



A Simulation of Calibration and Map-making Errors of the Tianlai Cylinder Pathfinder Array

Kaifeng Yu^{1,2,3}, Fengquan Wu^{1,2}, Shifan Zuo^{1,2}, Jixia Li^{1,2}, Shijie Sun^{1,2}, Yougang Wang^{1,2}, and Xuelei Chen^{1,2,3,4,5}
¹National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China; xuelei@cosmology.bao.ac.cn
²Key Laboratory of Radio Astronomy and Technology, Chinese Academy of Sciences, Beijing 100101, China
³School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China
⁴Key Laboratory of Cosmology and Astrophysics (Liaoning) & College of Sciences, Northeastern University, Shenyang 110819, China
⁵Center of High Energy Physics, Peking University, Beijing 100871, China

Received 2023 June 11; revised 2023 August 3; accepted 2023 August 9; published 2023 September 20

Abstract

The Tianlai cylinder array is a pathfinder for developing and testing 21 cm intensity mapping techniques. In this paper, we use numerical simulation to assess how its measurement is affected by thermal noise and the errors in calibration and map-making process, and the error in the sky map reconstructed from a drift scan survey. Here we consider only the single frequency, unpolarized case. The beam is modeled by fitting to the electromagnetic simulation of the antenna, and the variations of the complex gains of the array elements are modeled by Gaussian processes. Mock visibility data are generated and run through our data processing pipeline. We find that the accuracy of the current calibration is limited primarily by the absolute calibration, where the error comes mainly from the approximation of a single dominating point source. We then studied the m -mode map-making with the help of Moore–Penrose inverse. We find that discarding modes with singular values smaller than a threshold could generate visible artifacts in the map. The impacts of the residue variation of the complex gain and thermal noise are also investigated. The thermal noise in the map varies with latitude, being minimum at the latitude passing through the zenith of the telescope. The angular power spectrum of the reconstructed map show that the current Tianlai cylinder pathfinder, which has a shorter maximum baseline length in the North–South direction, can measure modes up to $l \lesssim 2\pi b_{\text{NS}}/\lambda \sim 200$ very well, but would lose a significant fraction of higher angular modes when noise is present. These results help us to identify the main limiting factors in our current array configuration and data analysis procedure, and suggest that the performance can be improved by reconfiguration of the array feed positions.

Key words: techniques: interferometric – telescopes – cosmology: observations – (cosmology:) diffuse radiation – radio continuum: general

1. Introduction

Neutral hydrogen (HI) is ubiquitous in our Universe, it provides a way to probe the early Universe, and can serve as a tracer of the large-scale matter distribution to reconstruct the expansion history of the Universe (e.g., Furlanetto et al. 2006; Pritchard & Loeb 2012). Making use of the baryon acoustic oscillation (BAO) which can be treated as a cosmological standard ruler, we can measure the dynamics of dark energy. In the past, the BAO has been measured in the optical galaxy redshift survey (e.g., Blake & Glazebrook 2003; Seo & Eisenstein 2003; Eisenstein et al. 2005). Radio observations would complement the optical surveys, and may also be applied to redshifts which so far has not been observed. A promising technique called 21 cm intensity mapping (IM) offers an economical and efficient way to probe this feature at our interested scale (e.g., Chang et al. 2008; Bull et al. 2015; Villaescusa-Navarro et al. 2017). With this technique, instead

of detecting individual galaxies, one can measure the aggregate emission from many galaxies in a patch of sky over a range of frequencies (i.e., redshift) to map the three-dimensional structure. This technique has been demonstrated by cross-correlating the data from the Green Bank Telescope (GBT) and the optical galaxy survey (Chang et al. 2010; Masui et al. 2013). Recently, from the MeerKAT survey, another detection of the cross-correlation power spectrum between the HI IM and optical survey has also been reported (Cunnington et al. 2023). Meanwhile, Paul et al. (2023) presents a direct detection of HI power spectrum on Mpc scale. Other existing or ongoing IM experiments focusing on the late-time cosmology include both single-dish telescopes and interferometers, such as FAST (Hu et al. 2020; Li et al. 2023), BINGO (Battye et al. 2013), CHIME (CHIME Collaboration et al. 2023), Tianlai (Chen 2012; Li et al. 2020), and the future next-generation radio telescope SKA (Santos et al. 2015) also can be used for HI IM observation. The future HI IM surveys will provide an

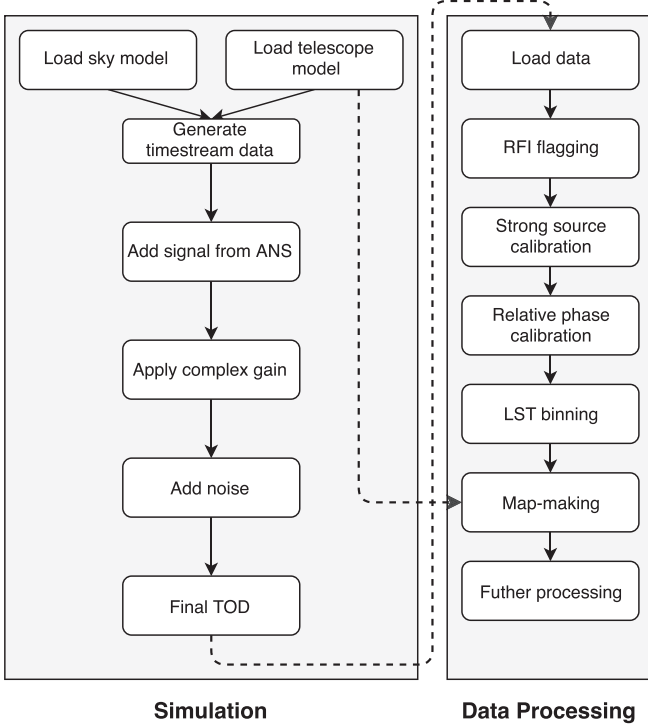


Figure 1. The flowchart of main steps in simulation and data processing of `tlpipe`.

unprecedentedly larger volume of the observable Universe than ever.

However, the 21 cm cosmology also faces great challenges including the contamination from the foreground which is 4–5 orders of magnitude brighter than the cosmological signal, and the complexity in data analysis (Liu & Shaw 2020). Although in principle the foreground radiation has a smooth spectrum while the 21 cm signal has a stochastic one, in practice, the observed spectrum of the foreground is not smooth, which is distorted by instrumental effects, such as the chromatic response of beam and the polarization leakage, which mix the spectral and spatial modes respectively (e.g., Bowman et al. 2009; Cunningham et al. 2021). The main strategies for dealing with foreground include subtraction, avoidance, and suppression (Chapman et al. 2016; Kerrigan et al. 2018). A number of foreground subtraction techniques have been developed, e.g., polynomial fitting (Wang et al. 2006), Principal Component Analysis (PCA) (Liu & Tegmark 2012), and Gaussian Process Regression (GPR) (Mertens et al. 2018).

The instrument response of the telescope is itself determined by calibration measurements. As any physical measurement, the calibration measurement also has its own measurement error, this error will propagate in the data processing and induce error in the final result. As the foreground is much stronger than the 21 cm signal, the error induced by the

instrument calibration could significantly affect the final result. One way to assess the impact of calibration error is to conduct an end-to-end simulation: various instrumental effects are modeled in the simulation to mimic the actual observation, the mock data generated this way is then processed with the same data processing pipeline, in order to check the data analysis algorithms and evaluate the effects of the calibration error on the final results.

The Tianlai project is an experiment aimed at exploring the hardware design and data analysis technique for 21 cm intensity mapping, it includes a cylinder array and a dish array. The basic performance of the Tianlai array has been analyzed in Li et al. (2020) for the cylinder pathfinder, and in Wu et al. (2021) for the dish pathfinder. In this paper, we investigate the impact of calibration error on the map-making of the Tianlai cylinder pathfinder array. This will also be useful in the interpretation of the observational data, and testing the validity and efficiency of algorithms in the data processing pipeline.

The complex gains of the array elements are calibrated by observing bright radio sources in the sky (absolute calibration), and a periodically broadcasting noise source (relative calibration). In this paper we mainly consider the errors in these calibrations, and how these errors would affect the synthetic maps made from the observation, using the *m*-mode analysis method. The main steps of simulation are shown in the left panel of Figure 1. We use a sky model to generate the mock interferometric visibility data, and then process it with the Tianlai data processing pipeline (`tlpipe`) (Zuo et al. 2021), to produce the calibrated data and sky map. We compare these with the input, to determine the error in the calibration, map-making, and power spectrum estimation.

The paper is organized as follows. We first describe the basic models which are used in our simulation, including the telescope beam, the sky model, the receiver noise, and gain variation in Section 2; then we describe the calibration procedure and apply the data processing pipeline to the mock data for calibration in Section 3. We then describe the *m*-mode formalism of map-making and the results in Section 4. Finally, we conclude in Section 5.

2. Simulation

For radio interferometers, the correlation of voltages between a pair of antenna receivers, or *visibility*, is related to the sky temperature as

$$V_{ij}(t) = g_i(t)g_j^*(t) \int T(\hat{n})A_i(\hat{n}; t)A_j^*(\hat{n}; t) \times e^{2\pi i \hat{n} \cdot \mathbf{u}_{ij}(t)} d^2\hat{n} + n_{ij}(t) \quad (1)$$

$$= g_i(t)g_j^*(t) \int d^2\hat{n} B_{ij}(\hat{n}; t)T(\hat{n}) + n_{ij}(t), \quad (2)$$

Table 1
Simulation Parameters

Parameters	Value
Latitude	44°15' N
Cylinder numbers	3
Cylinder width	15 m
Feeds	31, 32, 33 [from East to West]
Longest NS Length	12.4 m
System Temperature	90 K
Bandwidth	700–800 MHz
Channel width	122 kHz

where \hat{n} denotes the direction on the sky, $T(\hat{n})$ is the brightness temperature in that direction. $\mathbf{u}_{ij} = (\mathbf{r}_i - \mathbf{r}_j)/\lambda$ is the baseline vector between the two feeds i and j in unit of the wavelength, $A_i(\hat{n})$ is the primary beam of feed i , $g_i(t)$ is the complex gain of feed i to be calibrated, and $n_{ij}(t)$ is the receiver noise, which we assume to have zero mean. The beam transfer function $B_{ij}(\hat{n}; t)$ is

$$B_{ij}(\hat{n}; t) = A_i(\hat{n}; t) A_j^*(\hat{n}; t) e^{2\pi i \hat{n} \cdot \mathbf{u}_{ij}(t)}. \quad (3)$$

For simplicity, we will assume that all $A_i(\hat{n}; t)$ are the same in this paper. For the Tianlai cylinder which is fixed on ground, the rotation of Earth generates varying visibility.

2.1. Telescope Configuration and Beam Model

The Tianlai cylinder array pathfinder has three cylindrical reflectors with 15 m width, and there are 31, 32, 33 feeds spaced along the focal axis respectively, which is in the North–South direction, the longest distance between feeds on each cylinder is 12.4 m, though the cylindrical reflector itself is 40 m long. The purpose of the unequal spacing is to reduce the grating lobes, an effect appears when the minimum feed spacing is greater than half observational wavelength (Zhang et al. 2016b). The adopted simulation parameters are listed in Table 1.

