







Publication Year	2022
Acceptance in OA @INAF	2022-09-02T09:32:15Z
Title	The Correlation Calibration of PAPER-64 data
Authors	Tamirat G. Gogo; Yin-Zhe Ma; Piyanat Kittiwisit; Jonathan L. Sievers; Aaron R. Parsons; et al.
DOI	10.1093/mnras/stab3516
Handle	http://hdl.handle.net/20.500.12386/32541
Journal	MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY
Number	510

The correlation calibration of PAPER-64 data

Tamirat G. Gogo,^{1,2} Yin-Zhe Ma^{1,2}  ^{1,2}★ Piyanat Kittiwisit,^{1,2} Jonathan L. Sievers,^{1,3} Aaron R. Parsons,^{4,5} Jonathan C. Pober,⁶ Daniel C. Jacobs,⁷ Carina Cheng,⁴ Matthew Kolopanis^{1,7} ,⁷ Adrian Liu,^{8,9} Saul A. Kohn,¹⁰ James E. Aguirre^{1,10} ,¹⁰ Zaki S. Ali,⁵ Gianni Bernardi,^{11,12,13} Richard F. Bradley,^{14,15,16} David R. DeBoer,⁵ Matthew R. Dexter,⁵ Joshua S. Dillon^{1,4} ,⁴ Pat Klima,¹⁵ David H. E. MacMahon,⁵ David F. Moore,¹⁰ Chuneeta D. Nunhokee,⁴ William P. Walbrugh¹³ and Andre Walker¹¹

¹School of Chemistry and Physics, University of KwaZulu-Natal, Westville Campus, Private Bag X54001, Durban, 4000, South Africa

²NAOC-UKZN Computational Astrophysics Centre (NUCAC), University of KwaZulu-Natal, Durban, 4000, South Africa

³Department of Physics, McGill University, 3600 University Street Montreal, QC H3A 2T8, Canada

⁴Department of Astronomy, UC Berkeley, Berkeley, CA 94720, USA

⁵Radio Astronomy Laboratory, UC Berkeley, Berkeley, CA 94720, USA

⁶Department of Physics, Brown University, Providence, RI 02912, USA

⁷School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287, USA

⁸Department of Physics and McGill Space Institute, McGill University, Montreal, QC H3A 2T8, Canada

⁹CIFAR Azrieli Global Scholar, Gravity & the Extreme Universe Program, Canadian Institute for Advanced Research, 661 University Ave., Suite 505, Toronto, Ontario M5G 1M1, Canada

¹⁰Department of Physics and Astronomy, U. Penn., Philadelphia PA 19104, USA

¹¹INAF-Istituto di Radioastronomia, via Gobetti 101, I-40129, Bologna, Italy

¹²Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown, 6140, South Africa

¹³South African Radio Astronomy Observatory, Black River Park, 2 Fir Street, Observatory, Cape Town, 7925, South Africa

¹⁴Department of Electrical and Computer Engineering, U. Virginia, Charlottesville VA 22904, USA

¹⁵National Radio Astronomy Obs., Charlottesville VA 22904, USA

¹⁶Department of Astronomy, U. Virginia, Charlottesville VA 22904, USA

Accepted 2021 November 29. Received 2021 November 24; in original form 2021 September 27

ABSTRACT

Observation of redshifted 21-cm signal from the Epoch of Reionization (EoR) is challenging due to contamination from the bright foreground sources that exceed the signal by several orders of magnitude. Removal of this very high foreground relies on accurate calibration to keep the intrinsic property of the foreground with frequency. Commonly employed calibration techniques for these experiments are the sky model-based and the redundant baseline-based calibration approaches, which can suffer from sky-modelling error and array redundancy imperfection respectively. In this work, we introduce the hybrid *correlation calibration* (CORRCAL) scheme, which aims to bridge the gap between redundant and sky-based calibration by relaxing redundancy of the array and including sky information into the calibration formalisms. We apply the CORRCAL to the data of Precision Array for Probing the Epoch of Reionization (PAPER) experiment, which was pre-calibrated using redundant baseline calibration. We show about 6% suppression at the bin right on the horizon limit of the foreground wedge-like structure, relative to the pre-calibrated power spectra. This small improvement of the foreground power spectra around the wedge limit could be suggestive of reduced spectral structure in the data after CORRCAL calibration, which lays the foundation for future improvement of the calibration algorithm and implementation method.

Key words: instrumentation: interferometers – methods: observational – telescopes – cosmology: dark ages, reionization, first stars.

1 INTRODUCTION

The cosmic dark ages ended at around 100 million years after the big bang, marking the formation of the first structures such as galaxies and stars. These objects started to emit radiation that influences the nearby primordial intergalactic medium (IGM), which predominantly filled with the neutral hydrogen atom. As the number

of these very first sources increases, they emitted sufficient radiation to convert the neutral IGM into the fully ionized medium of the hydrogen atom eventually. The period that marks the transition of neutral IGM into the fully ionized state is known as the Epoch of Reionization (EoR) in the history of the Universe. Thus, investigating the era of EoR will answer the fundamental questions of the formation of the first structures and their properties.

The measurement of the radiation from the high-redshifted 21-cm emission line is employed to probe the EoR era. This emission line is due to the hyperfine transition of the neutral hydrogen atom in the

* E-mail: Ma@ukzn.ac.za

IGM empowered by the first sources. The information imprinted into the radiation associated with the redshifted 21-cm line can be used as a powerful tool in the modern cosmology to study the ionization state and large-scale temperature fluctuation of IGM, owing to the formation of the first structures (see e.g. Barkana & Loeb 2001; Oh 2001; Furlanetto, Oh & Briggs 2006; Morales & Wyithe 2010; Pritchard & Loeb 2012).

A direct probe of the EoR is the tomographic mapping of redshifted 21-cm signals from the neutral hydrogen atom (Madau et al. 1996; Barkana & Loeb 2001; Furlanetto et al. 2006; Morales & Wyithe 2010; Pritchard & Loeb 2012). Significant progress has made in the past decades, because several experiments dedicated to studying this signal have been built or are being upgraded. The first-generation 21-cm experiments, including the Murchison Widefield Array (MWA; Tingay et al. 2013), the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER; Parsons et al. 2010), the LOw-Frequency ARray (LOFAR; van Haarlem et al. 2013), and the GiantMeter wave Radio Telescope EoR experiment (GMRT; Paciga et al. 2011) are already operating and taking data. Observations from these first-generation experiments have set upper limits to the statistical power spectrum of 21-cm brightness temperature from cosmic reionization. In particular, results from MWA, LOFAR, and PAPER observations (e.g. Paciga et al. 2011; Dillon et al. 2014; Patil et al. 2017; Cheng et al. 2018; Kolopanis et al. 2019; Mertens et al. 2020) have placed upper limits on the statistical power spectrum of the 21-cm emissions over a broad range of redshifts, providing evidence for the heating of the IGM before reionization. The most recent result from MWA constrained the EoR power spectrum as $\Delta^2 = (43 \text{ mK})^2 = 1.8 \times 10^3 \text{ mK}^2$ for $k = 0.14 h \text{ Mpc}^{-1}$ at $z = 6.5$, which is the tightest measurement currently (Trott et al. 2020; Rahimi et al. 2021). Lessons learned from these experiments have led to the construction of the second-generation instruments, which include the Hydrogen Epoch of Reionization Array (HERA; DeBoer et al. 2017), the upgraded MWA (MWA-II; Wayth et al. 2018), and the LOFAR 2.0 Survey (Edler, de Gasperin & Rafferty 2021). With more collecting areas and improved electronics, high-significant measurements of the 21-cm power spectrum are anticipated from these second-generation instruments (Pofer et al. 2014; Liu & Parsons 2016; Wayth et al. 2018).

Detecting the 21-cm signal from EoR is challenging due to the weakness of 21-cm signal comparing to the brightness temperature of the foreground originated from synchrotron radiation from our Galaxy, strong point sources, and thermal bremsstrahlung from the H II region (Shaver et al. 1999; Santos, Cooray & Knox 2005; Furlanetto et al. 2006; Bernardi et al. 2009; Parsons et al. 2014). To disentangle strong foregrounds from weak 21-cm signal, several mitigation techniques have been developed (see e.g. Wang et al. 2006; Bowman, Morales & Hewitt 2009; Liu et al. 2009; Liu & Tegmark 2011; Parsons et al. 2012; Dillon, Liu & Tegmark 2013; Liu, Parsons & Trott 2014a; Beardsley et al. 2016; Pofer et al. 2016; Hothi et al. 2021). These techniques, in general, can be categorized into two approaches – foreground subtraction and foreground avoidance (Chapman & Jelić 2019). Foreground subtraction involves modelling of the foreground and subtracting them from the data. In contrast, foreground avoidance completely discards foreground contaminated data from analysis and tries to reconstruct the power spectra of EoR in the Fourier regime, which are not affected by foreground.

In both foreground mitigation and subtraction, accurate calibration of the array is critical because calibration errors can contaminate the Fourier mode of the EoR power spectrum (Morales et al. 2012; Ewall-Wice et al. 2017). Furthermore, simulations in Orosz et al. (2018)

suggested that calibration error introduced by quasi-redundancy in antenna positioning and slight variations of the beam responses between different antennas are significant sources of contamination of the EoR signal. Even in the limit of perfect antenna positioning and telescope response, the EoR window can still be contaminated if the instrument calibration is susceptible to sky-modelling error (Byrne et al. 2019).

The array configuration of PAPER is designed to enhance sensitivity for a first power-spectrum detection by measuring same Fourier modes on the sky with a group of redundant baselines. These redundant visibilities need to be calibrated accurately by using the redundancy information of the array. The calibration technique implemented for such kind of array layout is generally known as redundant baseline calibration (Wieringa 1992; Liu et al. 2010). This technique intends to calibrate measurements from the array of antennas configured on a regular grid spacing with maximum redundancy in baseline distribution and tightly packed with the drift-scan mode of observations. Then one can use statistical tool to calibrate the visibilities for independent baselines.

