. 2020, *Monthly Notices of the Royal Astronomical Society*, 493, 1662 3
- Mohan, N., & Rafferty, D. 2015, *Astrophysics Source Code Library*, ascl:1502.007 6
- Moore, D. F., Aguirre, J. E., Parsons, A. R., Jacobs, D. C., & Pober, J. C. 2013, *ApJ*, 769, 154 15
- Mozdzen, T. J., Mahesh, N., Monsalve, R. A., et al. 2019, *Monthly Notices of the Royal Astronomical Society*, 483, 4411 2
- Newburgh, L. B., Bandura, K., Bucher, M. A., et al. 2016, in *Proceedings of SPIE*, Vol. 9906, 99065X 2
- Offringa, A. R., van de Gronde, J. J., & Roerdink, J. B. T. M. 2012, *Astronomy and Astrophysics*, 539, A95 4
- Offringa, A. R., et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 444, 606 4
- Patil, A. H., Yatawatta, S., Zaroubi, S., et al. 2016, *Monthly Notices of the Royal Astronomical Society*, 463, 4317 14
- Patil, A. H., et al. 2017, *Astrophysical Journal*, 838, 65 3
- Paul, S., Santos, M. G., Chen, Z., et al. 2023, *arXiv:2301.11943* 2
- Pearson, T. J., & Readhead, A. C. S. 1984, *Annual Review of Astronomy and Astrophysics*, 22, 97 5
- Perdereau, O., Ansari, R., Stebbins, A., et al. 2022, *MNRAS*, 517, 4637 2, 3
- Pober, J. C., et al. 2013, *Astronomical Journal*, 145, 65 13
- Santos, M. G., Cooray, A., & Knox, L. 2005, *Astrophysical Journal*, 625, 575 7
- Shaw, J. R., Sigurdson, K., Sitwell, M., et al. 2015, *Physical Review D*, 91, 083514 2
- Shlens, J. 2014, *arXiv e-prints*, *arXiv:1404.1100* 8
- Spinelli, M., Bernardi, G., Garsden, H., et al. 2021, *Monthly Notices of the Royal Astronomical Society*, 505, 1575 2
- Spinelli, M., Carucci, I. P., Cunnington, S., et al. 2022, *Monthly Notices of the Royal Astronomical Society*, 509, 2048 2
- Team, C. 2022, *Publications of the Astronomical Society of the Pacific*, 134, 114501 4
- Townsend, J., Santos, M. G., Chatterjee, S., et al. 2026, *arXiv e-prints*, *arXiv:2602.23055* 2
- Villaescusa-Navarro, F., Genel, S., Castorina, E., et al. 2018, *Astrophysical Journal*, 866, 135 2
- Wilensky, M. J., Morales, M. F., Hazelton, B. J., et al. 2019, *Publications of the Astronomical Society of the Pacific*, 131, 114507 15
- Wilk, M. B., & Gnanadesikan, R. 1968, *Biometrics*, 55, 1 9
- Wolz, L., Blake, C., Abdalla, F. B., et al. 2017, *Monthly Notices of the Royal Astronomical Society*, 464, 4938 8

- Wu, F., Li, J., Zuo, S., et al. 2021, *Monthly Notices of the Royal Astronomical Society*, 506, 3455 2, 3, 6
- Wyithe, J. S. B., Loeb, A., & Geil, P. M. 2008, *Monthly Notices of the Royal Astronomical Society*, 383, 1195 2
- Yang, W., Wolz, L., Li, Y., et al. 2025, *ApJS*, 279, 32 2
- Yu, Z., Deng, F., Sun, S., et al. 2022, *Research in Astronomy and Astrophysics*, 22, 125007 3
- Zhang, J., Ansari, R., Chen, X., et al. 2016, *MNRAS*, 461, 1950 2
- Zhang, J., Liu, J., Wu, F., et al. 2021, *IEEE Antennas and Propagation Magazine*, 63, 98 2
- Zuo, S., Li, J., Li, Y., et al. 2021, *Astronomy and Computing*, 34, 100439 4