The primary beam pattern of the Tianlai cylinder pathfinder has not been measured, especially along the North–South direction, though along the East–West direction it can be derived by observing the transit of a strong radio source (Zuo et al. 2019; Li et al. 2020). Here we model the primary beam as a product of a Gaussian function along the North–South direction and a Fraunhofer diffraction solution along the East–West,

$$A(\hat{n}) = A_{\text{NS}}(\sin^{-1}(\hat{n} \cdot \hat{x}); \theta_{\text{NS}}) \\ \times A_{\text{EW}}(\sin^{-1}(\hat{n} \cdot \hat{y}), F, \theta_{\text{EW}})$$

where \hat{x} and \hat{y} are the unit vector pointing East and North, respectively, and F is the focal ratio. In the North–South

direction, the beam amplitude has the form

$$A_{\text{NS}}(\theta; \theta_{\text{NS}}) = \exp \left[-4 \ln 2 \left(\frac{\theta}{\theta_{\text{NS}}} \right)^2 \right] \quad (4)$$

where $\theta_{\text{NS}}(\nu) = \alpha \frac{\lambda}{D_{\text{NS}}}$ and $D_{\text{NS}} = 0.3$ m as the size of the Tianlai cylinder feeds. For the beam in the East–West direction, we take the form

$$A_{\text{EW}}(\theta; F, \theta_{\text{EW}}) \propto \int_{-\frac{W}{2}}^{\frac{W}{2}} A_D \\ \left(2 \tan^{-1} \left(\frac{x}{2FW} \right); \theta_{\text{EW}} \right) e^{-ikx \sin \theta} dx \\ \propto \int_{-\frac{1}{4F}}^{\frac{1}{4F}} A_D(2 \tan^{-1} u; \theta_{\text{EW}}) e^{-i \frac{4\pi F W u}{\lambda} \sin \theta} du, \quad (5)$$

where W is the width of the cylinder, and A_D is taken as

$$A_D(\theta; \theta_{\text{FWHM}}) = \exp \left[-\frac{\ln 2}{2} \frac{\tan^2 \theta}{\tan^2(\theta_{\text{FWHM}}/2)} \right],$$

which is consistent with Shaw et al. (2015). We choose the parameter α in θ_{NS} and F, θ_{EW} in A_{EW} by fitting the result of the electromagnetic (EM) field simulation (Sun et al. 2022).

In Figure 2, we show the fitted power beam $|A|^2$ with the EM simulated normalized gain at 750 MHz in the East–West direction and the North–South direction separately. We fit the data with the mean of the X and Y polarization for our unpolarized case, which gives the fitted parameters $\alpha = 1.04$, $F = 0.2$ and $\theta_{\text{EW}} = 2.74$. This may still differ from the actual case, but for the purpose of simulation it should be sufficient.

2.2. Sky Model

We generate the sky maps including foreground components and the cosmological 21 cm signal using the publicly available code `cora`⁶ (Shaw et al. 2014, 2015). The primary foreground components consist of diffuse synchrotron emission from the Galaxy and emission from the extragalactic point sources. The base map is generated by extrapolating the 408 MHz map (Haslam et al. 1982) with spectral index from Miville-Deschênes et al. (2008), and Gaussian random fluctuations in brightness and spectral index are added to account for small scale fluctuations. The emission of extragalactic point sources involves three components: bright point sources in the VLSS and NVSS catalogs; synthetic dimmer sources, constructed by drawing from the point source distribution in Di Matteo et al. (2002); unresolved background of dimmer sources, generated by drawing a Gaussian realization from an angular power spectrum for point sources model. The redshifted 21 cm signal is treated as a Gaussian fluctuation, generated according to a given angular power spectrum.

⁶ <https://github.com/radiocosmology/cora>

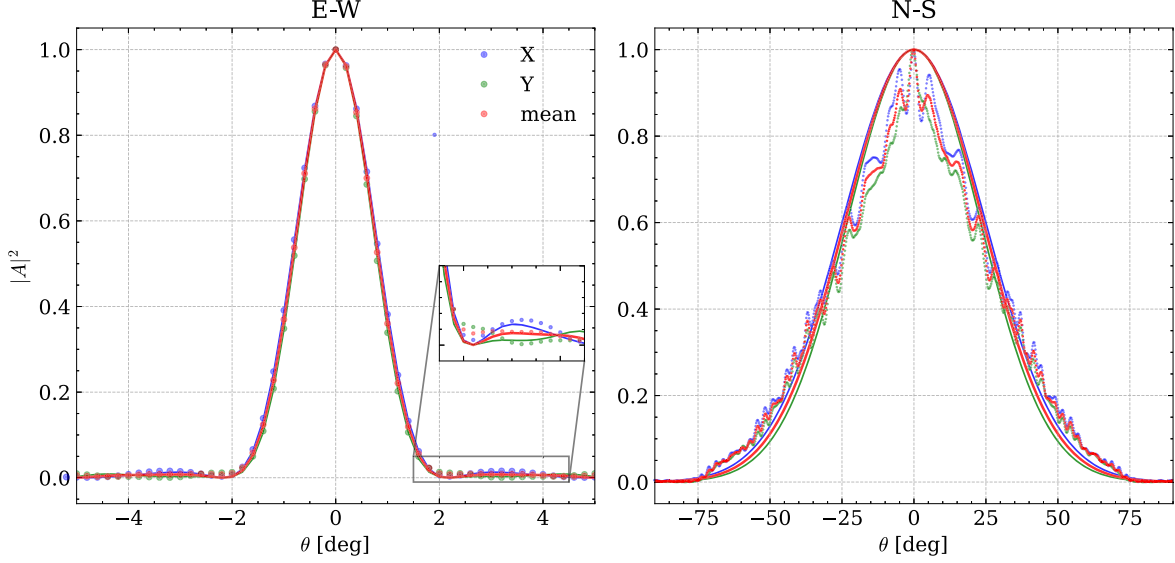


Figure 2. The EM simulated normalized power gain (dot) of the X (blue) and Y (green) polarization feed at 750 MHz and the corresponding fitted beam in the East–West direction (left) and North–South direction (right). The red line fits the average of the two polarizations, which is used for our unpolarized case.

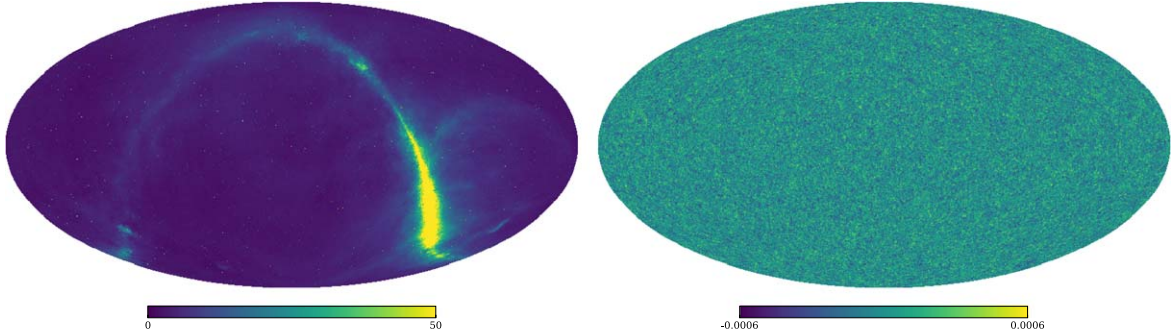


Figure 3. The simulated sky maps at 750 MHz of foreground including diffuse synchrotron emission and extragalactic point sources (left) and 21 cm signal (right).

We illustrate the simulated sky map of the foreground and 21 cm signal at 750 MHz in Figure 3, though in the present study, we will deal mainly with the much stronger foreground radiation. The maps are pixelated using the HEALPix scheme (Górski et al. 2005), with $N_{\text{side}} = 512$ corresponding to an angular resolution of $6.87'$.

2.3. Receiver Noise

We assume the noise in the visibility of each feed pair at a given frequency follows a Gaussian distribution, of which the variance at any particular time can be modeled as

$$\sigma_{\text{noise}} = \frac{\Omega T_{\text{sys}}}{\sqrt{2} \sqrt{t_{\text{int}} \Delta \nu}} \quad (6)$$

where $\Omega = \int |A(\hat{n})|^2 d^2 \hat{n}$ is the beam solid angle, T_{sys} is the system temperature, t_{int} is the integration time, and $\Delta \nu$ is the

bandwidth of frequency channel, the $\sqrt{2}$ in denominator is for an average of the two polarizations—in this work we neglect the difference in the two polarizations. As analyzed in Li et al. (2020), the average system temperature of the Tianlai cylinder pathfinder is 90 K. Currently for the Tianlai cylinder pathfinder $t_{\text{int}} = 4$ s, $\Delta \nu = 122$ kHz. For our simulated beam, $\Omega \approx 0.034$ sr. So we get $\sigma_{\text{noise}}/\Omega = 0.091$ K per beam, or $\sigma_{\text{noise}} \approx 0.0031$ K · sr for a one day observation. In the sampling of the complex noise data, both the standard deviation in the real and imaginary part are taken as $\sigma_{\text{noise}}/\sqrt{2}$.

2.4. Complex Gain Fluctuation and Cable Delay

We model the complex gain of each input channel as

$$g(\nu, t) = |g(\nu, t)| e^{i(2\pi \nu \tau + \varphi(\nu, t))} \quad (7)$$

where τ is the instrumental time delay, and φ is a residue term in the phase which is nonlinear with respect to the frequency ν .

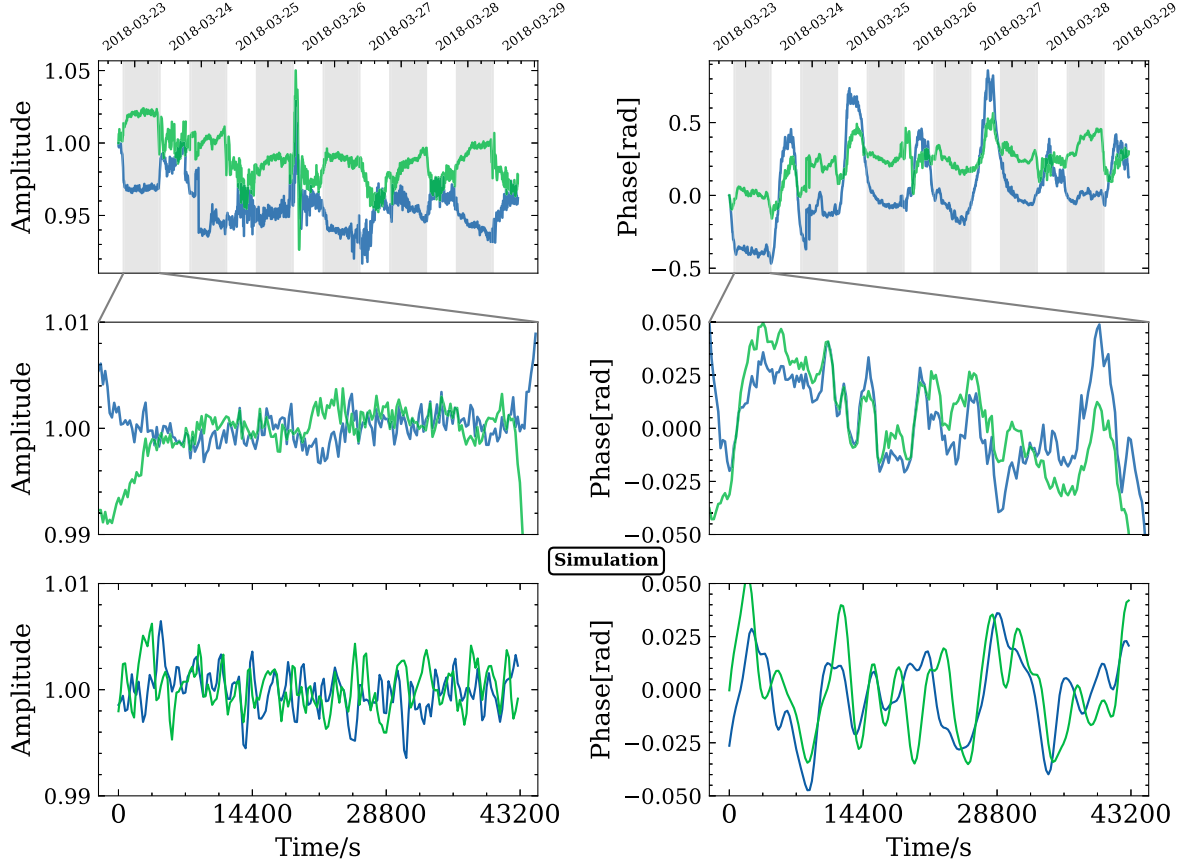


Figure 4. The variation of the relative amplitude (top left) and phase (top right) of the complex gain for two feed channels (plotted in blue and green curves) of the Tianlai cylinder pathfinder in six consecutive days starting from 2018/03/22. The shadow area corresponds to the observation in the nighttime(12:00-0:00 UTC). The middle panels shows a zoom up of the data of the first night. In the bottom panels we show an example of our simulated complex gain variation.