In this work, we introduce a recently developed calibration scheme called *correlation calibration* (CorrCal^1), first proposed in Sievers (2017). CorrCal aims to bridge the gap between sky-based and redundant calibration by incorporating partial sky information and information on the non-redundancies of the array into the calibration solutions. We use this new scheme to re-calibrate data from the 64-element PAPER array (hereafter PAPER64), which is a 9-h integration data set. Ali et al. (2015) calibrated this data set with a redundant baseline calibration scheme and published power spectrum results. This work used the same initial calibration techniques employed in Ali et al. (2015) but has an independent pipeline from that point on. We will show that CorrCal can potentially constrain the wedge-like structure of foreground in the Fourier space.

The rest of paper is organized as follows. In Section 2, we give a general review of common calibration approaches for radio interferometric measurements. In Section 3, we present CorrCal calibration scheme. In Section 4, we apply our method to the 9-h PAPER-64 data set. In Section 5, we present the results of the CorrCal calibration. The conclusion will be in the last section.

2 A REVIEW OF CALIBRATION FORMALISM

In radio interferometry, the correlated signal (visibility) from two independent receiving elements (antennas) is the sum of the system noise and the actual sky signal (or true sky visibility) multiplied by the complex gain parameters of the antennas. The gain parameters depend on the electromagnetic properties of the antennas, whereas the additive noise is mostly dominated by the thermal noise of the receiver system and the sky components.

Performing a calibration is to solve for the complex gain parameters to recover the true sky signal from the measured signal. The mathematical form of this statement is the measurement equation,

$$d_{pq} = g_p g_q^* v_{pq} + n_{pq}. \quad (1)$$

Here, g_p and g_q are the unknown complex gain parameters of antenna p and antenna q , which depend on the frequency f . v_{pq} is the true sky visibility corresponding to the baseline between antenna p and antenna q , for which we ultimately want to recover from the measured visibility (observed data) d_{pq} . n_{pq} is the system noise on this baseline,

¹The prototype version of the software located on <https://github.com/sievers/correal2.git> is being used for this work.

which is usually assumed to be Gaussian. The asterisk superscript * indicates the complex conjugate. All terms in equation (1) are implicitly time-dependent. The antenna gain g can also be written as a complex exponential form with amplitude a and phase ϕ for each of the antenna p (or q),

$$g_p = a_p \exp(i\phi_p). \quad (2)$$

Note that we used 'i' to indicate the imaginary unit $\sqrt{-1}$ throughout the paper.

For multiple baselines, the measurement equation can be written in a more compact form by using vector and matrix notations

$$\mathbf{d} = \mathbf{G}\mathbf{v} + \mathbf{n}. \quad (3)$$

Here, \mathbf{d} is a vector containing the measured visibilities of different baselines. \mathbf{v} is vector of the corresponding true sky visibility. \mathbf{G} is a diagonal gain matrix with $g_p g_q^*$ as its elements placed diagonally, and \mathbf{n} is a vector containing noise of the corresponding baseline formed from antenna p and q . For simplicity, we will omit the frequency (and time) dependency in the vector and matrix notations through the rest of the paper unless otherwise stated.

Most calibration software packages developed so far to solve for the gain parameter \mathbf{G} can be categorized into either one of the two conventional approaches, i.e. *sky-based* calibration or *redundant-baseline* calibration.² In the sub-sections that follow, we present the underlying mathematical formalism of these two calibration approaches and discuss their inherent advantages and disadvantages.

2.1 Sky-based calibration

Sky-based calibration uses a model visibility that approximates the true sky visibility, $m_{pq} \approx v_{pq}$ to solve for the gain solutions. equation (1) can be written equivalently in the matrix form,

$$\mathbf{d} = \mathbf{G}\mathbf{m} + \mathbf{n}. \quad (4)$$

The model visibility is usually constructed from known sources and the telescope beam response. Solving for the gains \mathbf{G} is then essentially equivalent to minimizing the χ^2 function of the difference between the measured visibility \mathbf{d} and the model visibility \mathbf{m} ,

$$\chi^2 = (\mathbf{d} - \mathbf{G}\mathbf{m})^\dagger \mathbf{N}^{-1} (\mathbf{d} - \mathbf{G}\mathbf{m}). \quad (5)$$

Here, \mathbf{N} is an ensemble average noise matrix in which its diagonal elements are the noise variance σ_{pq}^2 on the baseline formed from antenna p and antenna q ,

$$\begin{aligned} \mathbf{N} &= \langle \mathbf{m}\mathbf{m}^\dagger \rangle \\ &= \sigma_{pq}^2 \mathbf{I}, \end{aligned} \quad (6)$$

where the angle brackets $\langle \dots \rangle$ denote ensemble average, and in the second equality in equation (6) assumes that the noise matrix is diagonal, i.e. the noise on visibilities of different baselines is *uncorrelated*. \mathbf{I} is the identity matrix, and \dagger denotes a complex conjugate transpose. The measured visibility can be corrected (or calibrated) by using the gain solutions estimated from optimization of equation (5). However, the precision of sky-based calibration is highly dependent on the sky models, which are imperfect due to missing sources or an inaccurate beam model. This imperfection can lead to calibration errors that introduce foreground contamination in the cosmological Fourier space for 21-cm signal (Barry et al. 2016; Beardsley et al. 2016; Ewall-Wice et al. 2017; Byrne et al. 2019).

²It is worth noticing that Liu et al. (2010) investigated the Taylor expansion of the sky to deal with direction dependence or non-redundancy.

2.2 Redundant calibration

Redundant calibration relies on multiple copies of identical baselines in arrays with the regular layout to relatively solve for both the gain parameter and the true sky visibility from an over-determined system of linear equations without the need for the sky model (Wieringa 1992). The measurement equation for redundant calibration replaces the true sky visibilities v_{pq} with visibility terms that are constrained to be equal across redundant baselines u_α , where α indexes the redundant baseline sets (redundant groups),

$$d_{pq} = g_p g_q^* u_\alpha + n_{pq}. \quad (7)$$

Since the number of measured visibility d_{pq} is greater than the number of u_α that averaged over redundant group of the array and the gain parameter g , this equation becomes over-determined and can be solved, provided there are sufficient redundant baselines. The solutions for equation (7) can be then determined by minimizing the χ^2 function of the form,

$$\chi^2 = (\mathbf{d} - \mathbf{G}\mathbf{u}_\alpha)^\dagger \mathbf{N}^{-1} (\mathbf{d} - \mathbf{G}\mathbf{u}_\alpha). \quad (8)$$

In contrast to equation (5), the redundant visibility parameter \mathbf{u}_α here is not pre-defined and will be solved iteratively alongside with the antenna gain factor $\mathbf{G} \equiv g_p g_q$. To minimize the χ^2 function, a least-square estimator is usually utilized. We direct the interested reader to the existing literature (e.g. Liu et al. 2010; Zheng et al. 2014; Dillon et al. 2018, 2020) for detailed discussion of the least-square solver implementation. Minimizing equation (8) gives us a power spectrum estimation

$$\hat{\mathbf{x}} = [\mathbf{A}^\top \mathbf{N}^{-1} \mathbf{A}]^{-1} \mathbf{A}^\top \mathbf{N}^{-1} \mathbf{d}. \quad (9)$$

Here, the matrix \mathbf{A} , which depends on the array configuration, maps the measured visibility \mathbf{d} to the calibration parameters \mathbf{G} and \mathbf{v} for which we want to solve. The vector $\hat{\mathbf{x}}$ contains the least-square estimates for per antenna gain parameter and per-redundant group averaged visibility. Equation (9) is solvable iteratively, using the fitted calibration parameter as the fiducial guesses for the next cycle of the fitting, until the solutions converge. The method is implemented in the `Omnical`³ software package (Zheng et al. 2014), which was used to calibrate PAPER and the MWA Phase II observations (Parsons et al. 2014; Ali et al. 2015; Li et al. 2018; Kolopanis et al. 2019).

Due to the *degeneracies* in the measurement equation (equation (7)), the least-square estimator of equation (8) will not converge to a unique solution. Here, the term degeneracies refers to the linear combinations of gains and visibilities that redundant calibration cannot solve for, which results in the χ^2 function invariant across different choices of the degenerate parameters. In-depth discussions of degeneracy issues in redundant baseline calibration can be found in Liu et al. (2010), Zheng et al. (2014), Dillon et al. (2018), and Byrne et al. (2019).

In general, redundant baseline calibration suffers from four types of degeneracies for array elements configured on the same plane. First, if one multiplies the amplitude of the antenna gain by a constant factor A and divides the sky parameter by A^2 , the χ^2 remains unchanged so that the overall amplitude, A , of the instrument cannot be solved. Secondly, moving the phase of the gain parameter by certain amount Δ , i.e. $g_p = |g_p| e^{i\phi_p} \rightarrow |g_p| e^{i(\phi_p + \Delta)}$ does not affect the χ^2 due to the cancellation of $g_p \times g_q^*$ in equation (1). The third and fourth degeneracies are caused by phase gradients Δ_x and Δ_y along x and y directions, respectively, for a planar array. For

³<https://github.com/jeffzhen/omnical.git>

example, if one shifts $g_p = |g_p|e^{j\phi_p} \rightarrow |g_q|e^{j(\phi_q + \Delta_x x_q + \Delta_y y_q)}$, it does not affect the χ^2 if it is complemented by shift in sky parameter $V_{pq}^{\text{true}} = |V_{pq}^{\text{true}}|e^{j\phi_{pq}} \rightarrow |V_{pq}^{\text{true}}|e^{j(\phi_{pq} - \Delta_x b_x - \Delta_y b_y)}$, where x_q and y_q are the coordinates for antenna j , and $b_x = x_q - x_p$ and $b_y = y_q - y_p$ are x and y coordinates of a baseline vector \mathbf{b} connecting antenna p and q , respectively. So, the phase gradients Δ_x and Δ_y are still degenerate in the relative calibration stage. In a sense, inclining the whole array in either direction (x or y) is exactly equivalent to moving the phase centre of the sources on the sky in opposite direction. Hence, these two effects can cancel each other and result in the phase-gradient degeneracies, which in turn results in the sources appearing to be offset from the phase centre.

To fix the degeneracy issues, the redundant calibration can be divided into two calibration steps. These are relative calibration and absolute calibration. First, relative calibration is performed using the algorithm that we have just described to solve the antenna gains and the true sky. However, these solutions are degenerate, and an overall amplitude, an overall phase, and phase gradient calibration parameters remain unknown. Then, absolute calibration is performed to set these degeneracies to a reference and obtain the final calibration solutions. One particular approach for absolute calibration, as discussed in Byrne et al. (2019), is to map a pre-defined sky model of bright point sources to the degenerate parameters. However, setting degeneracies using the sky information in redundant calibration is difficult because incompleteness of the sky model can introduce frequency-dependent gain errors, which may lead to spectral structures in the otherwise smooth observations (Barry et al. 2016; Byrne et al. 2019).

Thus, the fundamental shortcomings of redundant calibration are imperfection of array redundancy and sky-modelling error. During the absolute calibration stage, the sky-modelling error enters into the redundant calibration. Redundant calibration assumes that the same sky is seen by all baselines in the same redundant group, requiring these baselines to have the same physical length and orientation. This assumption also dictates that all antenna elements in the array have the same primary beam responses. In the real situation, both are impossible to achieve for the level of accuracy required. These non-redundancies will result in calibration errors that lead to more foreground contamination in the cosmological Fourier space for 21-cm signal (Ewall-Wice et al. 2017; Orosz et al. 2018). The powerful alternative is to include the relaxed-redundancy information and position of sources into the calibration formalism (Sievers 2017).

3 CORRELATION CALIBRATION (CorrCal)

Sievers (2017) proposed a hybrid calibration scheme called `CorrCal`, which aims to further bridge the gap between sky-based and redundant calibrations by taking into account sky and array information into the calibration algorithm. The `CorrCal` framework relies on the assumption that sky information is statistically Gaussian, which is a good approximation (Sievers (2017)). This assumption allows `CorrCal` to relax the redundancy of the array and including the known sky information in its formalism through covariance-based calculation of χ^2 . The covariance matrix thus forms a *statistical model* of the sky power spectrum, weighted by the instrumental response. We describe the underlying mathematical formalism of `CorrCal` in this section.

Given a vector of the measured visibilities \mathbf{d} defined in equation (3), its covariance Σ follows,

$$\Sigma = \langle \mathbf{d}\mathbf{d}^\dagger \rangle. \quad (10)$$

Using the definition of \mathbf{d} from equation (3), Σ can be expanded as,

$$\begin{aligned} \Sigma &= \langle (\mathbf{G}\mathbf{v} + \mathbf{n})(\mathbf{G}\mathbf{v} + \mathbf{n})^\dagger \rangle \\ &= \mathbf{G}\langle \mathbf{v}\mathbf{v}^\dagger \rangle \mathbf{G}^\dagger + \mathbf{G}\langle \mathbf{n}\mathbf{n}^\dagger \rangle + \langle \mathbf{n}\mathbf{v}^\dagger \rangle \mathbf{G}^\dagger + \langle \mathbf{n}\mathbf{n}^\dagger \rangle. \end{aligned} \quad (11)$$

By assuming that the instrumental noise vector \mathbf{n} does not correlate with the true sky visibility vector \mathbf{v} , the middle two terms drop out. The last term describes the noise covariance, which is generally assumed to be Gaussian and uncorrelated between baselines; therefore, it can be written as a noise matrix \mathbf{N} as defined in equation (6). The first term describes the correlation of the sky multiplied by the gains. If we define the variance–covariance matrix of the true sky as

$$\mathbf{C} = \langle \mathbf{v}\mathbf{v}^\dagger \rangle, \quad (12)$$

equation (11) takes

$$\Sigma = \mathbf{G}\mathbf{C}\mathbf{G}^\dagger + \mathbf{N}. \quad (13)$$

Under an assumption that the sky is random Gaussian distributed, we note that by virtue of the central limit theorem, the multitude of random phenomena that produce the random character of the observed data \mathbf{d} implies that their distributions are nearly Gaussian in general. Thus, the likelihood \mathcal{L} of data \mathbf{d} , given their covariance matrix Σ , takes the standard form of a multidimensional Gaussian distribution,

$$\mathcal{L}(\mathbf{d}, \Sigma) \propto \exp \left[-\frac{1}{2} \mathbf{d}^\dagger \Sigma^{-1} \mathbf{d} \right]. \quad (14)$$

Using equation (13) in this equation, we obtain the likelihood

$$\mathcal{L}(\mathbf{d}, \mathbf{C}|\mathbf{G}) \propto \exp \left[-\frac{1}{2} \mathbf{d}^\dagger (\mathbf{G}\mathbf{C}\mathbf{G}^\dagger + \mathbf{N})^{-1} \mathbf{d} \right]. \quad (15)$$

Taking the logarithm of both sides of equation (15), we obtain the log-likelihood function $\log \mathcal{L}$,

$$\log \mathcal{L}(\mathbf{d}, \mathbf{C}|\mathbf{G}) \propto -\frac{1}{2} \mathbf{d}^\dagger (\mathbf{G}\mathbf{C}\mathbf{G}^\dagger + \mathbf{N})^{-1} \mathbf{d}. \quad (16)$$

Maximizing the likelihood function \mathcal{L} is equivalent to minimizing the (negative) log-likelihood $\log \mathcal{L}$. The term on the right-hand side of equation (16) measures the square of the distances from predicted (or true) visibility to the measured visibility in the standard deviation unit. Minimizing this term against \mathbf{G} is equal to maximizing the likelihood of true visibility being observed. The exponential term is thus simply equal to the χ^2 , given by

$$\chi^2 = \mathbf{d}^\dagger (\mathbf{G}\mathbf{C}\mathbf{G}^\dagger + \mathbf{N})^{-1} \mathbf{d}. \quad (17)$$

The covariance-based calculation of χ^2 in this equation is the underlying mathematical formalism of `CorrCal`. We further add an additional term $\chi^2 \rightarrow \chi^2 + (\sum_i \text{Im}(g_i))^2 + (\sum_i \text{Re}(g_i) - N_{\text{ant}})^2$ as a regularization factor to bound the gain solution. Here, the χ^2 function does not explicitly depend on the sky parameter or written in terms of the redundant sets of visibility, but still, the array and sky information are integrated into the covariance matrix \mathbf{C} of the true sky visibility.

Since we are modelling the data as Gaussian distributed, a form of the χ^2 function in equation (17) is different from the one defined in equation (5) and equation (8), which explicitly depends on the sky parameters \mathbf{m} and \mathbf{u}_α , respectively. However, the `CorrCal` approach of keeping the array and sky information in the \mathbf{C} matrix allows one to reproduce an exact copy of the redundant calibration algorithm in a limiting case. For instance, all baselines within the perfectly redundant group see the same sky, so the effective noise \mathbf{N}_{eff} for this group is written as the sum of per-visibility diagonal noise variance matrix \mathbf{N} and covariance between visibilities. That is,

$$\mathbf{N}_{\text{eff}} = \mathbf{N} + (\mathbf{a}\mathbf{1}) \otimes (\mathbf{a}\mathbf{1})^\dagger, \quad (18)$$

where a is a parameter controlling the variance of the sky signal, $\mathbf{1} = [1, 1, 1, \dots]$ is a vector of ones, and an operator \otimes denotes the outer product of vectors. Applying the Woodbury identity [see equation (24)] to $\mathbf{N}_{\text{eff}}^{-1}$, and taking $a \rightarrow \infty$, the χ^2 becomes the same to one for redundant calibration algorithm.

Although complete information of the true sky can never be obtained, we can incorporate partial sky information into the covariance matrix by using known sky models. Thus, we further split the covariance matrix in equation (17) into two components as per suggestion in Sievers (2017),

$$\mathbf{C} = \mathbf{S} + \mathbf{R}, \quad (19)$$

where the matrix \mathbf{S} contains the covariance of the expected visibilities of known sky sources, whereas \mathbf{R} is the covariance matrix of the visibilities within the redundant group that contains everything else that is not in \mathbf{S} , i.e. ‘the rest,’ which is mostly the diffuse sky signal from our Galaxy’s emission, weighted by the instrumental response. Note that equation (19) assumes that the sky components in \mathbf{S} and \mathbf{R} are uncorrelated since most point sources are extragalactic. With this definition, equation (17) becomes

$$\chi^2 = \mathbf{d}^\dagger (\mathbf{G}(\mathbf{S} + \mathbf{R})\mathbf{G}^\dagger + \mathbf{N})^{-1} \mathbf{d}. \quad (20)$$

Note that the explicit formalisms of \mathbf{S} and \mathbf{R} will be presented in Section 3.1 in details.

With the redundant group diffuse sky \mathbf{R} and sources \mathbf{S} expressed as vector outer products, equation (19) can be rewritten as

$$\mathbf{C} = \mathbf{s}\mathbf{s}^\dagger + \mathbf{r}\mathbf{r}^\dagger, \quad (21)$$

where \mathbf{r} and \mathbf{s} are N -dimensional vectors for the diffuse sky and source components, respectively. We substitute equation (21) for \mathbf{C} in equation (20) to obtain a χ^2 function

$$\chi^2 = \mathbf{d}^\dagger (\mathbf{N} + \mathbf{G}(\mathbf{s}\mathbf{s}^\dagger + \mathbf{r}\mathbf{r}^\dagger)\mathbf{G}^\dagger)^{-1} \mathbf{d}. \quad (22)$$

Applying the gain matrix to the source and redundant vectors in this equation, we have

$$\chi^2 = \mathbf{d}^\dagger (\mathbf{N} + \hat{\mathbf{s}}\hat{\mathbf{s}}^\dagger + \hat{\mathbf{r}}\hat{\mathbf{r}}^\dagger)^{-1} \mathbf{d}, \quad (23)$$

where $\hat{\mathbf{s}} = \mathbf{G}\mathbf{s}$, $\hat{\mathbf{r}} = \mathbf{G}\mathbf{r}$, so $\hat{\mathbf{s}}^\dagger = \mathbf{s}^\dagger\mathbf{G}^\dagger$ and $\hat{\mathbf{r}}^\dagger = \mathbf{r}^\dagger\mathbf{G}^\dagger$.

Inversion of the matrices inside the parenthesis term in equation (23) can be done with Woodbury inversion formula (Woodbury 1950), which we write here in the special case of a Hermitian matrix if both \mathbf{B}^{-1} and $\mathbf{I} + \mathbf{v}^\dagger\mathbf{B}^{-1}\mathbf{v}$ are invertible

$$(\mathbf{B} + \mathbf{v} \otimes \mathbf{v}^\dagger)^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1}\mathbf{v}(\mathbf{I} + \mathbf{v}^\dagger\mathbf{B}^{-1}\mathbf{v})^{-1}\mathbf{v}^\dagger\mathbf{B}^{-1}, \quad (24)$$

where \mathbf{B} and the identity matrix \mathbf{I} are square matrices with dimension $N \times N$, while \mathbf{v} is N -dimensional vector.