Due to temperature variation, we found τ varies at the level of a few to a few tens of picoseconds (Zuo et al. 2019), corresponding to millimeter to centimeter changes in the optical cable, and accounts most of the instrument phase variation. The top and middle panels of Figure 4 show the normalized amplitude and phase of the gain for the X-polarization of two feeds on the Tianlai array, which are fairly typical. As we can see from the figure, although there are large variations in the phase during the day time, the phase variation during the night is relatively small and the relative amplitudes are quite stable.

In our simulation, we model the complex gain of one frequency channel as (we omit the variable ν here)

$$g(t) = (1 + \Delta g(t))e^{i2\pi\nu\phi(t)}$$

where we have combined the time delay τ and residue phase φ into a single phase ϕ , as we consider here a single frequency channel. In the actual instrument, a large part of the change in time delay comes from the change of cable length due to temperature variation, so it is not completely random but dependent on the diurnal variation of temperature. However, these changes will be corrected by the calibration procedure,

what really matters is the residue for the calibration. So it is sufficient here to model this change as a Gaussian process. We generate the fluctuations in amplitude Δg and phase Φ separately from multivariate normal distributions,

$$\Delta g \sim \mathcal{N}(\mathbf{0}, \Sigma^{\text{amp}}), \quad \Phi \sim \mathcal{N}(\mathbf{0}, \Sigma^{\text{phs}}).$$

The fluctuations should be correlated in time, we model this by adopting the covariance function Σ^{amp} and Σ^{phs} as

$$\Sigma_{ij} = \sigma^2 \exp\left(-\frac{(t_i - t_j)^2}{2\xi^2}\right), \quad (8)$$

where σ and ξ gives the fluctuation amplitude and correlation length respectively.

Inspired by the variation in Figure 4, we take $\sigma^{\text{amp}} = 0.002$ for the amplitude part, and $\sigma^{\text{phs}} = 3$ picosecond for the phase part. It is more difficult to get the correlation length, as we do not have sufficiently accurate measurement. Here we take the correlated time to be $\xi^{\text{amp}} = 300$ s, and $\xi^{\text{phs}} = 900$ s. This model is not derived by a fitting the observation data, but they do give variations comparable in magnitude to the actual case,

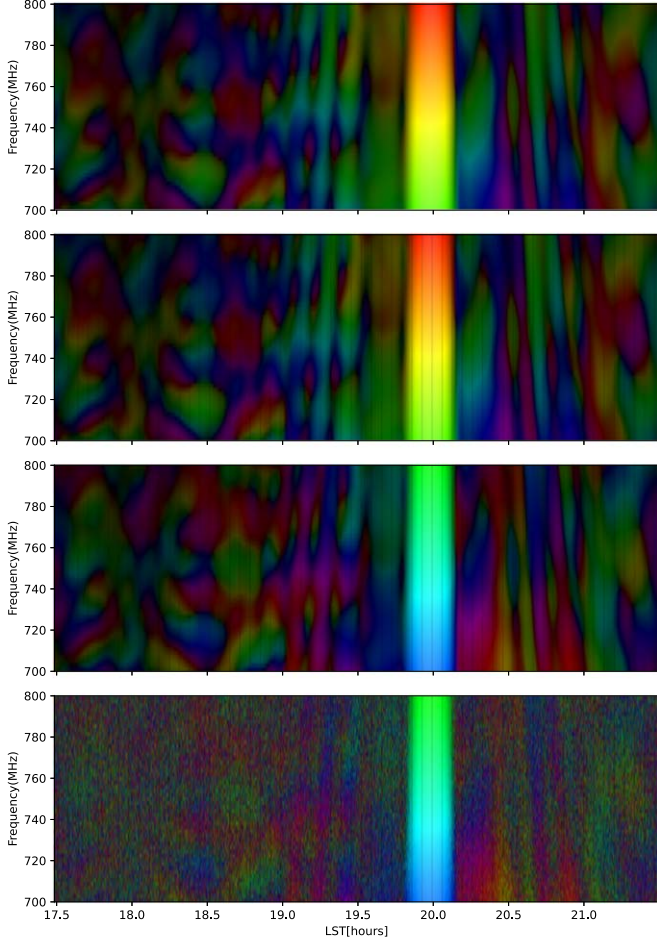


Figure 5. The timestream data of baseline A1–A31 in 4 hr from 700 MHz to 800 MHz at each main stage of applying instrumental effects. From top to bottom, (1): The underlying visibility; (2): Add signal from noise source (the periodic vertical line); (3): Apply the cable delay; (4): Add Gaussian noise with $N_{\text{day}} = 1$.

so that we could have some idea of the effect of variation of the complex gains.

The simulated visibility of a duration of 4 hr for a pair of distant feeds on the same cylinder (A1–A31) is shown as a waterfall in Figure 5. The visualization scheme is from SageMath,⁷ where the brightness and color represents the magnitude and phase of the complex visibility. In the top panel, we show the ideal visibility generated from the sky map. The pattern is produced by the sky model, and the bright feature is the transit of the bright source (Cygnus A) across the primary beam. The color changes with frequency, as we would expect for a baseline of fixed physical length which induce different phase at different wavelength. In the second panel from the top, we added the periodically broadcasting noise source (described

in the next section), which appear as regularly spaced fine vertical lines. The third panel from the top shows the addition of gain variation. In the bottom panel, we added noise, as would be obtained from a single day observation. Note that in the real data, there would be cross-talks between the different receiver channels, which would induce “correlated noise” which persist over time and show up as horizontal features in the waterfall plot (see Li et al. 2020), here we ignore such noise.

3. Calibration

After generating the mock visibility data using the sky model and telescope model described above, we pass the simulated data to the Tianlai data processing pipeline `tlpipe`⁸ (Zuo et al. 2021). Its main tasks include data distribution, RFI flagging, calibration, local sidereal time (LST) binning, map-making, and some other utilities for data selection and analysis. The sketchy data processing procedure from handling raw data to scientific products is shown in the right panel of Figure 1. Below we summarize the calibration tasks.

In an interferometer array, the amplitude and instrument phase of the receiver gain g_i may vary in time, and the data need to be calibrated. This can be done by observing known sources. Two kinds of calibration sources are used for the Tianlai array: (1) strong point source on the sky when they are transiting over the fixed Tianlai primary beam, which can provide calibration for both the amplitude and phase of the receiver gain, we will call this absolute calibration; (2) an artificial calibrator noise source (CNS) to calibrate the relative phase change in the complex gain, which we will call relative calibration. In `tlpipe`, the complex gain is calibrated for each frequency channel individually. We will briefly summarize the algorithm here, refer to Zuo et al. (2019) for more details.

3.1. Absolute Calibration

An absolute calibration can be performed with the help of a bright point source on the sky. When it transits through the main beam of the telescope, we assume the signal from it when transiting dominates the visibility data, which is

$$V_{ij}^0 = S_c g_i^* g_j^* A_i^*(\hat{n}_0) A_j^*(\hat{n}_0) e^{i2\pi\hat{n}_0 \cdot (u_i - u_j)} \quad (9)$$

where S_c is the flux of the strong point source, \hat{n}_0 is its position. Due to RFI or receiver malfunction, some of the data may be corrupted and appear as outliers. In the presence of outliers and noise, the whole data can be written in matrix form as $V = V^0 + S + N$, where S is a sparse matrix which might come from residual RFI or missing values, N represents the contribution from noise. We can perform a Stable Principal

⁷ https://doc.sagemath.org/html/en/reference/plotting/sage/plot/complex_plot.html

⁸ <https://github.com/TianlaiProject/tlpipe>

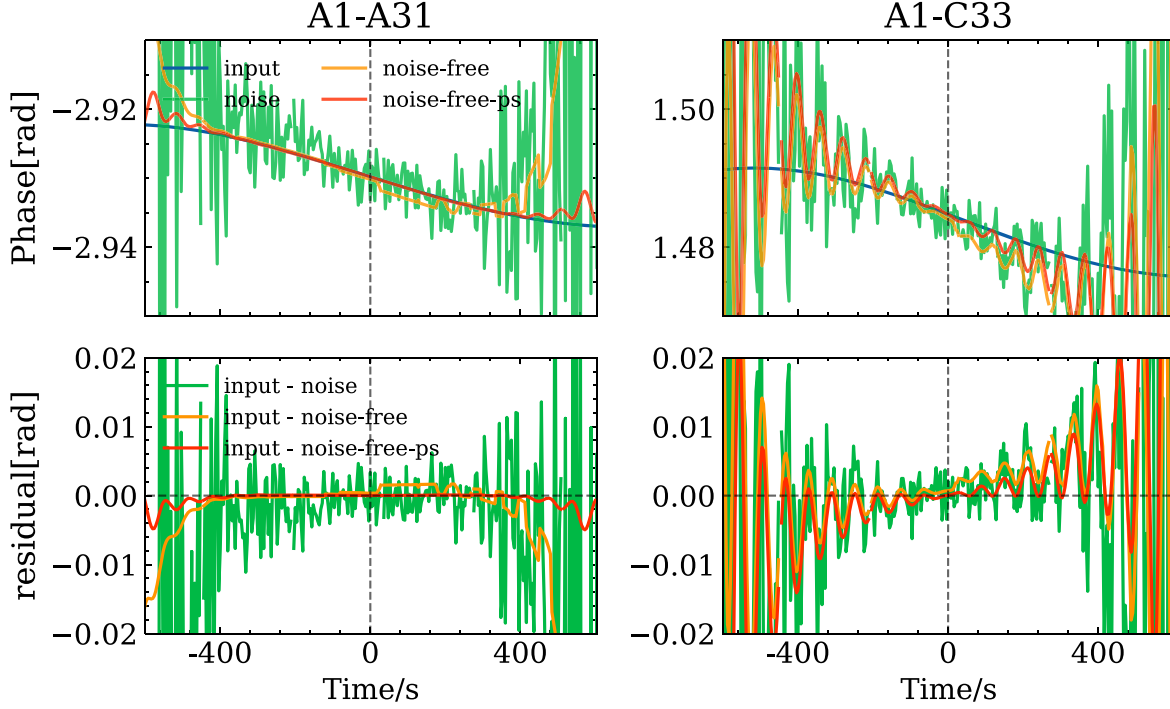


Figure 6. The solution of absolute calibration when transiting Cyg A for the full sky data with/without noise, and the sky with four brightest point sources only of a short baseline (left column) and a longer baseline (right column). The top row shows the input complex gain (blue), the solved gain for the full sky data with noise (green) and noise-free (orange), and for the four point sources sky noise-free data (red). The bottom row shows the difference between input and these solved gain.

Component Analysis (SPCA) on V to extract V^0 , then

$$V_{ij}^0 = S_c G_i G_j^*,$$

where $G_i = g_i A_i(\hat{n}_0) e^{-2\pi i \hat{n}_0 \cdot u_i}$ or in matrix-vector form

$$V^0 = S_c G G^*. \quad (10)$$

Note that the vector G is an eigenvector of V^0 corresponding to its largest eigenvalue.

After solving G with the eigenvector decomposition method during the transit of bright point sources, we can derive the corresponding complex gains during this period. In our experiment, we have mainly used the Cygnus A (Cyg A) as the calibrator, though Cassiopeia A (Cas A), Virgo A (Vir A) and Taurus A (Tau A) are also used occasionally. It is the brightest source in the sky, and is unresolved by the Tianlai cylinder pathfinder. We plot the input and solved phase for two baselines (A1–A31 and A1–C33) during the Cyg A transit in the top panel of Figure 6, and the residue in the bottom panels. In each plot, we show an idealized case where the sky has only a dominating point source and without noise, a case where the sky is given by our fiducial sky model but noise-free, and the more realistic case with noise.