For the sake of computational tractability, the implementation of the Woodbury inversion to equation (23) in `CORRCAL` follows the following steps:

(i) First, the Woodbury inversion can be applied to Γ^{-1} , where $\Gamma = (\mathbf{N} + \hat{\mathbf{r}}\hat{\mathbf{r}}^\dagger)$ for each redundant group separately, ignoring the source vectors \mathbf{s} . To apply Woodbury identity to Γ^{-1} , let $\mathbf{B} = \mathbf{N}$ and $\mathbf{v} = \mathbf{r}$. Then, the inverted form of Γ is kept separate for each redundant group and saved in a factored form by applying the Cholesky factorization of the matrix $[\mathbf{I} + \mathbf{v}^\dagger\mathbf{B}^{-1}\mathbf{v}]^{-1}$ in equation (24) for the redundant vectors \mathbf{r} .

(ii) Finally, using the source vectors \mathbf{s} , and the already factored form of Γ^{-1} from the first step, the Woodbury inversion can be computed using $(\Gamma + \hat{\mathbf{s}}\hat{\mathbf{s}}^\dagger)^{-1}$.

3.1 Determining the sky and noise variance–covariance matrices

To perform the `CORRCAL`, we first need to estimate the matrices \mathbf{R} , \mathbf{S} , and \mathbf{N} . Thus, in this section, we will derive these variance–covariance matrices explicitly.

3.1.1 Diffuse sky component matrix (\mathbf{R})

For an instrument with a small field of view, the covariance matrices can be directly constructed from the visibility in the two-dimensional (2D) UV plane. However, most 21-cm experiments, including PAPER, use receiving elements that have wide field of views. Thus, the curvature of the sky becomes important. To account for this effect and simplify the calculation, we will first project the visibility function into the spherical harmonic space before calculating the covariance.

Given a vector $\mathbf{u} = \mathbf{b}/\lambda$ that expresses the baseline vector \mathbf{b} between antennas p and q in wavelength λ , the measured cross-correlation signal from these antennas, or the visibility, ignoring noise term, is given by,

$$v_{pq}(\mathbf{u}|f) = \int d\Omega I(\hat{\mathbf{r}}, f) B(\hat{\mathbf{r}}, f), \quad (25)$$

where

$$\begin{aligned} B(\hat{\mathbf{r}}, f) &= A_p(\hat{\mathbf{r}}, f) A_q^*(\hat{\mathbf{r}}, f) \exp[-2i\pi\mathbf{u} \cdot \hat{\mathbf{r}}] \\ &= |A_p(\hat{\mathbf{r}}, f)|^2 \exp[-2i\pi\mathbf{u} \cdot \hat{\mathbf{r}}]. \end{aligned} \quad (26)$$

In this equation, $I(\hat{\mathbf{r}}, f)$ is the flux density measured over a solid angle Ω subtended by the observed region of the sky in the direction of a unit vector $\hat{\mathbf{r}}$. $A_p(\hat{\mathbf{r}}, f)$ and $A_q(\hat{\mathbf{r}}, f)$ are the primary beam of antenna p and q , respectively. Here, the squaring of the antenna beam term $A_p(\hat{\mathbf{r}}, f)$ in the second equality of equation (26) comes from the fact that we assume that all antennas have the same primary beam response. The antenna position information is integrated into the exponential factor. Thus, the second line of the equation carries array information (position and beam response).

Expanding the terms for the sky ($I(\hat{\mathbf{r}}, f)$) and array ($B(\hat{\mathbf{r}}, f)$) over the spherical harmonics, we have

$$I(\hat{\mathbf{r}}, f) = \sum_{\ell m} a_{\ell m}(f) Y_{\ell m}(\hat{\mathbf{r}}), \quad (27)$$

$$B(\hat{\mathbf{r}}, f) = \sum_{\ell m} b_{\ell m}(f) Y_{\ell m}(\hat{\mathbf{r}}), \quad (28)$$

where $a_{\ell m}$ and $b_{\ell m}$ are the amplitudes of the flux density and the beam function for the spherical harmonics $Y_{\ell m}$, respectively. The angular number $\ell = 0, 1, 2, 3, \dots$, and the azimuthal number $m = -\ell, -\ell + 1, \dots, \ell - 1, \ell$ follow standard convention, allowing $2\ell + 1$ values of m for each value of ℓ . Using equation (27) and equation (28), the visibility in equation (25) can be expressed in spherical harmonics as,

$$\begin{aligned} v_{pq}(f) &= \int d\Omega \sum_{\ell m} a_{\ell m}^*(f) Y_{\ell m}^*(\hat{\mathbf{r}}) \sum_{\ell' m'} b_{\ell' m'}(f) Y_{\ell' m'}(\hat{\mathbf{r}}) \\ &= \sum_{\ell \ell' m m'} a_{\ell m}^*(f) b_{\ell' m'}(f) \int d\Omega Y_{\ell m}^*(\hat{\mathbf{r}}) Y_{\ell' m'}(\hat{\mathbf{r}}) \\ &= \sum_{\ell m} a_{\ell m}^*(f) b_{\ell m}(f), \end{aligned} \quad (29)$$

where we have used the orthogonality of spherical harmonics to turn the integral in the second line into a Kronecker delta function. Equation (29) is the spherical harmonics version of the visibility equation in the absence of noise.

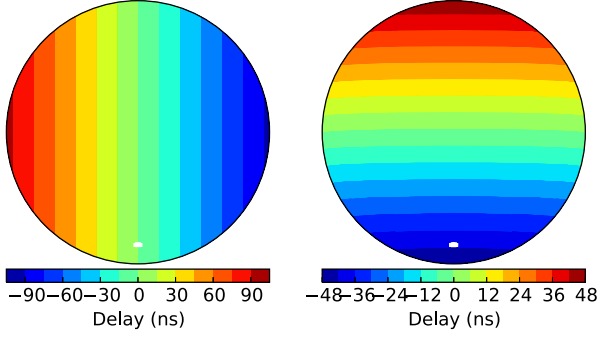


Figure 1. A delay term ($\mathbf{b} \cdot \hat{\mathbf{r}}/c$) pattern in nanoseconds (ns) over PAPER latitude for baselines pointed along the East–West (left) and the North–South (right) directions. These baselines are calculated from the actual PAPER-64 antenna coordinates that are shown in Fig. 8.

We calculated the array function that is defined in equation (26) using the exponential argument ($\mathbf{b} \cdot \hat{\mathbf{r}}/c$) in equation (25) and the PAPER-64 beam model depicted in Fig. 2. The simulated exponential argument is shown in Fig. 1 for the PAPER-64 baselines oriented along the East–West and the North–South directions. To generate the beam model of the PAPER-64 array within the frequency range of 120–168 MHz, we use the Astronomical Interferometry in Python (AIPY),⁴ a software package primarily developed for PAPER/HERA data analysis. Fig. 2 shows that the AIPY-simulated PAPER-64 beam model projected on the HEALPIX⁵ coordinates at three arbitrary right ascensions for frequency of 150 MHz. We will use this beam model for all calculations of the covariance matrices.

Using the expected visibility function defined in equation (29), in this section, we construct the diffuse sky covariance matrix \mathbf{R} . While determining this matrix, all degrees of freedom including frequencies and baselines must be taken into accounts. We estimate this matrix from the relation

$$\begin{aligned} \mathbf{R}_{(u,f):(u',f')} &= \langle v^{\text{diff}}(\mathbf{u}, f) v^{\text{diff},*}(\mathbf{u}', f') \rangle \\ &= \left\langle \sum_{\ell m} a_{\ell m}^{\text{diff},*}(f) b_{\ell m}^{\text{diff}}(\mathbf{u}, f) \right. \\ &\quad \times \left. \sum_{\ell' m'} a_{\ell' m'}^{\text{diff}}(f') b_{\ell' m'}^{\text{diff},*}(\mathbf{u}', f') \right\rangle \\ &= \sum_{\ell m} \sum_{\ell' m'} b_{\ell m}^{\text{diff}}(\mathbf{u}, f) \langle a_{\ell m}^{\text{diff},*}(f) a_{\ell' m'}^{\text{diff}}(f') \rangle \\ &\quad \times b_{\ell' m'}^{\text{diff},*}(\mathbf{u}', f'), \end{aligned} \quad (30)$$

where ‘diff’ is to denote that $v_{pq}(f)$ is coming from the diffuse sky component and the angle bracket $\langle \dots \rangle$ shows the averaging over an ensemble of realizations of the fluctuation. With the assumption that the sky is statistically isotropic Gaussian random field, the covariance for sky flux density spherical harmonics coefficients becomes

$$\begin{aligned} \langle a_{\ell m}^{\text{diff}}(f) a_{\ell' m'}^{\text{diff},*}(f') \rangle &= \langle |a_{\ell m}^{\text{diff}}(f, f')|^2 \rangle \delta_{\ell \ell'} \delta_{m m'} \\ &= C_{\ell}^{\text{diff}}(f, f') \delta_{\ell \ell'} \delta_{m m'}, \end{aligned} \quad (31)$$

where $C_{\ell}^{\text{diff}}(f, f') = \langle |a_{\ell m}^{\text{diff}}(f, f')|^2 \rangle$ is the diffuse sky angular power spectrum, which depends only on multipole number ℓ that corresponds to the angular scale $\theta = 180^\circ/\ell$, and δ is the Kronecker delta function. This is because the assumption of isotropy ensures

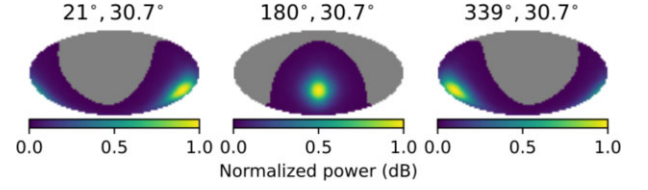


Figure 2. The AIPY-simulated antenna beam pattern of PAPER-64 array at 150 MHz at different longitudes, plotted for the PAPER latitude (30.7°S) at Karoo desert, South Africa.