As we can see from the figure, during the transit of Cyg A, the solutions are basically good within ± 300 s of the transit, when the source is within the primary beam and dominates the received signal, though for long baselines we can see some

small oscillations at the 10^{-3} rad level. Beyond this timescale, the source is outside the primary beam and not dominating, and as a result the solved phase has large errors. Within this window, the model with only point source yields very accurate phase, but for realistic sky model there is still some residue, and with the thermal noise this error is larger.

The phase and residue for all 96 feeds at the moment of the Cyg A is plotted in Figure 7. In the top row of Figure 7, we show the phase of the complex gains for the 96 receivers during the transit, which are randomly distributed in the $(-\pi, \pi)$ range. We show the input, the calibration result without adding noise, and the calibration result with the addition of noise. The differences are small, so in the top panel of Figure 7 the points of these cases coincide. The difference between the input and the two calibration solutions are shown in the middle panel of Figure 7.

As we can see, even in the noise-free case, the calibration solution is still a little different from the input, at the level of $10^{-3} \sim 10^{-4}$ rad. To identify the source of this error, we also simulated the case where the sky consists only four well separated points sources (Cyg A, Cas A, Vir A, Tau A), and the visibility is only dominated by one of them at a time. In this simplified case, the solved phase is almost equal to the input. However, once we take the realistic sky radiation model into account, we see there is error in the solved phase, because this

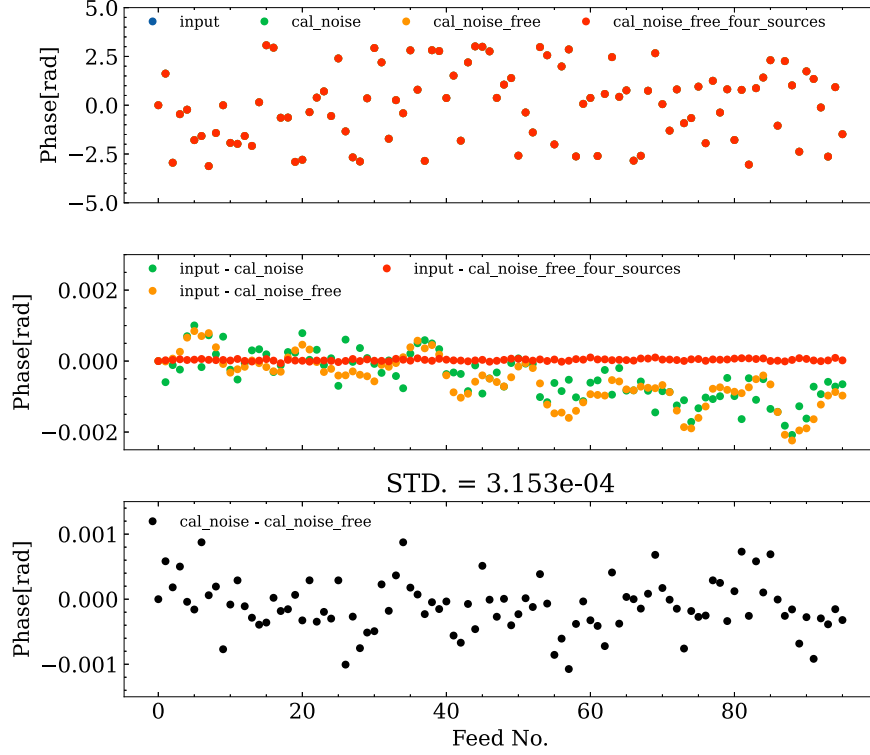


Figure 7. The phase of the complex gains of the feeds as determined by an absolute calibration at the Cyg A transit. Top: The phase of the complex gain for the 96 feeds at the time of Cyg A transit. The points show the input phase (blue), the solved phase without noise (orange), the solved phase with noise (green), and the solved noise with simplified sky model of only four strong point sources. Middle: The residual between the input phase and the solved phase. Bottom: The difference between the solved phase for the noise data and the noise-free data.

contribution is neglected in Equations (9) and (10). We also tested adding the expected thermal noise, this will cause some additional error, but is smaller than the error from over-simplified sky model. In the bottom panel of Figure 7, the difference between the case without thermal noise and with thermal noise is shown, which is at the 10^{-4} level. For comparison, the closure phase test indicates an error level of 10^{-3} rad for the actual Tianlai cylinder array data (Li et al. 2020), comparable or slightly larger than the simulation result.

Note that for an interferometer array, the size of errors in the calibrated instrument phases depends both on the calibration precision for individual baseline and on the array scale N . During the calibration process, each baseline provides an equation to constrain the phase, and the number of baselines scales as $O(N^2)$, while the number of unknown variables scales as $O(N)$. For comparable calibration precision of an individual baseline, the resulting error of the instrument phase achieved with a large array can be much smaller than that with a small array. However, for a very large array this scaling may be broken due to effects not considered in the interferometer model, such as cross-coupling between array elements, or direction-dependent instrument phase, etc., which would limit the precision of the calibration in large arrays. In our present

case, the Tianlai cylinder array has an $N \sim 10^2$ which is larger than many previous arrays, so the error is correspondingly smaller. Still, the estimated error size is comparable with what we find in the actual data as indicated by the closure phase measurement, so we believe our present modeling is adequate and captures the main source of error.

3.2. Relative Calibration

As the point source calibration could only be made when a known bright point source is transiting, and there are only a few such sources in the sky, we use the relative phase calibration for the rest of the time.

An artificial CNS is placed nearby to broadcast broadband noise-like signal periodically. In this simulation, we model its visibility during the broadcast as

$$V_{ij}^{\text{CNS}} = S_{\text{CNS}} \frac{A_i(\hat{n}_i) A_j(\hat{n}_j)}{\sqrt{\Omega_i \Omega_j}} \frac{r^2}{r_i r_j} e^{ik(r_i - r_j)} \\ = C e^{ik(r_i - r_j)}, \quad (11)$$

where $k = 2\pi/\lambda$, r_i and r_j are the distance from CNS to feed i and feed j . For the real case, this may not be completely accurate, due to near field effect and the reflections from nearby

terrain. However, in the relative calibration we do not make any usage of this knowledge, so it should not affect the result.

In the `tlpipe`, the visibility when CNS is on and its adjacent visibility when CNS is off are assumed to be

$$V_{ij}^{\text{on}} = g_i g_j^* (V_{ij}^{\text{sky}} + V_{ij}^{\text{CNS}}) + n_{ij}^{\text{on}} \quad (12)$$

$$V_{ij}^{\text{off}} = g_i g_j^* V_{ij}^{\text{sky}} + n_{ij}^{\text{off}} \quad (13)$$

we can assume $n_{ij}^{\text{on}} - n_{ij}^{\text{off}} \approx 0$ when compared with the CNS term, then

$$V_{ij}^{\text{on}} - V_{ij}^{\text{off}} = g_i g_j^* V_{ij}^{\text{CNS}} \quad (14)$$

As the CNS and feeds are fixed, the phase change in its visibility would come from the variation of the complex gain. Regardless the exact CNS visibility, as long as it does not change, the phase of the observed visibility is

$$\Phi_{ij} = \text{Arg}(V_{ij}^{\text{on}} - V_{ij}^{\text{off}}) \approx \text{Arg}(g_i g_j^*) + \text{Arg}(V_{ij}^{\text{CNS}}) \quad (15)$$

and its change depends only on the variation of the complex gain and does not depend on the exact form of the CNS induced visibility,

$$\Delta\Phi_{ij}(t) \equiv \Phi_{ij}(t) - \Phi_{ij}(t_0) \quad (16)$$

where $\Phi_{ij}(t)$ is the phase at time t , and t_0 is the time of the last absolute calibration. We can then calibrate the visibility as

$$V_{ij}^{\text{rel-cal}} = e^{-i\Delta\Phi_{ij}} V_{ij}^{\text{off}} \quad (17)$$

We can estimate the minimum magnitude of the error of relative calibration as induced by the measurement noise: for the complex visibility of our unpolarized case, the measurement error in each of the real and imaginary components is about $\sigma_{\text{noise}}/\sqrt{2} \approx 0.0022 \text{ K} \cdot \text{sr}$. In the absolute or relative calibration, if the visibility induced by the calibration source is V , and $|V| \gg \sigma$, then the measured visibility has a phase error of

$$\delta\phi \approx \frac{\sigma}{|V|} \quad (18)$$

The CNS signal strength is about $10^2 \text{ K} \cdot \text{sr}$ in our simulated visibility so the error $\delta\phi_{\text{rel}}$ induced by the thermal noise of measurement is

$$\delta\phi_{\text{rel}} \sim \frac{0.0022}{100} \approx 2.2 \times 10^{-5} \text{ rad} \quad (19)$$

Figure 8 shows the phase of the complex gain of a typical baseline after the two-step (absolute+relative) calibration. In the top panel we show the phase evolution over one day, and in the second row show the phase during the transit of Cyg A. Both the input and calibration results (with noise or without noise) are shown. In the third row, we show the difference between the input and the calibration solution. Here we plot two cases: the case with noise in both the absolute and relative calibration; and the noise-free case, where we ignore the noise in both the absolute and relative calibration. As noted before,

the solution has a phase error due to numerical error at a level of 10^{-4} rad, which is set largely by the absolute calibration. Due to this error from the absolute calibration, the two solutions are different from the beginning, but this difference is almost constant, as the phase variation is precisely corrected by the relative calibration procedure. Finally, in the bottom row, we show the difference between the solution with and without noise, this shows an error of about 2.3×10^{-5} , which is consistent with our estimates of relative calibration error from above. In addition to the error induced by thermal noise, there could be other errors, e.g., from reflection and multi-path effects, which we will ignore in the present work.

Our simulation result shows that the absolute calibration could fix the randomly distributed instrument phase, and for the daily instrumental phase variations at the 10^{-2} rad level, the relative calibration could reduce it significantly. However, it also shows that during the absolute calibration, there is an error at the $10^{-4} \sim 10^{-3}$ rad level, which will persist subsequently. This error is partly due to the fact that the sky is actually not dominated by a single radio source, and partly due to thermal noise. The absolute calibration accuracy might be improved by jointly fitting the phase with the data obtained during the whole transit process instead of a single epoch data at the transit point, and also perhaps by adopting more sophisticated sky model.

The relative calibration would track how the phase vary from the initial phase, but would not improve on the initial phase determination. The error in the relative calibration is however smaller, at the 10^{-5} level.

4. Map-making Results

The output of an interferometer array is the visibility data V_{ij} , from this data the sky brightness temperature map $T(\hat{n})$ can be obtained by solving the inverse problem of the visibility Equation (2) (Zheng et al. 2017). While map-making is not absolutely necessary for obtaining the 21 cm power spectrum, which can in principle be obtained directly from the visibility data, a map is more intuitive than the visibilities, so it provides a good check for potential problems and systematic effects in the observation.

For drift scan surveys, where the telescope is fixed on the ground and sky drift by due to the rotation of Earth, an m -mode decomposition method can be used, which simplifies the problem and reduces the amount of computation significantly (Shaw et al. 2014, 2015; Zhang et al. 2016a, 2016b). In the present work we will simulate the map-making processing using the m -mode formalism.

4.1. The m -mode Formalism

The m -mode formalism works with the drift scan telescope, and has been used in the sky reconstruction for the Tianlai dish and cylinder arrays using simulated data (Zhang et al.