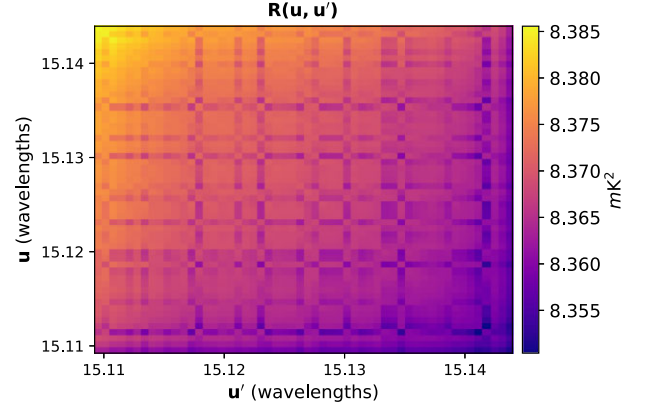


Figure 3. Covariance matrix between baselines obtained from the diffuse sky component at 150 MHz for a redundant group containing about 30-m length baselines. The axes of this figure spanning the true measured variation in the baseline lengths for the baselines in the 30-m group. The entries of the covariance are not precisely equal as displayed in the figure. The covariance between baselines drops off as baselines become less redundant.

that $\langle |a_{\ell m}^{\text{diff}}(f, f')|^2 \rangle$ is a function of only ℓ , not m . Therefore, there is no correlation in the m modes. Thus, the covariance matrix [equation (30)] can be simplified to

$$\mathbf{R} = \sum_{\ell m} b_{\ell m}^{\text{diff}}(\mathbf{u}, f) b_{\ell m}^{\text{diff},*}(\mathbf{u}', f') C_{\ell}^{\text{diff}}(f, f'). \quad (32)$$

Equation (32) is a general expression for the covariance matrix corresponding to the diffuse sky. It depends on the baseline and antenna beam model through a function defined in equation (26). The measured antenna positions (instead of assuming that they are perfectly redundant) are integrated into this equation. Hence, through this expression, the covariance-based calculation of χ^2 in equation (20) carries array information during optimization processes. Fig. 3 displays the covariance matrix between visibilities at 150 MHz in a given redundant group of baselines whose length is about 30 m.

In equation (32), however, the term $C_{\ell}^{\text{diff}}(f, f')$ is not known. The measured data covariance matrix should be constructed from the same redundant group that \mathbf{R} had formed to estimate $C_{\ell}^{\text{diff}}(f, f')$. The data covariance matrix can be determined from

$$\mathbf{D}^{\alpha} = \langle \mathbf{d}^{\alpha}(\mathbf{u}, f) \mathbf{d}^{\alpha \dagger}(\mathbf{u}', f') \rangle, \quad (33)$$

where \mathbf{D}^{α} is a covariance matrix of measured visibility data \mathbf{d}^{α} from the redundant group α . To determine the $C_{\ell}^{\text{diff}}(f, f')$, thus, we equate the covariance matrix \mathbf{D}^{α} with the diffuse sky matrix \mathbf{R} from the same redundant group α in equation (32). Note that the diffuse galactic emissions dominate visibilities from shorter baselines, while visibilities measured by longer baselines are unlikely to be dominated by diffuse emissions, so we only equate \mathbf{D} with \mathbf{R} (i.e. $\mathbf{R}^{\alpha} \approx \mathbf{D}^{\alpha}$) if both are computed from the same redundant group α .

⁴<https://github.com/HERA-Team/aipy.git>

⁵<https://github.com/healpy/healpy.git>

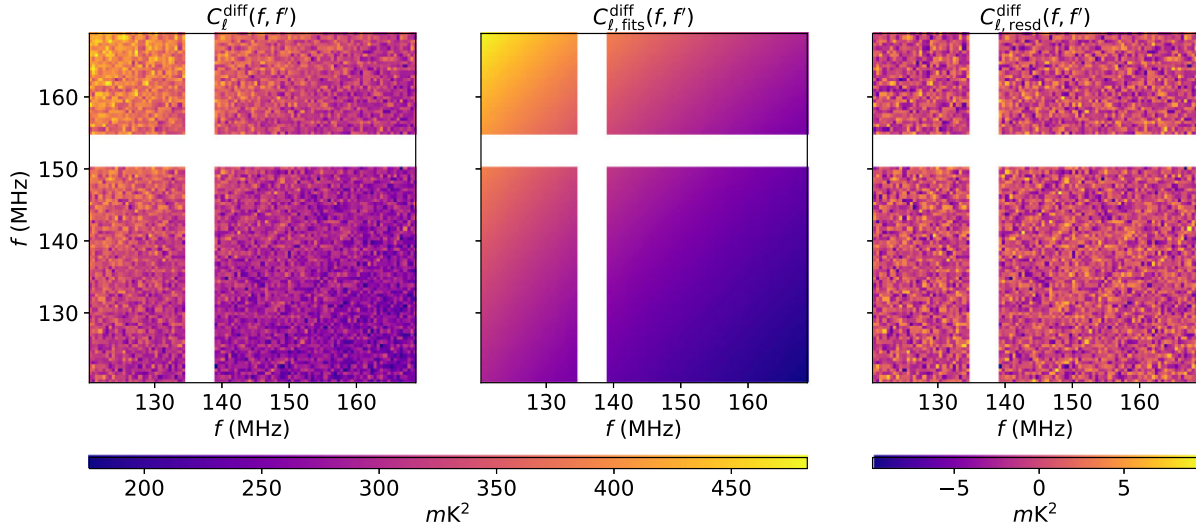


Figure 4. Left: Cross-frequency sky power spectrum $C_{\ell}^{\text{diff}}(f, f')$ obtained from equation (34). Middle: The fitted power spectrum $C_{\ell, \text{fits}}^{\text{diff}}(f, f')$. Right: The residuals $C_{\ell, \text{resd}}^{\text{diff}}(f, f') = C_{\ell}^{\text{diff}}(f, f') - C_{\ell, \text{fits}}^{\text{diff}}(f, f')$. In all plot, we assume that the ℓ -dependence of $C_{\ell}^{\text{diff}}(f, f')$ is the same. The whiteout regions are corresponding to the flagged frequency channels from the data.

This assumption may avoid mode-mixing between baselines with different lengths. It follows that,

$$C_{\ell}^{\text{diff}}(f, f') \approx \frac{D^{\alpha}}{\sum_{\ell m} b_{\ell m}^{\text{diff}}(\mathbf{u}, f) b_{\ell m}^{\text{diff}}(\mathbf{u}', f')}. \quad (34)$$

To pull $C_{\ell}^{\text{diff}}(f, f')$ out of the summation sign in equation (32), we assume that for a given redundant block α , the sky is slowly changing function of $\ell \approx 2\pi|\mathbf{u}|$. In other words, this means that the angular power spectrum of the sky does not evolve significantly over the baseline length within the redundant group.

Since sky power spectrum is real, we discard the complex part of $C_{\ell}^{\text{diff}}(f, f')$ in equation (34). Then, we implement the power-law curve fitting into equation (34) to model the angular cross-frequency power spectrum statistically. We make explicit assumptions about the functional form of the fitting function based on the fact that the power spectrum of the diffuse emission follows a simple power law with frequency (Santos et al. 2005). We implement the `SCIPY` optimization routine for curve fitting to fit data with the pre-specified function. With the fitting parameters A (amplitude in mK^2) and the spectral index α , the fitted power spectrum takes

$$C_{\text{fits}}^{\text{diff}}(f, f') = A \left(\frac{ff'}{f_0^2} \right)^{-\alpha}, \quad (35)$$

where the fitted parameters $A = 249.69 \text{ mK}^2$ and $\alpha = 1.5$, and the $f_0 = 150 \text{ MHz}$ is the reference frequency. Since the power spectrum of the diffuse emission does not evolve significantly over baselines in the quasi-redundant group, we assume that the fitted power spectrum is not a function of ℓ .

Then, the fitted angular power spectrum $C_{\text{fits}}^{\text{diff}}(f, f')$ (see Fig. 4) for that particular quasi-redundant group will be inserted in place of $C_{\ell}^{\text{diff}}(f, f')$ in equation (32) to estimate the model covariance matrix of the group. Likewise, one needs to derive the best-fitting cross-frequency power spectrum statistically from the respective quasi-redundant sets to estimate the diffuse sky covariance matrix in each redundant group. In terms of the fitted $C_{\text{fits}}^{\text{diff}}(f, f')$ parameter, the diffuse sky covariance matrix \mathbf{R} finally takes

$$\mathbf{R} = C_{\text{fits}}^{\text{diff}}(f, f') \sum_{\ell} b_{\ell}^{\text{diff}}(\mathbf{u}, f) b_{\ell}^{\text{diff},*}(\mathbf{u}', f'). \quad (36)$$

We use equation (36) to represent the diffuse sky component for χ^2 optimization process in equation (17). The dimension of the covariance matrix \mathbf{R} depends on the number of baseline $N_{\mathbf{u}}$ and number of frequency channels N_f , i.e. its dimension is equal to $N_{\mathbf{u}}N_f \times N_{\mathbf{u}}N_f$.

3.1.2 Point source component matrix (\mathbf{S})

The partial information about the point sources on the sky is one of the components the `CORRCAL` calibration scheme employs to estimate the best likelihood antenna gain parameters. Simulation in Sievers (2017) showed that the inclusion of the bright sources information in `CORRCAL` has a significant effect in reducing both amplitude and phase calibration error. However, the phase error has reduced remarkably (about a factor of 5 better than non-source-informed correlation calibration) when the correlation information of point sources has implemented in Sievers (2017), which substantially improves the quality of phase calibration.

In this study, we use the published astronomical catalogues in Jacobs et al. (2013) to model the known sources. Using sources information in this catalogue, we form the matrix \mathbf{S} that carries the sources information.