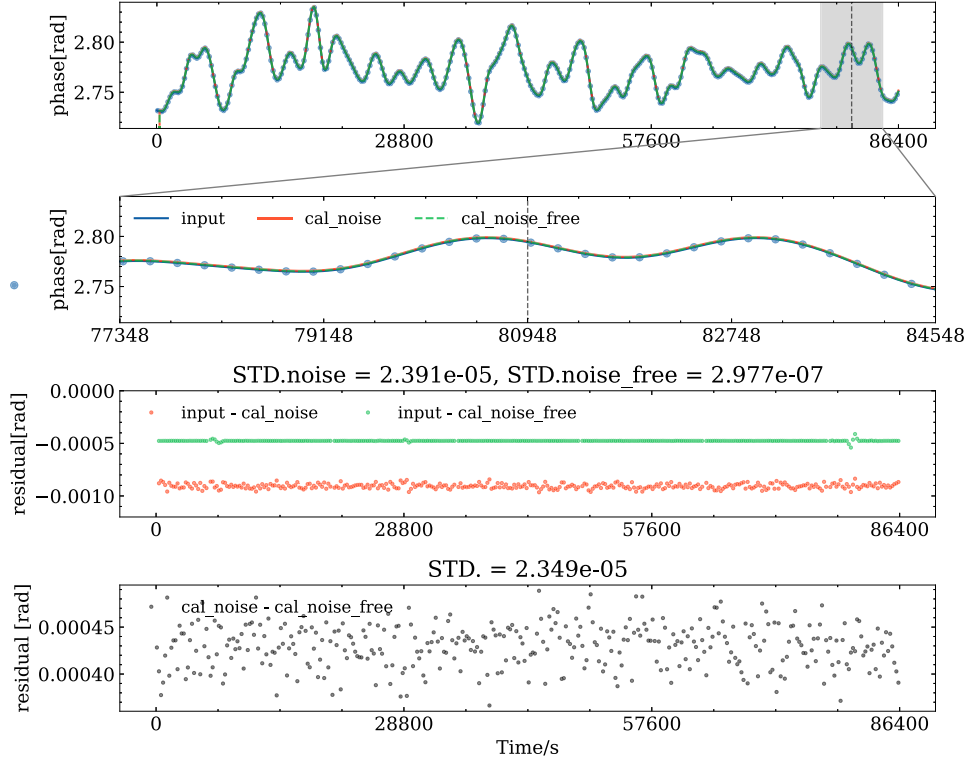


Figure 8. The solution to the phase variation and the residual after calibration for one baseline (A1–A21 in this case). The dots in all panels indicate the on-time of CNS. Top Row: the solved phase for noise data (orange) and noise-free data (green dashed) for a whole day data compared with the input (blue). Second Row: Zoom in on the period when Cyg A is transiting. Third Row: The residual between the input and the calibration solution. Bottom Row: The difference between the solved phase of the two cases, the variation of which is from the noise.

2016a, 2016b), and the analysis of OVRO-LWA data (Eastwood et al. 2018, 2019). We summarize it briefly below.

Take the fact that the measured visibility changes periodically with the sidereal day as the Earth rotates, we can replace the time dependence of visibility with the dependence on the sidereal hour angle φ as $\varphi(t)$, then $V_{ij}(t)$ and $B_{ij}(\hat{n}; t)$ in Equations (2) and (3) are expressed as $V_{ij}(\varphi)$ and $B_{ij}(\hat{n}; \varphi)$. Taking advantage of the periodicity of φ , we can Fourier transform $V_{ij}(\varphi)$ with respect to the φ as

$$V_m^{ij} = \int_0^{2\pi} \frac{d\varphi}{2\pi} V_{ij}(\varphi) e^{-im\varphi} \quad (20)$$

where V_m^{ij} is the m -mode indexed m , the Fourier conjugate variables to the sidereal hour angle φ . The m -modes V_m^{ij} correspond to the components of the visibility $V_{ij}(\varphi)$ varying on different timescales, larger values of m correspond to the faster varying components. V_0^{ij} is the average of the visibility over sidereal time.

Take the spherical harmonics transform of the beam transfer matrix $B_{ij}(\hat{n}; \varphi)$ and the sky brightness temperature $T(\hat{n})$,

$$B_{ij}(\hat{n}; \varphi) = \sum_{lm} B_{lm}^{ij}(\varphi) Y_{lm}^*(\hat{n}) \quad (21)$$

$$T(\hat{n}) = \sum_{lm} a_{lm} Y_{lm}(\hat{n}), \quad (22)$$

substitute them into Equation (2) and combine with Equation (20). Ignore the complex gain for now, we have

$$V_m^{ij} = \sum_{lm'} \int \frac{d\varphi}{2\pi} B_{lm'}^{ij}(\varphi) a_{lm'} e^{-im\varphi} + n_m^{ij}. \quad (23)$$

Using the property of the drift scan telescope, of which the pointing changes with the rotation of Earth, the beam transfer function can be written as $B_{lm'}^{ij}(\varphi) = B_{lm'}^{ij}(0) e^{im'\varphi}$. Substitute it into Equation (23), it gives

$$V_m^{ij} = \sum_l B_{lm}^{ij} a_{lm} + n_m^{ij}. \quad (24)$$

For any $m > 0$, the positive and negative m -modes V_m^{ij} and V_{-m}^{ij} are independent data. For the real-valued sky temperature, $a_{l,-m} = (-1)^m a_{lm}^*$, then

$$V_{-m}^{ij*} = \sum_l (-1)^m B_{l,-m}^{ij*} a_{lm} + n_{-m}^{ij*}. \quad (25)$$

Equation (23) can be rewritten in matrix form

$$\mathbf{v} = \mathbf{B}\mathbf{a} + \mathbf{n}, \quad (26)$$

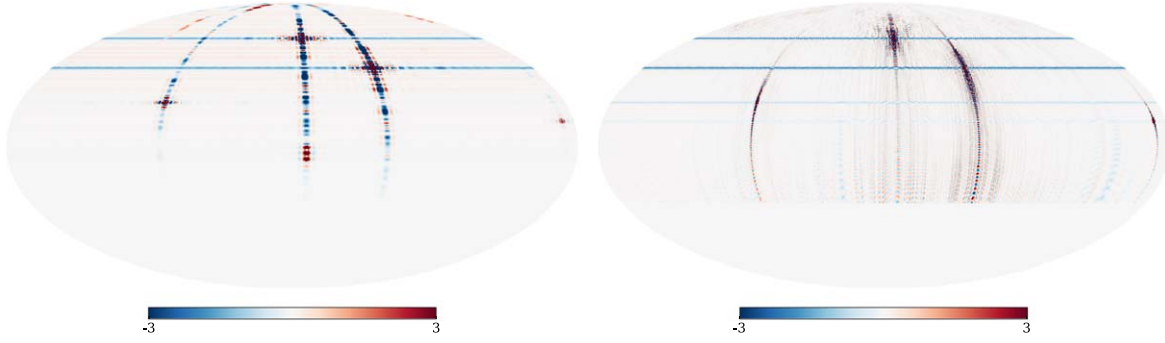


Figure 9. The dirty map (left) and reconstructed map (right) for the case of four bright sources (except the bright sources, all sky radiation is set to zero in the input map). The maps have been smoothed to an approximate resolution of Tianlai cylinder pathfinder in the EW direction with a Gaussian beam filter of 0.0133 rad.

where \mathbf{v} , \mathbf{B} , \mathbf{a} , \mathbf{n} are the matrix notation for the $V_m^{ij,\pm}$, $B_{lm}^{ij,\pm}$, a_{lm} , $n_m^{ij,\pm}$ respectively. It is noted that \mathbf{B} is a block diagonal matrix, which allows the operation on Equation (26) can be performed m -by- m , so the operation on large matrices can be split into smaller ones.

Due to the fixed size of the telescope, its sensitivity to larger l and m decreases rapidly beyond a scale, which reduces Equation (24) to a finite sum. We limit the largest measurable spherical harmonic degree l by the largest dimension of the array D_{\max} , and m is determined by the largest distance in the E-W direction D_{E-W} , which are taken as $l \lesssim 2\pi D_{\max}/\lambda$ and $m \lesssim 2\pi D_{E-W}/\lambda$. For Tianlai cylinder pathfinder at 750 MHz, we take $l_{\max} = 742$ and $m_{\max} = 715$ in our simulation.

The least-squares solution to Equation (26) which minimizes $\|\mathbf{v} - \mathbf{B}\mathbf{a}\|^2$ is

$$\hat{\mathbf{a}}_{\text{LS}} = (\mathbf{B}^* \mathbf{B})^{-1} \mathbf{B}^* \mathbf{v}, \quad (27)$$

as there are unmeasured modes of the sky due to incomplete uv -coverage, the rank of matrix \mathbf{B} is usually not full, therefore the matrix $(\mathbf{B}^* \mathbf{B})$ is not invertible. A solution of least-squares solution can be written in terms of the Moore–Penrose pseudo-inverse (denoted with $+$) as

$$\hat{\mathbf{a}}_{\text{LS}} = \mathbf{B}^+ \mathbf{v}. \quad (28)$$

The Moore–Penrose pseudo-inverse regularizes these singular modes (singular values below a threshold) to zero. The threshold is chosen to ensure a good reconstruction. For maps with low noise, low threshold could be adopted for the inversion. It would be determined by some experiment.

4.2. Map Reconstruction

To test this algorithm, we first generate a simple sky map, in which all pixels are set to zero, except for a few bright points placed at different places in the sky, and run the simulation. In the left panel of Figure 9, we show the dirty map, which is obtained by

$$\mathbf{a}_{\text{dirty}} = \mathbf{B}^* \mathbf{v}. \quad (29)$$

In the right panel of Figure 9, we show the reconstructed map obtained from Equation (28), which is smoothed to an approximate resolution of Tianlai cylinder pathfinder in the EW direction with a Gaussian beam filter with FWHM $\sim 0.3997 \text{ m}/30 \text{ m} \approx 0.0133 \text{ rad}$, corresponding to the maximum resolution of the beam in the East–West direction.

The dirty map reflects the response of the array for the drift scan observation. We can see that there are very extended sidelobes along the North–South direction, and also shorter sidelobes in the East–West direction at each of the bright radio sources. These are basically along the longitude and latitude lines, due to the structure of the cylinder array and its drift scan mode of observation. The sidelobes have complicated structures, which would be frequency-dependent, and can affect the extraction of the 21 cm signal.

In the reconstructed map, the instrument response is deconvolved, though it is not perfect and depends on the deconvolution algorithm. We see that in the reconstructed image, the sidelobes are not as strong as in the dirty map, especially the sidelobes along the East–West direction are significantly reduced. However, the sidelobe is not completely eliminated, especially in the North–South direction. Compared with the dirty map, the position of the sidelobe actually shifted, showing that while reducing the sidelobe in some parts of the map, the imperfect deconvolution may also introduce new fake features at other parts of the map.

There are also lines along the whole latitude lines passing through the bright sources, but they are not sidelobes. These are produced because we have subtracted the $m = 0$ modes, which do not vary and would be subtracted as correlated noise in the real data. However, this means practically that the average along each latitude must be zero, so for the latitudes where the bright sources are located, the other places are set to a negative value.