The source position is implicit in time and usually expressed in terms of right ascension (RA) and declination (Dec) of catalogue source position. The time dependence of sources position vector $\hat{\mathbf{r}}_i$ thus comes from source motion relative to the observer meridian, which best visualized using the Hour-Angle (HA) and the observer local sidereal time (LST). Because HA defines the amount of time since the source transited the observer meridian, hence, it tells how distant a source is from the meridian. Based on the HA values ($\text{HA} = \text{LST} - \text{RA}$) of each source from the catalogue, we use sources that are above the horizon during observations time to create \mathbf{S} . We then project each source position on to the sphere. Then, we predict the visibilities in the spherical sky from a set of point sources with known positions. Using the source statistics in this catalogue, we calculate the predicted visibility with the help of equation (26) for known sources. That is, suppose the point source profile is a Dirac-delta function on the sky $\hat{\mathbf{r}}_i$ ($i = 1, 2, \dots, N_{\text{ps}}$, N_{ps} is the total number

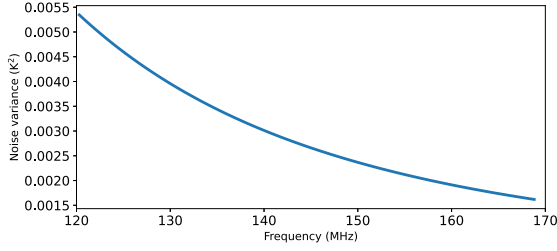


Figure 5. Noise variance estimated from the system temperature of PAPER.

of point sources), then the specific intensity [equation (25)] can be written as

$$I(\hat{\mathbf{r}}, f) = \sum_{i=1}^{N_{\text{ps}}} I(\hat{\mathbf{r}}_i, f) \delta_{2\text{D}}(\hat{\mathbf{r}}, \hat{\mathbf{r}}_i), \quad (37)$$

where $\delta_{2\text{D}}(\hat{\mathbf{r}}, \hat{\mathbf{r}}_i)$ is a 2D Dirac-delta function. Substituting this function into equation (26), we obtain the visibilities vector for point sources

$$\mathbf{v}^{\text{ps}}(\mathbf{u}, f) = \sum_{i=1}^{N_{\text{ps}}} I(\hat{\mathbf{r}}_i, f) |A_{pq}(\hat{\mathbf{r}}_i, f)|^2 e^{-i2\pi\mathbf{u}\cdot\hat{\mathbf{r}}_i}. \quad (38)$$

To simulate beam function for sources, we use the known position on the sky from the established catalogue in Jacobs et al. (2013) and the PAPER-64-simulated beam model that is displayed in Fig. 2. Taking the outer product between the visibility vector $\mathbf{v}^{\text{ps}}(\mathbf{u}, f)$ and its complex conjugate transpose $\mathbf{v}^{\text{ps},\dagger}(\mathbf{u}, f)$, we form the covariance information for the discrete sources,

$$\mathbf{S} = \mathbf{v}^{\text{ps}}(\mathbf{u}, f) \mathbf{v}^{\text{ps},\dagger}(\mathbf{u}, f). \quad (39)$$

Thus, we will use this source information covariance matrix to calculate the χ^2 function in equation (20).

3.1.3 Noise matrix (N)

In this work, we estimate the per-diagonal noise variance matrix \mathbf{N} from the *radiometer equation*

$$\sigma_{\text{rms}} = \frac{T_{\text{sys}}}{\sqrt{2\Delta f \tau}}, \quad (40)$$

where $\Delta f = 97$ kHz is the frequency bandwidth in kHz, $\tau = 49.2$ s is the integration time. The system temperature T_{sys} from the relation defined in equation (23) of Cheng et al. (2018) is,

$$T_{\text{sys}} = 180 \text{ K} \left(\frac{f}{0.18 \text{ GHz}} \right)^{-2.55} + T_{\text{revr}}, \quad (41)$$

where f is frequency, and T_{revr} is a receiver temperature which equals to 144 K. The noise variance is shown in Fig. 5.

3.2 Eigendecomposition of sky signal covariance matrix

The amplitude of \mathbf{R} in equation (36) for a perfectly redundant group containing two baselines is equal to a parameter controlling the variance of the sky, C_ℓ , times a square matrix of ones. That is,

$$\mathbf{R} = C_\ell^2 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad (42)$$

However, for a quasi-redundant group, all the off-diagonal elements of the correlation matrix \mathbf{R} are slightly different from one. As shown in Fig. 3, all the elements of the covariance matrix are not

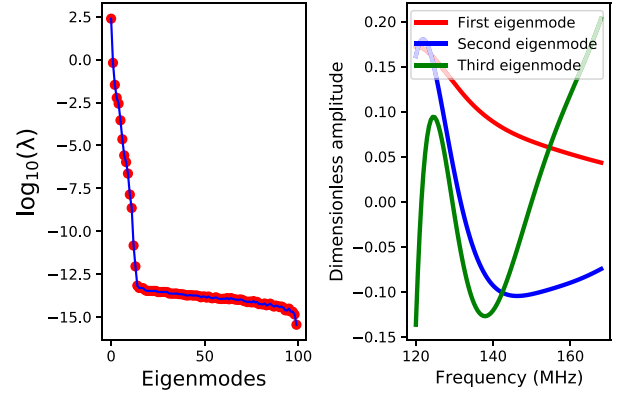


Figure 6. Left: Eigenvalues of cross-frequency covariance matrix described in equation (36). The vertical axis represents the \log_{10} eigenvalues, while the horizontal axis labels the corresponding principal component index. Right: The first three eigenmodes for \mathbf{R} corresponding to the first three eigenvalues for the frequency range from 120 to 168 MHz.

precisely equal. It shows that the correlation between baselines decreases as they become less redundant. In that case, we approximate \mathbf{R} from its eigendecomposition in our analysis following Sievers (2017). If each redundant group in the array is not represented by few eigenmodes that are much less than the visibilities in the group, it shows that there is no sufficient redundancy in the array and therefore any calibration scheme that relies on the array redundancy may not give the accurate results. In the eigendecomposition analysis, few modes of the principal components accounted for most of the variances in the original data. Thus, those eigenmodes of \mathbf{R} corresponding to sufficiently large eigenvalues could replace its original data with very little loss of information. In the left-hand panel of Fig. 6, we show the eigenvalues of cross-frequency covariance matrix described in equation (36). The vertical axis represents the \log_{10} eigenvalues, while the horizontal axis labels the corresponding principal component numbers. The sharp falloff in the eigenvalues against the eigenmode index suggests that by measuring a few modes of an eigenvalue, one can account for most of the variation in the original data that forms \mathbf{R} . In addition, all the eigenvalues are greater than zero because \mathbf{R} is a positive-definite covariance matrix. In the right-hand panel of Fig. 6, we depict the first three eigenmodes for \mathbf{R} corresponding to the first three eigenvalues for a frequency range from 120 to 168 MHz. Specially, the first and the second eigenmodes are more important to explain the diffuse sky covariance between frequency channel because they are a relatively smooth function of frequency.

We keep eigenmodes with amplitude more than 10^{-5} times the largest. This allows the keeping of $\sim 2-3$ complex eigenmodes per quasi-redundant block. These few modes could represent most of the variations in a given redundant group α . For instance, the proportion of variance related to the eigenvalues λ_1 and λ_2 can be calculated from

$$\left(\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \right) \times 100, \quad (43)$$

where p is the total eigenmodes. If the value of equation (43) is greater than the threshold percentage 99 per cent, we use eigenmodes corresponding to eigenvalues λ_1 and λ_2 to create a matrix that carries information about weights contributed by the variables in the original data to these eigenmodes. Variables with the highest correlation (in absolute value) with a principal component receive the highest weight in that component. This vector can be calculated from the statistical

relation

$$\mathbf{p} = \mathbf{e}_i \sqrt{\lambda_i}, \quad (44)$$

where $(\mathbf{e}_i, \lambda_i)$ is the eigenvector–eigenvalue pairs for \mathbf{R} , and the subscript i is the principal component index. Using the vector outer product of \mathbf{p} thus we estimate a matrix that represents the diffuse sky covariance, given by

$$\mathbf{R}_{\text{PCA}} = \mathbf{p} \otimes \mathbf{p}^\dagger, \quad (45)$$

where \mathbf{R}_{PCA} is the diffuse sky covariance matrix calculated from the principal component analysis of \mathbf{R} for a given redundant block α . The dimension of \mathbf{R}_{PCA} would significantly be reduced if very few eigenmodes could represent most of the variation in the original data. For instance, if we keep only two eigenmodes of \mathbf{R} for a given redundant group containing four baselines, a complex vector \mathbf{p} could be

$$\mathbf{p} = [p_1, p_2, 0, 0], \quad (46)$$

where p_1 and p_2 are complex elements of a vector \mathbf{p} that retained from eigenmode analysis. The outer product of this vector by its complex conjugate transpose takes

$$\mathbf{p} \otimes \mathbf{p}^\dagger = \begin{bmatrix} p_1 p_1^* & p_1 p_2^* & 0 & 0 \\ p_2 p_1^* & p_2 p_2^* & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (47)$$

Here, the operator $*$ denotes a complex conjugate. Adding the noise to the above matrix, we have

$$\mathbf{\Gamma} = \mathbf{N} + \mathbf{p} \otimes \mathbf{p}^\dagger. \quad (48)$$

All off-diagonal elements of the noise variance matrix \mathbf{N} , which has the same size as $\mathbf{p} \otimes \mathbf{p}^\dagger$, are zero. Thus, $\mathbf{\Gamma}$ has only a relatively small number of non-zero elements and it can be considered to be sparse because most of its elements are zero. Therefore, it is wasteful to use the general dense matrix algebraic methods to inverting $\mathbf{\Gamma}$ since most of the $O(N^3)$ operations devoted to inverting the matrix involve zero operands. Hence, in terms of memory storage and computational efficiency, performing the sparse matrices operation traversing only non-zero elements has significant advantages over their dense matrix counterparts. Taking advantage of the sparse structure of the correlation matrix $\mathbf{\Gamma}$, `CorrCal` utilizes the sparse matrix operations to increase its computational performance in a large data set.

To implement the sparse matrix computations in `CorrCal`, per-redundant group correlation matrices $\mathbf{\Gamma}$ are stored as sub-matrices on the diagonal of the square block-diagonal matrix. These sub-matrices will be square, and their dimensions vary depending on the number of baselines in the redundant group. Since the off-diagonal elements of the block-diagonal matrix are zero, and there can be many sparse sub-matrices stored as its diagonal elements, the overall matrix must be sparse. It can be easily compressed and thus requires significantly less storage, which also enhances the computational speed.

3.3 Implementation of `CorrCal`

The crucial steps to perform `CorrCal` are grouping data according to their corresponding redundant baseline and determining the matrices \mathbf{R} , \mathbf{S} , and \mathbf{N} to each of these redundant groups. Then, one needs to set up the sparse matrix representations for these matrices and determine the χ^2 function and its gradient information using the `CorrCal` routines. Utilizing these ingredients as inputs, `CorrCal`

finally adopts the `Scipy`⁶ built-in optimization algorithm to fit antenna gains. We summarize the procedures of implementation of `CorrCal` as follows:

(i) First, we need to create the sparse matrices representing \mathbf{R} , \mathbf{S} , and \mathbf{N} set up to each redundant group using the `CorrCal` routine, `sparse_2level()`. The following arguments must be supplied to this routine to create sparse matrices:

- (a) The real-imaginary separated per-redundant group diffuses sky vector \mathbf{p} . In this vector, we make all its entries equal to zero except those entries corresponding to eigenmodes that we kept per redundant block [see equation (46)].
- (b) The real-imaginary separated vector \mathbf{s} that carries information about the known point sources.
- (c) Per-diagonal noise variance matrix \mathbf{N} .
- (d) A vector that contains indices setting off the redundant group.

(ii) Secondly, the χ^2 function and its gradient information should be determined using the `CorrCal` subroutines. The arguments passed to these sub-routines in order to calculate the χ^2 function and the gradient information of χ^2 are:

- (a) A vector containing the real-imaginary separated initial guess for antenna gains.
- (b) Per-redundant group measured visibility data vector. The real-imaginary parts should be separated such that imaginaries follow their respective reals [e.g. see vectorized visibility in equation (49)].
- (c) The sparse matrix representations for \mathbf{R} , \mathbf{S} , and \mathbf{N} from the first step.
- (d) A vector containing per-visibility antenna indices.

(iii) Finally, using the χ^2 , its gradient information, and a vector of the initial guess for antenna gains that scaled by a large number as an input, `CorrCal` adopts the `Scipy` built-in non-linear conjugate gradient algorithm, `Scipy.optimize.fmin_cg()`, to solve for antenna gains.

3.3.1 Antenna gain solutions as a function of frequency–bandpass solution

For the EoR studies, highly precise frequency-dependent bandpass calibration plays a crucial role. If one’s calibration solution has a noise-like structure along the frequency axis, it may introduce fine-frequency structure to the otherwise smooth sky observations even if the instrument’s response is the slow function of observing frequencies. It poses the serious challenges towards efforts to isolate spectrally structured weak 21-cm signal from spectrally smooth strong foreground signal. Several strategies have proposed for ensuring smooth calibration solutions as a function of frequency. For instance, simulation in Ewall-Wice et al. (2017) and Orosz et al. (2018) has shown that downweighting measurements from longer baselines in a redundant baseline calibration solution would eliminate the migration of fine spectral structure from longer baselines to the shorter baselines regimes. This effect is because the longer baselines have more chromatic nature than the shorter ones as shown in Morales et al. (2012). Other strategies fit the calibration solutions with smooth function in frequency for a bandpass calibration.

⁶<https://github.com/scipy/scipy.git>

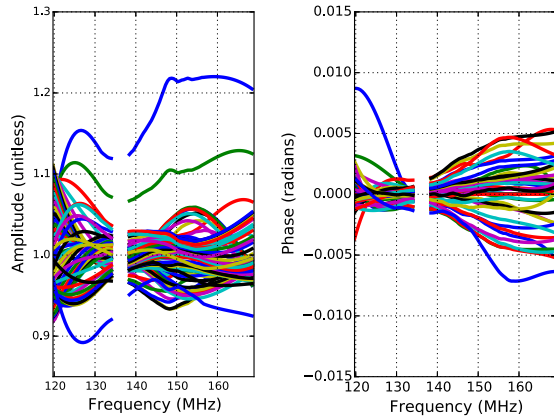


Figure 7. The bandpass gain solution derived from a snapshot of 10-min observation (JD2456242.25733) of PAPER-64 elements. Amplitude (left) and phase (right) of the gain solutions against frequency. Each line is a different antenna. Discontinuity in the gain solution shows the flagged frequency channel affected by RFI contamination.

However, the natural formalism of `CorrCal` would allow one to fit for gain solutions as a function of all frequency channels at once. It means that instead of solving an independent set of equations for every single frequency, visibility data \mathbf{d} from different frequency channels put together in a longer vector as,

$$\mathbf{d} = [\operatorname{Re}\{v(f_1)\}, \operatorname{Im}\{v(f_1)\}, \operatorname{Re}\{v(f_2)\}, \operatorname{Im}\{v(f_2)\}, \dots], \quad (49)$$

where $\operatorname{Re}\{v(f_1)\}$, $\operatorname{Im}\{v(f_1)\}$, $\operatorname{Re}\{v(f_2)\}$, and $\operatorname{Im}\{v(f_2)\}$ are the real (Re) and imaginary (Im) parts of the visibility measurements corresponding to frequency channel f_1 and f_2 , respectively. Concatenating visibility from all frequency channels in a longer vector like equation (49) enforces `CorrCal` to execute covariance information between frequencies. This imposes one to solve for smooth bandpass gain solutions for all frequencies as a single calibration step as depicted in Fig. 7. To some extent, the frequency structure of the gain solutions in Fig. 7 is comparable to the autocorrelation result in Li et al. (2019) and Byrne et al. (2019). Once these bandpass calibration solutions are found, we divide these solutions to the raw data for `CorrCal` calibration.

The correlation calibration suffers from the same degeneracy problem as with redundant baseline calibration presented in Section 2.2. Therefore, the gain solutions in `CorrCal` are degenerate, and these degeneracy issues will have to be dealt with. In this work, the overall amplitude degeneracy in `CorrCal` gain solutions is constrained by setting the average absolute value of the gains equal to 1. Because we are applying `CorrCal` to data that have already been absolutely calibrated, we do not want to rescale the overall amplitude. To set the overall phase degenerate parameter Δ , we define the reference antenna. We then set the overall phase by bringing the argument of the gain of the reference antenna equal to zero. The phase gradient degeneracy issues are resolved using the known point sources information on the sky. Then, we solve for phase gradient parameters by fitting them with the phase solutions of gain that have been obtained using sources information. That means,

$$\chi_{\text{phase}}^2 = \sum_{p=1}^N (\operatorname{Arg}[\hat{g}_p(f)] - \Delta_x b_{x,p} - \Delta_y b_{y,p}), \quad (50)$$

where $b_{x,p}$ and $b_{y,p}$ are the x and y components of the baseline b for antenna p . Then, we subtract the fitted gradient parameters off

from the phase calibration solutions across the array to set for phase gradient issues.

3.4 Comparisons of `CorrCal` to other methods

The `CorrCal` approach has a distinct feature to alleviate issues like imperfections of array redundancy and miss-modelled sky catalogue that *redundant baseline* and *sky-based* calibration techniques suffer from. In `CorrCal`, sources with known positions can be included as a prior even if their fluxes are not precisely known, because the position of the sources is better known than their intensities during observation time and frequency. This concept allows `CorrCal` to assume that the source catalogue is complete and well modelled with precisely known source positions. The method works well if the sky-modelling error could be mostly caused by uncertain/missed source fluxes in the *sky-based* calibration approach.

The covariance-based formulation of χ^2 in `CorrCal` provides significant flexibility by relaxing the assumption of explicit redundancy to accounts for imperfections of array redundancy. For instance, if baselines in a quasi-redundant group are assumed to be alike but they are not perfectly alike, then the `CorrCal` method incorporates this offset from perfection in its formalism by suppressing the off-diagonal elements of covariance matrix within the group. However, the redundant baseline method assumes that the baselines are perfectly redundant in the group, and therefore all the off-diagonal elements of the covariance matrix in the group are exactly equal [e.g. equation (42)].

4 THE PAPER-64 DATA SET

PAPER is a low-frequency radio interferometer experiment dedicated to probing the EoR through the measurements of the 21-cm power spectrum (Parsons et al. 2010). It was the first 21-cm array experiment with a redundant array configuration, where antennas were arranged in a regular grid to generate multiple copies of the same baselines. The redundant baselines probe the same modes on the sky, resulting in significant improvement in the sensitivity of the 21-cm power spectrum; thus, it is reasonable to use redundant calibration. The array was operating at the South African SKA site in the Karoo desert until its decommissioning in 2015. It was first deployed with 16 antennas and continued to grow, eventually increasing to 128 antennas in 2015. PAPER has 1024 frequency channels across the 100-MHz band. It observes between 100 and 200 MHz, corresponding to the 21-cm signal from redshifts 6–13, with a frequency resolution of ~ 97 kHz. Data from the 32-elements PAPER array have been analysed in Parsons et al. (2014), which results in one of the very first upper limits on the 21-cm power spectrum at $z = 7.7$. Analyses of the data set from the 64-elements PAPER array (hereafter PAPER-64) in Cheng et al. (2018) and Kolopanis et al. (2019) have placed significant upper limits on the power spectrum amplitude of 21-cm signal.

We will perform a re-calibration of the PAPER-64 data set taken from Ali et al. (2015) with `CorrCal`. This data set has been calibrated with redundant calibration using the `OmniCal` software and then absolute calibrated. The absolute calibration has been made by fitting to the bright sources such as Fornax A, Pictor A, and the Crab Nebula to set the overall phase. Using the Pictor A calibrator, the overall amplitude in these data has been fixed. The detailed discussions about the absolute calibration stage for these data are presented in Ali et al. (2015). The data set has dual polarization products. We use one night of observations taken between 2012

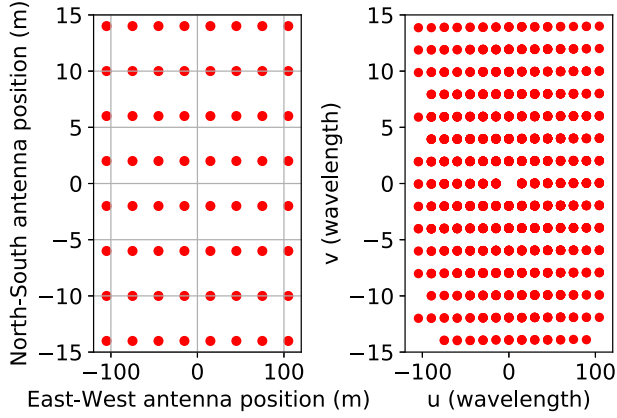


Figure 8. Antenna position (left) and UV coordinate of the PAPER-64 array. The North–South and East–West antenna coordinates are in meter, and the UV coordinates are in wavelengths at 150 MHz (right).