We then generate an input map from the sky model, and also reconstructed maps with the mock data. In this work, we set the Moore–Penrose pseudo-inverse threshold to be $\epsilon \times \max(\sigma_i)$, where σ_i is the singular values of matrix \mathbf{B} , and ϵ is a preset

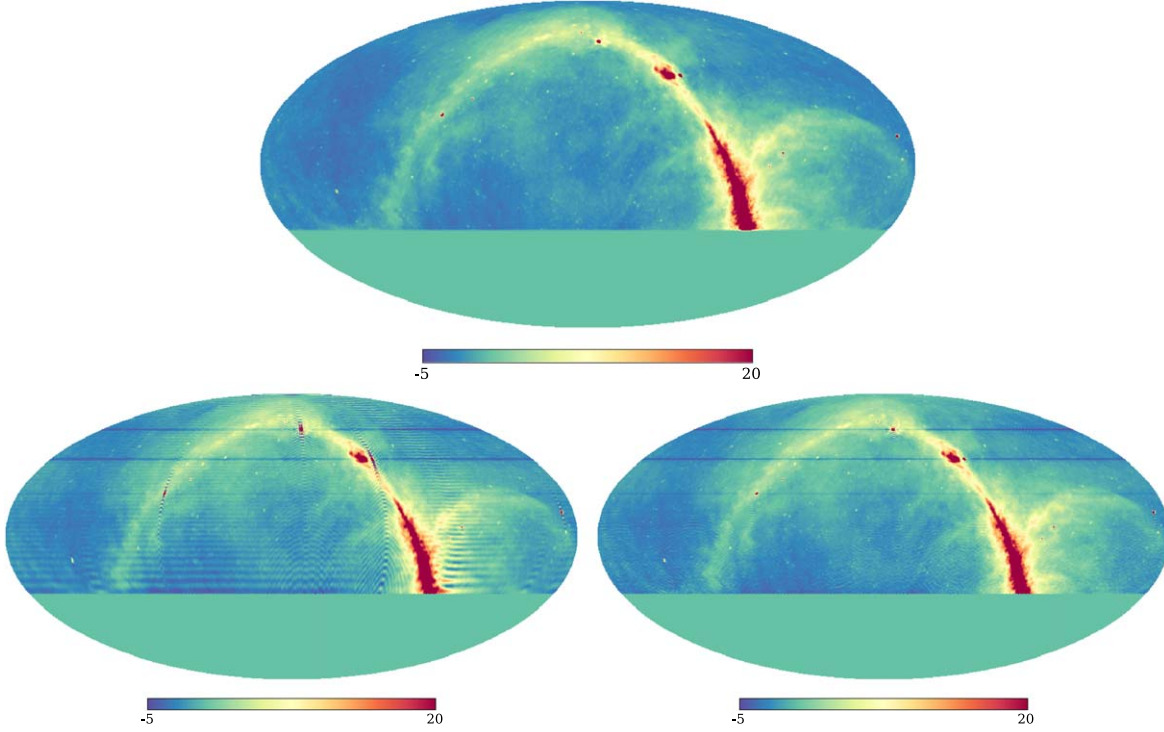


Figure 10. The filtered input sky map (top) and the reconstructed sky maps (bottom) in Mollweide project, the area below south of -30° which is close to or below the horizon of the location of Tianlai array is masked. Bottom left: the reconstructed map from the data without gain variation and noise with threshold $\epsilon = 10^{-3}$; Bottom right: $\epsilon = 10^{-5}$. All maps have been smoothed as we do in Figure 9.

threshold value, which we take here as 10^{-3} . The resulting maps are plotted in Figure 10 in the Mollweide projection, the area below south of -30° is mask in consideration of the field of view of Tianlai Array.

As our telescope is sensitive only to modes with $l < 600$ (this is a conservative number, it is likely smaller than this), even with perfect reconstruction we will still loose the $l > 600$ modes, we therefore compare with a map made from the original sky map shown in Figure 3, but with all modes of $l > 600$ removed, which we will refer to as the (low-pass) filtered input map, even though the actual input map which we use to generate the visibilities does have the $l > 600$ modes. This is done by first computing the spherical harmonics coefficients of it, setting those components with $l > 600$ to zero, and transform back. Similarly, we also discard all the $m = 0$ modes, which are prone to error as they do not vary over time. This also removes the $l = 0$ mode, i.e., the global mean temperature of the map. As a result, some pixels of the map become negative. This map is shown in the top panel of Figure 10.

In the bottom left of Figure 10 we show the reconstructed sky maps in the ideal case with threshold $\epsilon = 10^{-3}$, i.e., from the mock visibilities generated by the sky model, without any noise or gain variation. The reconstructed map is also smoothed with a Gaussian filter with FWHM ~ 0.0133 rad. As we can

see, there are differences from the input map. Besides the longitudinal sidelobes and negative latitude lines around the bright points which we noted in Figure 9, another notable feature is the horizontal stripes extended from the brighter part of the galactic plane, which form a comb-like structure.

In the bottom right of Figure 10, we plot the reconstructed map in the ideal case with a smaller threshold $\epsilon = 10^{-5}$, for this case, the comb-like artifact near the galactic plane disappears, and generally a more accurate map is reconstructed. This is because when we impose the cut off threshold, modes with small eigenvalues or singular values are discarded. These modes however also contain some information about the sky. Comparing with the input map, setting these modes to zero is equivalent to adding their negative. Unlike modes generated by noise, however, these modes are regular and show up clearly in the map. In the left panel of Figure 11, we plot the difference between the maps of the two thresholds. The differences show most clearly near the brightest part of the sky, i.e., near the bright point sources, and the brightest part of the galactic plane. The comb-like structure are mainly due to missing small m (e.g., $m < 10$) modes caused by the pseudo-inverse truncation. If we denote these missing small m (e.g., $m < 10$) modes as Δa_{lm} , the missing map component would be $\sum_{lm} \Delta a_{lm} Y_{lm}(n)$. We plot the map made with only the $m < 10$ and $l < 200$ modes in the right panel of Figure 11, and one can see clearly the

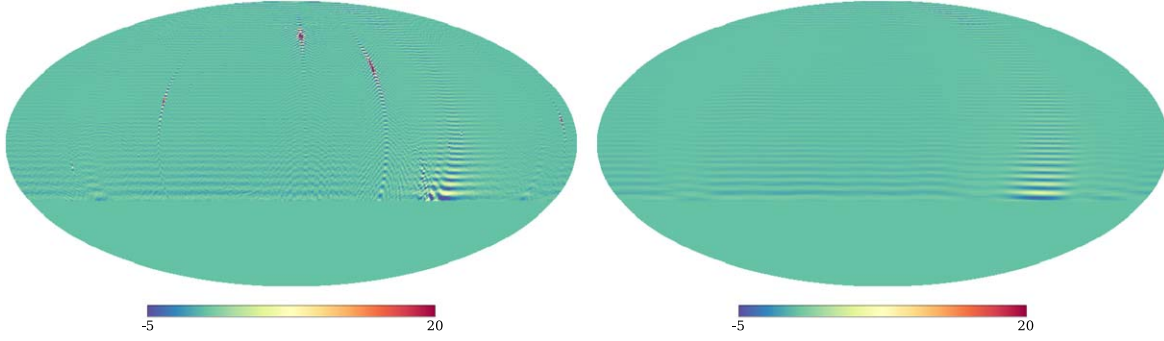


Figure 11. Left: The difference of the two maps in the bottom of Figure 10. Right: The difference of the two reconstructed maps with the $0 < m < 10$ and $l < 200$ modes only.

comb-like structure. On the other hand, the structure around the few bright point sources does not show up in this case, which shows that they are generated mostly by the high m modes.

We also show the spherical harmonic coefficients a_{lm} for the difference map of $\epsilon = 10^{-3}$ and $\epsilon = 10^{-5}$ case in Figure 12. As can be seen from the figure, the largest a_{lm} of this differential plot are those $m < 10$ modes, concentrated near the ℓ -axis. We expect such small m modes will produce sizable features along the latitude lines, which would appear as the comb-like structure we see in the reconstructed map with $\epsilon = 10^{-3}$.

From the above analysis, we see that a smaller ϵ is desirable in the reconstruction. However, in the presence of noise, the value of ϵ cannot be too small, otherwise the reconstructed map will be affected greatly by the noise, often rendering the whole solution unusable. The threshold $\epsilon = 10^{-3}$ ensures that even with relatively large noise, for example those for a single day observation, the reconstruction could still proceed smoothly. However, if the noise in the visibility could be reduced by adding the the data collected at the same local sidereal time (LST) on different days, it may be possible to impose smaller ϵ value and obtain more accurate maps.

4.3. Calibration Error and Noise

With the map-making error understood, we now consider the impact of gain variation and noise. We use the method outlined in Section 3 to calibrate the gains, and use the calibrated visibilities to make the reconstructed map, as shown by the left panel of Figure 13. This map is very similar to the reconstructed map in the ideal case shown in Figure 10. We also add noise, and run through the same procedure in the presence of noise, and the result is shown in the right panel of Figure 13. The noise in the visibility measurement also translates to fluctuations in the map.

To see the impact of these more clearly, we plot the differences of these maps in Figure 14, using the same color scale. In the top left we show the difference between the reconstructed map in the ideal case and the input map. As noted

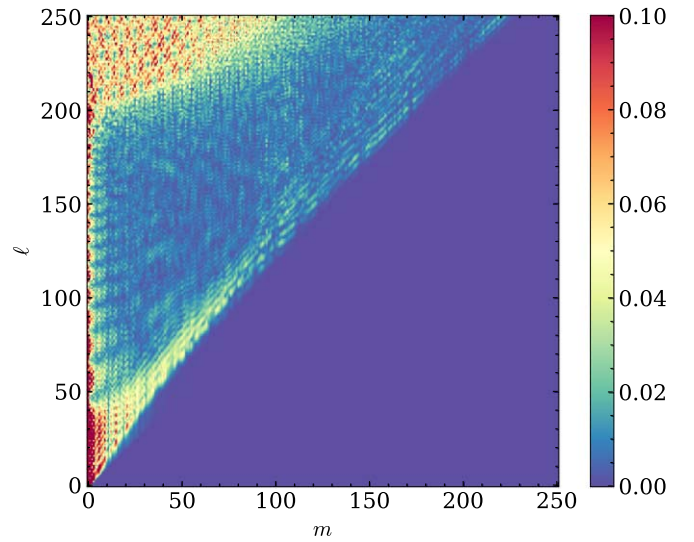


Figure 12. The amplitude of the spherical harmonics a_{lm} of the difference map.

in Section 4.2, there are large differences near the bright sources, and we can also see the horizontal stripes including the comb-like structure near the galactic plane.

In the top right panel, we plot the difference between the reconstruction with gain variation but without making the relative calibration. The differences are small, as the gain variations are not large by itself. Indeed, we found from the actual observational data during night that even without the relative calibration, we can still make quite good maps. In the current simulation we have adopted a gain fluctuation model which have fluctuation level similar to the real case, so it is not surprising to find this result. The residue difference in the map are again located mainly at the bright spots, or their sidelobes which are along the longitude and latitude lines. Once relative calibration is done, these residue differences also disappear (bottom left), showing that the relative calibration can effectively mitigate this error.

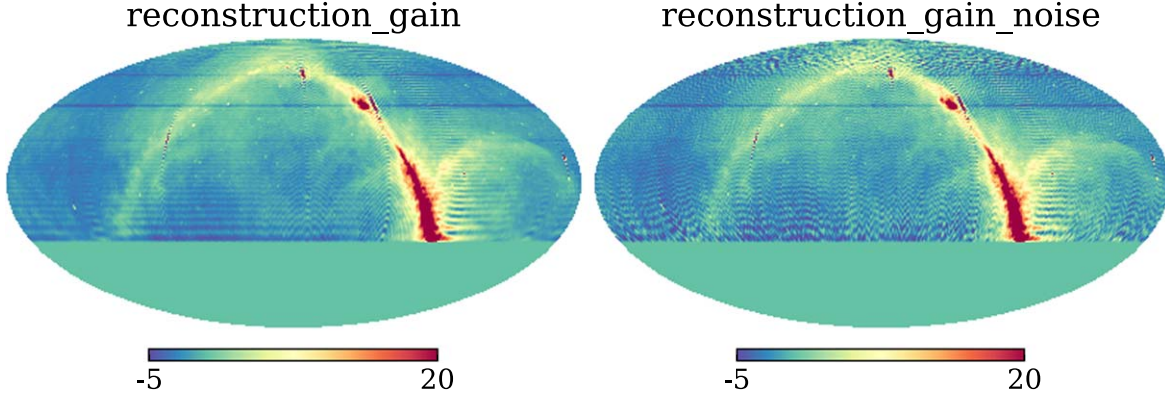


Figure 13. Left: the reconstructed map with complex gain variation after calibration. Right: the reconstructed map with complex gain variation and noise after calibration.

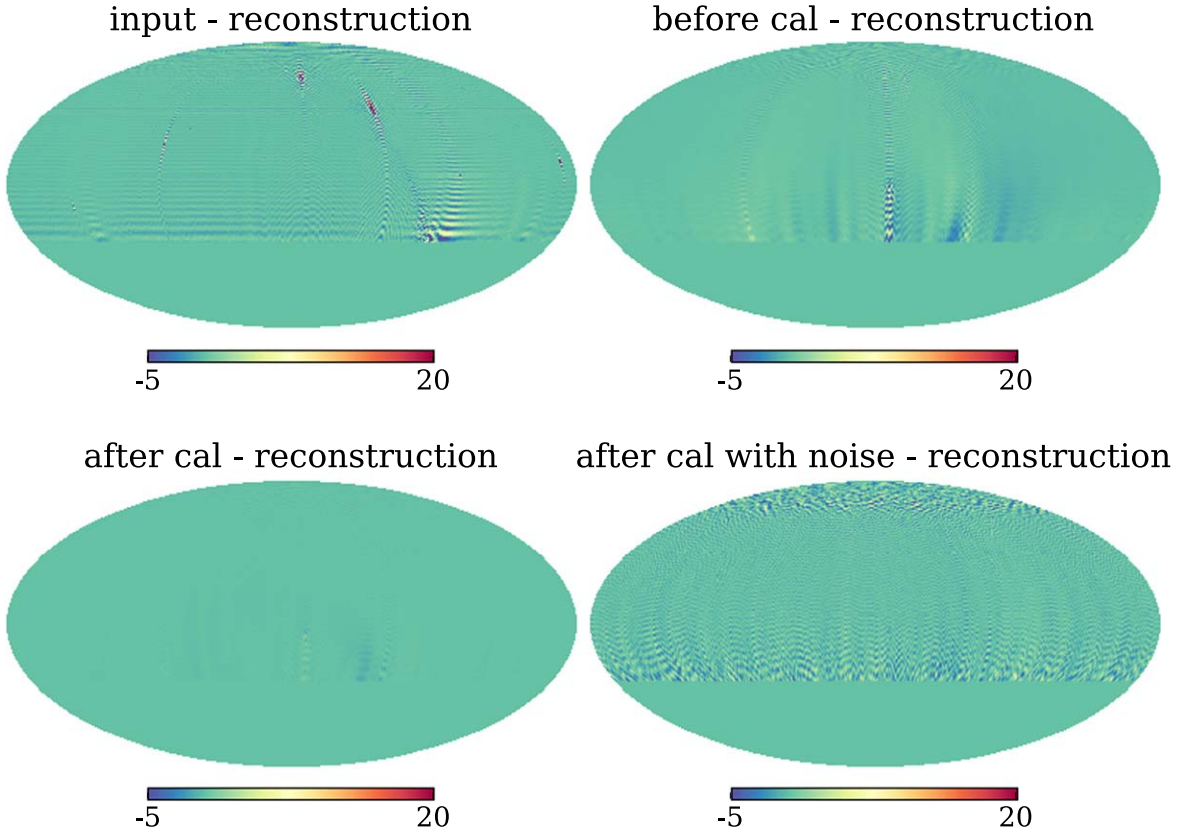


Figure 14. The difference between reconstructed maps.

Finally, in the bottom right, we show the difference in the reconstructed map with both calibrated gain variation and thermal noise. The impact of the noise is uniformly distributed across all right ascensions, but as we can see from the figure it is not uniform in the North–South direction. All the full-sky maps are computed using HEALPix and illustrated with Mollweide projection, thus it is not due to unequal pixel area.

The noise is lowest around mid-latitude ($\sim 45^\circ$), and increases toward south and north. This suggests that it is due to the sensitivity of the telescope, as it has the best sensitivity for the part of the sky which is right above the telescope, but the sensitivity degrades on the two sides. Note here we plotted the case for the error of one day's data, and its magnitude could be reduced by stacking more days of observations.

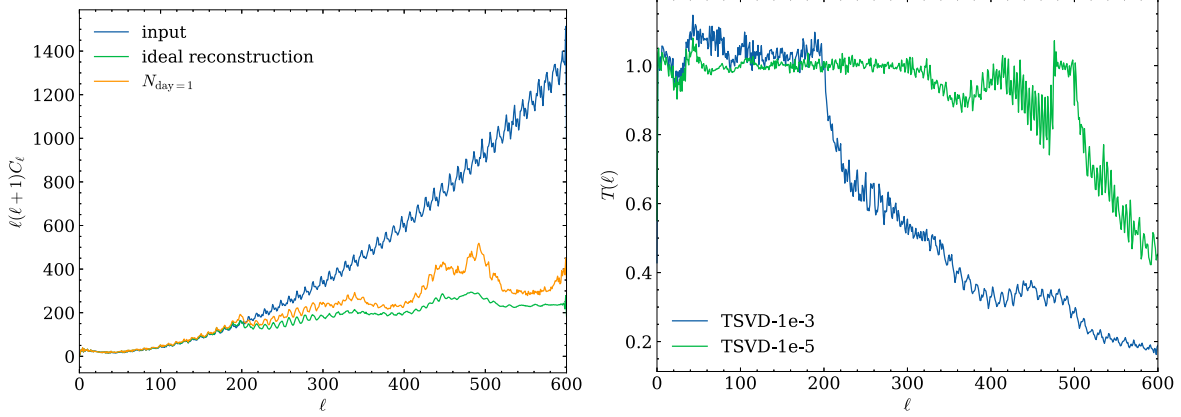


Figure 15. Left: The angular power spectrum for the input (blue) and reconstructed maps from the data with complex gain variation and noise (orange), and the ideal case (green, i.e., calibrated perfectly, noise-free) with $\epsilon = 10^{-3}$. Right: The transfer function $C_\ell/C_\ell^{\text{input}}$ for the ideal case with $\epsilon = 10^{-3}$ and 10^{-5} .

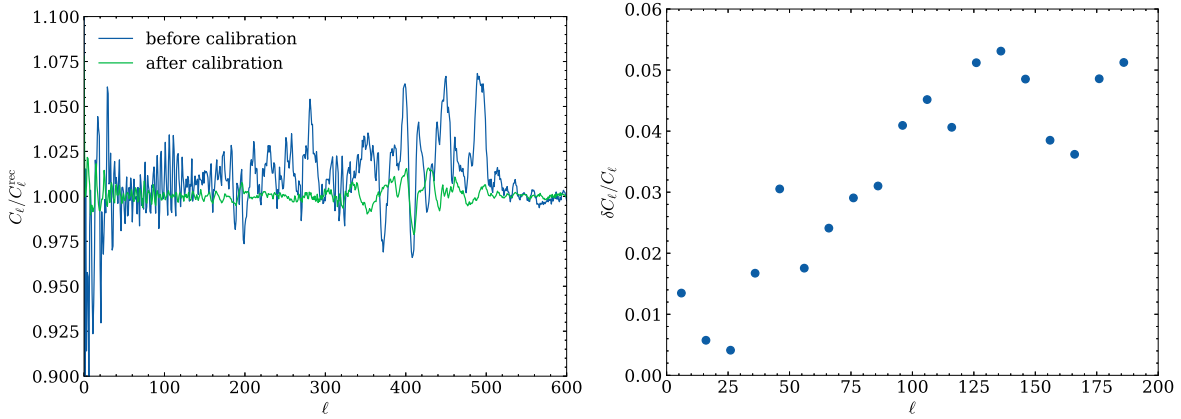


Figure 16. The impact of gain variation and noise. Left: ratio of angular power spectrum with gain variation to the ideal case, before and after relative calibration. Right: The angular band power for $\Delta\ell = 10$ induced by the noise.

4.4. Angular Power Spectrum

Finally, we assess how these errors would affect the angular power spectrum, which is defined as $C_l \delta_{ll'} \delta_{mm'} \equiv \langle a_{lm} a_{l'm'}^* \rangle$, and under the ergodic assumption,

$$C_l = \frac{1}{2l+1} \sum_{m=-l}^l |a_{lm}|^2. \quad (30)$$

An optimal estimation of the angular power spectrum for incomplete sky coverage has been discussed in Alonso et al. (2019). Here, we simply calculate the angular power spectrum of the input map and the reconstructed map by masking the part south of -30° latitude.

In the left panel of Figure 15 we plot the angular power spectrum for the input map (blue curve), for the reconstructed map in the ideal case (green curve), and the reconstructed map with calibrated gain variation and a noise corresponding to one day observation (orange curve), where $\epsilon = 10^{-3}$ is applied

when reconstructing the maps. The wiggles on the angular power spectrum are produced by some point sources in the sky. In the right panel of Figure 15 we plot the transfer function $T(l) = C_l/C_l^{\text{input}}$, which quantifies the amount of loss in the angular power spectrum under ideal condition for $\epsilon = 10^{-3}$ and 10^{-5} .

We can see that generally, at $l \lesssim 200$ the angular power spectrum is well recovered, and the transfer function $T(l) \sim 1$ for both ϵ values, which means that most of the modes are well recovered, though there are fluctuations due to the reconstruction errors discussed in Section 4.2. At $l \gtrsim 200$, however, for our fiducial case of $\epsilon = 10^{-3}$, the angular power spectrum falls systematically below that of the input map, and the transfer function drops significantly below 1, showing that at such angular scales and precision the spherical harmonics are not well recovered. For the $\epsilon = 10^{-5}$ case, the transfer function remains flat at $l \gtrsim 200$, only begins to decline significantly at $l \gtrsim 500$. This is because the discarded modes have small

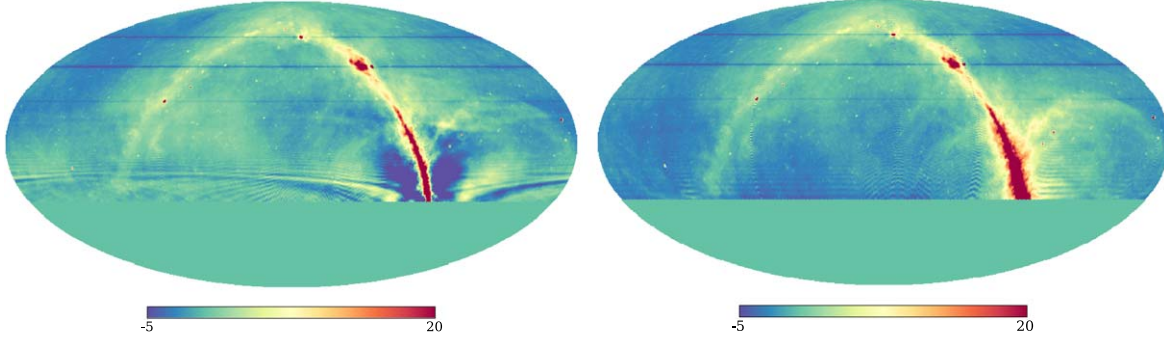


Figure 17. The reconstructed map with threshold $\epsilon = 10^{-3}$ for the two expanding configurations. Left: configuration 1; Right: configuration 2.

singular values, which would be unrecoverable in the presence of noise.

Note that the mode probed by a baseline b is $l \lesssim 2\pi b/\lambda$. For our array, the longest baseline in the North–South direction is about 12.4 m, corresponding to $l \sim 195$. Thus, the array could not probe some of the modes at $l > 195$ in the North–South direction, and in reconstruction such modes will be missing, reducing the total power. This may be the reason why the angular power spectrum of the reconstructed map falls systematically below that of the input.

Next we consider the impact of the gain variation and noise. In the left panel of Figure 16 we plot the ratio of the angular power spectrum with gain variation to that of the ideal case. The absolute calibration is performed to determine the initial phase. We plot both the case without relative calibration (blue curve, labeled as “before calibration”) and the case with relative calibration (green curve, labeled as “after calibration”). We can see that in both cases the average of the angular power spectrum ratio is around 1. There are 10^{-2} level fluctuations in the angular power spectrum ratio for the case without relative calibration, showing that even without the relative calibration, the array could still make reasonably good measurement. This fluctuation is significantly reduced with the introduction of relative calibration.

The addition of noise does not significantly change the angular power spectrum, even for the relatively large noise level of $N_{\text{day}} = 1$, i.e., simply the data of one day. One can expect this error would be further reduced with integration. The angular power spectrum with noise has a shape similar to the reconstructed angular power spectrum in the ideal case, namely it is well recovered at $l \lesssim 200$, though slightly larger than the ideal case, which is as would be expected for a random noise power. We plot the band power $\delta C_l / C_l^{\text{recon}}$ for bins of $\Delta l = 10$ in the right panel of Figure 16, where $\delta C_l = C_l^{+\text{noise}} - C_l^{\text{recon}}$ is the power induced by the noise. Here we have included the effect of gain variation which has been corrected by the relative calibration. This band power is at a few percent level for a one

day observation, but it can be reduced with longer integration time.

4.5. Reconfiguration

The analysis of the angular power spectrum transfer function in the last subsection indicates that a lack of long North–South baselines may be the reason for the error in reconstruction at $l \gtrsim 200$. The Tianlai pathfinder cylinder is 40 m long, and the reason we have only used 12.4 m in the center section of the cylinder is due to the limited funding available, and the choice of one wavelength for the shortest spacing. As a result, the longest North–South baseline is much shorter than that of the East–West (30 m). To check whether this is indeed responsible for the abrupt drop of transfer function at $l \gtrsim 200$, we consider two cases for a different array configuration, where the maximum baseline length in the North–South is about 30 m, the same as in the East–West:

1. Configuration 1: we simply increase the spacing of feeds on each cylinder to $30 \div 12.4 \approx 2.42$ times the original spacing;
2. Configuration 2: we keep the position of central feeds on each cylinder unchanged, but move the six outer feeds on two ends further away, and the distance to the outermost feed of the central part follows a geometric progression $d_i = \Delta d^0 q^i$ ($i = 1, 2, 3, 4, 5, 6$), where Δd^0 is the current spacing between feeds, $q = \left(\frac{9 + 6\Delta d^0}{\Delta d^0} \right)^{1/6}$.

In Figure 17, we plot the reconstructed map with $\epsilon = 10^{-3}$ for the two configurations in the ideal(noise-free) case, and in Figure 18 we plot the transfer function of angular power spectrum for the two cases.

For configuration 1, the comb-like structure in Figure 10 disappeared, but the southern edge of the map is not well reconstructed, there is a deep blue region around the brightest part of the galactic plane, where it should be bright. We also see that the transfer function is much improved at $l \gtrsim 200$, and even for $\epsilon = 10^{-3}$ it only starts to deviate from one at $\gtrsim 450$.

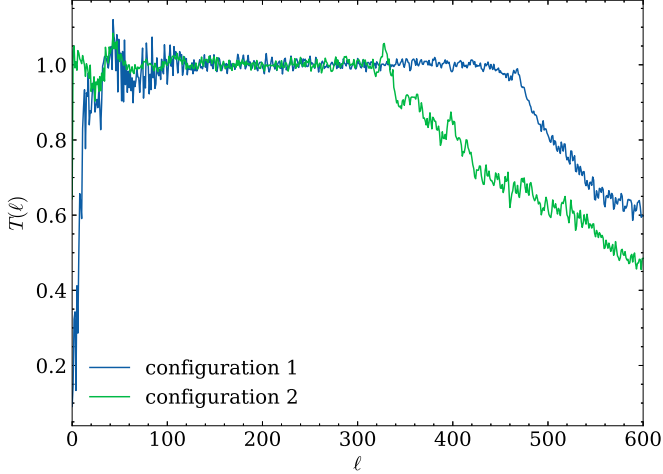


Figure 18. The transfer function $C_\ell/C_\ell^{\text{input}}$ with $\epsilon = 10^{-3}$ for the two North–South expanding configurations.

However, there is a deep pit at $l < 20$. This shows that with all North–South spacing increased by the same 2.42 factor, there is a dearth of short North–South baselines which probe the large scale variation in the North–South direction. It seems that the new defect in the reconstructed map is due to the missing of these short baselines.

For configuration 2, where the maximum baseline along the North–South direction is equal to that of East–West, but some short baselines are preserved, the structure around the galactic plane is much better reconstructed. Even though we still adopted $\epsilon = 10^{-3}$, comparing with Figure 10 the comb-like artifact is significantly reduced. Compared with configuration 1, the deviation of transfer function for the configuration 2 starts at a lower l value, but the $l < 20$ modes are better reconstructed, as indicated by the fact that the transfer function remains near one at small l . So we conclude that at least for a single frequency, the array in the new configuration 2 has better performance than both the original configuration and configuration 1.

5. Conclusion

We have made a simulation for the observation of the Tianlai cylinder pathfinder array. This is done by first generating mock visibility from a `cora` based sky model, and an unpolarized telescope beam model. The mock visibilities are then processed through the data processing pipeline `tlpipe`, which has been used in the processing of the real data. In addition to the ideal case, we have also simulated random drifting of the complex gains of the array elements by Gaussian process, and the thermal noise according to the measured system temperature. To correct for the variation of the complex gains, we also simulated the absolute and relative calibrations used in the Tianlai array. The absolute calibration is performed by making

transit observation of the bright source Cyg A, which gives both the amplitude and phase of the complex gain. The relative calibration is performed by measuring a periodically broadcasting artificial calibrator noise signal to track the change of instrument phase. We have also added measurement noise for these measurements. Our simulation shows that the current absolute calibration algorithm produces an error at the $10^{-4} \sim 10^{-3}$ radian level. This error is due largely to the fact that the actual sky is not dominated by a single bright source, even when it is transiting. The error induced by the thermal noise in the subsequent relative calibration is relatively small, only at the 10^{-5} rad level. In order to further improve the measurement precision of the array, it is important to make improve the absolute calibration algorithm, probably by taking the additional sky source contribution into account.

We then use the mock data to make maps of the sky by using the m -mode analysis, where the observed visibility data are decomposed into m -modes, and the spherical harmonics of the sky temperature fluctuations are solved from these m -modes. This solution has singularities due to incomplete uv coverage. In the present work, we have used the Moore–Penrose pseudo-inverse method, to obtain the solution. In this method, modes with singular values smaller than a certain threshold are discarded. Due to the incomplete sampling of the uv coverage, the reconstruction of the sky map is not perfect. There are sidelobes along the longitudinal lines. The $m = 0$ modes which do not vary with time cannot be recovered well from observation, and if we set these $m = 0$ modes to zero, negative stripes along the latitude line passing through bright sources are generated. Furthermore, the Moore–Penrose pseudo-inverse algorithm may introduce some artifacts in the reconstructed map. While it is possible to suppress such artifact by adopting a smaller threshold parameter ($\epsilon < 10^{-5}$), in the presence of noise this is impractical. We will discuss other map-making algorithms elsewhere. In the present study, we focus on the map error induced by the calibration process and the noise. We find that the calibration error induced relatively small error in the map, which are most prominent near bright sources. The thermal noise is minimum in the part of the sky which is right above the telescope location, i.e., around decl. of $\delta = 44^\circ$, but increases toward north and south, where the sensitivity of the telescope degrades.

We also investigated the angular power spectrum measurement. We find that the angular power spectrum is well recovered at $l \lesssim 200$. However, for $l > 200$, noise could introduce large error in the reconstructed map, and the angular power spectrum is systematically smaller for the higher modes. This is not surprising, as a baseline of length b probes only modes within $l < 2\pi b/\lambda$, and the maximum length of the Tianlai cylinder pathfinder array baseline in the North–South direction is 12.4 m, corresponding to $l \sim 195$. For higher l , there are modes which will not be sampled by the current Tianlai cylinder pathfinder array. We considered reconfigure

the array, and find that an improvement can be achieved by moving six feeds on each end of the cylinder to increasingly larger spacings, to achieve longer baseline in the North–South direction, while preserving the shorter spacing in the center section of the cylinder. We also studied the impact of the thermal noise, and find that even in a single day observation it would be small compared with sky signal.

The present investigation has some limitations. We have only studied the case of a single frequency, and ignored polarization. We have also limited the simulation to the current working algorithm of the Tianlai pipeline `tlpipe`, which may not be optimal. It is conceivable that these algorithms could be further improved to reduce the numerical error. The array elements are treated as independent and identical. In fact, there are also cross-couplings between the different array elements, and each array element also has a slightly different response, which make the response of the telescope more complicated and introduce nonlinearity. Such effects would be investigated in future studies.

Acknowledgments

We thank Profs. Reza Ansari, Albert Stebbins, and Peter Timbie for discussions. This work is supported by the National SKA Program of China, No. 2022SKA0110100. Some of the simulation work is supported by the National Natural Science Foundation of China under grant No. 1 220 030 249.

References

- Alonso, D., Sanchez, J., Slosar, A. & LSST Dark Energy Science Collaboration 2019, *MNRAS*, **484**, 4127
- Battye, R. A., Browne, I. W. A., Dickinson, C., et al. 2013, *MNRAS*, **434**, 1239
- Blake, C., & Glazebrook, K. 2003, *ApJ*, **594**, 665
- Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2009, *ApJ*, **695**, 183
- Bull, P., Ferreira, P. G., Patel, P., & Santos, M. G. 2015, *ApJ*, **803**, 21
- Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, *Natur*, **466**, 463
- Chang, T.-C., Pen, U.-L., Peterson, J. B., & McDonald, P. 2008, *PhRvL*, **100**, 091303
- Chapman, E., Zaroubi, S., Abdalla, F. B., et al. 2016, *MNRAS*, **458**, 2928
- Chen, X. 2012, *IJMPs*, **12**, 256
- CHIME Collaboration, Amiri, M., & Bandura, K. 2023, *ApJ*, **947**, 16
- 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. 2023, *MNRAS*, **518**, 6262
- Di Matteo, T., Perna, R., Abel, T., & Rees, M. J. 2002, *ApJ*, **564**, 576
- Eastwood, M. W., Anderson, M. M., Monroe, R. M., et al. 2018, *AJ*, **156**, 32
- Eastwood, M. W., Anderson, M. M., Monroe, R. M., et al. 2019, *AJ*, **158**, 84
- Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, *ApJ*, **633**, 560
- Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, *PhR*, **433**, 181
- Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, *ApJ*, **622**, 759
- Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1982, *A&AS*, **47**, 1
- Hu, W., Wang, X., Wu, F., et al. 2020, *MNRAS*, **493**, 5854
- Kerrigan, J. R., Pober, J. C., Ali, Z. S., et al. 2018, *ApJ*, **864**, 131
- Li, J., Zuo, S., Wu, F., et al. 2020, *SCPMA*, **63**, 129862
- Li, Y., Wang, Y., Deng, F., et al. 2023, *ApJ*, **954**, 139
- Liu, A., & Shaw, J. R. 2020, *PASP*, **132**, 062001
- Liu, A., & Tegmark, M. 2012, *MNRAS*, **419**, 3491
- Masui, K. W., Switzer, E. R., Banavar, N., et al. 2013, *ApJL*, **763**, L20
- Mertens, F. G., Ghosh, A., & Koopmans, L. V. E. 2018, *MNRAS*, **478**, 3640
- Miville-Deschênes, M. A., Ysard, N., Lavabre, A., et al. 2008, *A&A*, **490**, 1093
- Paul, S., Santos, M. G., Chen, Z., & Wolz, L. 2023, arXiv:2301.11943
- Pritchard, J. R., & Loeb, A. 2012, *RPPh*, **75**, 086901
- Santos, M., Bull, P., Alonso, D., et al. 2015, *Advancing Astrophysics with the Square Kilometre Array (ASKA14)*, **19** (Giardini Naxos: POS)
- Seo, H.-J., & Eisenstein, D. J. 2003, *ApJ*, **598**, 720
- Shaw, J. R., Sigurdson, K., Pen, U.-L., Stebbins, A., & Sitwell, M. 2014, *ApJ*, **781**, 57
- Shaw, J. R., Sigurdson, K., Sitwell, M., Stebbins, A., & Pen, U.-L. 2015, *PhRvD*, **91**, 083514
- Sun, S., Li, J., Wu, F., et al. 2022, *RAA*, **22**, 065020
- Villaescusa-Navarro, F., Alonso, D., & Viel, M. 2017, *MNRAS*, **466**, 2736
- Wang, X., Tegmark, M., Santos, M. G., & Knox, L. 2006, *ApJ*, **650**, 529
- Wu, F., Li, J., Zuo, S., et al. 2021, *MNRAS*, **506**, 3455
- Zhang, J., Ansari, R., Chen, X., et al. 2016a, *MNRAS*, **461**, 1950
- Zhang, J., Zuo, S.-F., Ansari, R., et al. 2016b, *RAA*, **16**, 158
- Zheng, H., Tegmark, M., Dillon, J. S., et al. 2017, *MNRAS*, **465**, 2901
- Zuo, S., Li, J., Li, Y., et al. 2021, *A&C*, **34**, 100439
- Zuo, S., Pen, U.-L., Wu, F., et al. 2019, *AJ*, **157**, 34