Table 1. Summary of observational parameters and their values used in this work.

Parameter	Value
PAPER array location	30.7°S, 21.4° E
Observation dates	2012 November 10 to 2012 November 11
Observing mode	Drift-scan
Time resolution	~42.9 s
Frequency range	120 MHz–168 MHz
Frequency resolution	0.49 MHz
Number of antenna	61
Field of view	60°
Visibility polarization	XX, YY
Shortest baseline	4 m
Longest baseline	212 m

November 10 (JD2456242.17382) and 2012 November 11 (JD 2456242.65402). As depicted in Fig. 8, the PAPER-64 array system comprises 64 dipole antennas arranged on a regular grid spacing with the total number of baselines equal to $64 \times (64 - 1)/2 = 2016$ and thus measure 2016 visibilities. For our calibration, we use all baselines except those associated with three antennas that were flagged during the observation due to known spectral instability. The observation parameters of this data set are summarized in Table 1. We direct readers to Ali et al. (2015) for more details of observation strategies, flagging of radio frequency interference (RFI), channelization of data, antenna cross-talk minimization, and other technical information of the PAPER array. Our method proposes a new possibility to fully gauge the systematics and noise of radio interferometry.

5 RESULTS

5.1 Post-CorrCal spectral structure of data

After CorrCal implementation, we have tried to compare the structure of data as a function of frequency. Fig. 9 depicts the real part of visibility against frequency for a single integration time of a given baseline. The careful examination of the plot shows that the real visibility gets spectrally smoother (red curve) after CorrCal re-calibration. It signifies the reduction of the spectral structure of the visibility measurements after re-calibration. Considering that,

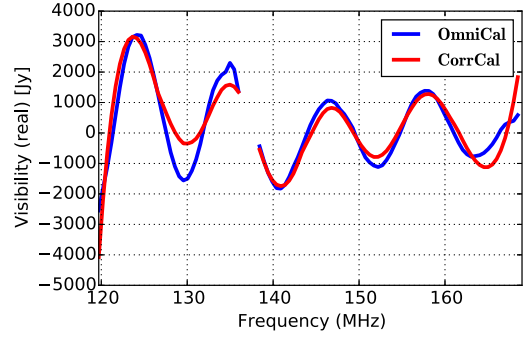


Figure 9. The real part of visibility data (JD2456242.25733) as a function of frequency at a single timestamp for an EW baseline of length about 30 m for both calibrations approaches. Line discontinuity corresponds to the flagged frequency channels. The spectral structure looks slightly reduced after CorrCal (red curve).

in the following section, we present the effect of re-calibration on the delay transformed power spectra (Parsons & Backer 2009).

5.2 Delay-space power spectra

The statistical analysis of the Fourier mode of the power spectrum (k_{\perp} , k_{\parallel}) for low-frequency observations is a powerful tool to probe the EoR (Morales & Hewitt 2004). This method shows the existence of distinctive regions dominated by 21-cm signal and foreground contaminants in a 2D line of sight (k_{\parallel}) and transverse co-moving ($k_{\perp} = \sqrt{k_x^2 + k_y^2}$) Fourier plane. The spectrally smooth foreground is expected to contaminate the region of low k_{\parallel} values, while the vast majority of k_{\parallel} values are free from these contaminants (Liu et al. 2014a). Nevertheless, several studies have shown that foreground contamination may go further to the higher k_{\parallel} values because of chromatic interaction of interferometry with intrinsically smooth foreground (Morales et al. 2012; Parsons et al. 2012; Vedantham, Shankar & Subrahmanyam 2012; Thyagarajan et al. 2013; Dillon et al. 2014; Liu et al. 2014a; Liu, Parsons & Trott 2014b). This effect leads to the distinctive wedge-like structure in k_{\perp} and k_{\parallel} Fourier space, leaving the foreground-free EoR window beyond the wedge (see results in Figs. 10 and 11 and discussions there).

To convert the measured visibility to 2D power spectrum in k_{\parallel} and k_{\perp} space, the per-baseline delay transform method is used. The technique was first introduced by Parsons & Backer (2009). As thoroughly presented in Parsons et al. (2012), the delay transformed visibility is obtained by inverse Fourier transforming of each baseline along the frequency axis. The method allows the smooth foreground to localize at the central region between the negative and positive geometric delay limits, and any spectrally unsmooth structure such as 21-cm signal may spread across delay axis (Parsons et al. 2012). We will review the mathematical formalism of the per baseline delay-transform method in Appendix A. In our analysis, we use Stokes I visibilities⁷ taken by PAPER-64 array from Julian date 2456242.17382 to 2456242.65402 (2012 November 10–11) to form the 2D power spectrum using delay transform method, a total of about 9-h observation over frequencies 140–160MHz.

⁷Note that these are not true Stokes I would rather the average of XX and YY visibilities (see Section A).

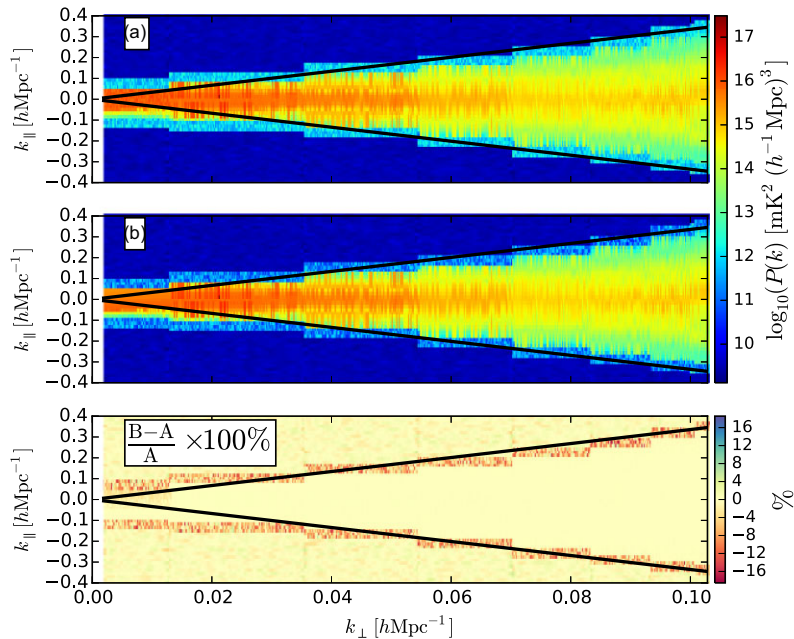


Figure 10. Delay power spectra on either side of k_{\parallel} direction. Foreground wedges are confined under the region bounded by the black line or the horizon line. (A) The delay power spectra for data already calibrated with OmniCal. (B) Power spectra after CorrCal re-calibration. The lower panel displays linear scale percentage deviation. The negative percentage deviation in power at the bin right on the horizon shows that the wedge has become tighter after CorrCal re-calibration.

Following Parsons et al. (2012) and Pober et al. (2013), we apply the Blackman–Harris window function during the delay-transform stage to minimize the foreground leakage caused by the finite bandwidth of frequency. This technique allows the EoR window to be mostly dominated by 21-cm brightness temperature by reducing the foreground spectral leakage to the smallest possible amount (Parsons et al. 2012; Pober et al. 2013). We employ the one-dimensional (1D) CLEAN algorithm on the delay-transformed data to reduce the effect introduced by RFI flagging and to increase the resolution in delay space (Pober et al. 2013). We then cross-multiply the consecutive integration time to avoid the noise bias. After these stages, we form the cylindrical delay power spectrum in k space using equation (A5).

Fig. 10 shows the delay power spectra where the positive and negative values of k_{\parallel} have not been averaged together. This form of power spectra is expected because emissions from the negative and positive delay directions are proportional to the negative and positive k_{\parallel} , respectively. The delay power spectra from emissions near-zero delays (where $k_{\parallel} \sim 0$) are confined within the central region of the 2D delay power spectra. It corresponds to foreground emissions from the peak primary beam response of the instrument.

However, from Fig. 10, we see that the strength of power spectra around zero delay decreases as one moves from the lower k_{\perp} to higher k_{\perp} regions. Furthermore, for both calibration approaches, more emission goes beyond the horizon limit for shortest baselines (smallest k_{\perp} values). This effect is because shorter baselines resolve out less of the Galactic synchrotron so that emission will be brighter and sidelobes extend further in k_{\parallel} (Pober et al. 2013).

The black diagonal line represents the horizon limit determined by baseline length and the source location relative to the main field of view of the array, marking the boundaries of the foreground

wedge. Its analytic expression has been defined in equation (A7). The spectral-smooth emissions are supposed to be confined within that limit, and any emissions that are intrinsically unsmooth with frequency are moving beyond the horizon limit. The calibration error plays a non-negligible role in this regard in imparting the spectral structure to emissions that were originally spectrally smooth as discussed earlier. This unnatural spectral structure, imparted due to calibration error or other effects, will affect the EoR window by scattering power beyond the wedge limits. To see CorrCal’s impact on the delay power spectra, we compare foreground power in a wedge region of the power spectrum, and outside of it in which the 21-cm signal is dominating by taking the percentage deviation between power after CorrCal and OmniCal, shown in the bottom panel of Fig. 10. It is apparent that the power spectra inside and outside the wedge region are comparable for both calibration approaches. However, we observe a consistent drop in power at the bin right on the horizon after CorrCal. Most importantly, this is consistent with some of the analyses done with MWA in (Li et al. 2018; Li et al. 2019; Zhang et al. 2020) that demonstrate redundant calibration’s ability to make small improvements in the modes nearest (but not inside) the wedge. This phenomenon suggests a reduction in the spectral structure of the data after re-calibration with CorrCal.

The well-known wedge-shaped foreground power spectra from our analysis are shown in Fig. 11. We generate this form of power spectra by averaging power from both positive and negative k_{\parallel} directions and folding it over the centre of the zero-delay line (where $k_{\parallel} = 0$) of Fig. 10. A similar feature of power spectra has been observed from PAPER-64 data in Pober et al. (2013), as displayed in Fig. 3 of their work. Moreover, Kohn et al. (2016) have reported a wedge-like structure from the analysis of PAPER-32 data. In our analysis, Figs 11(a) and (b), respectively, display the wedge-shaped power spectra to data already calibrated using OmniCal

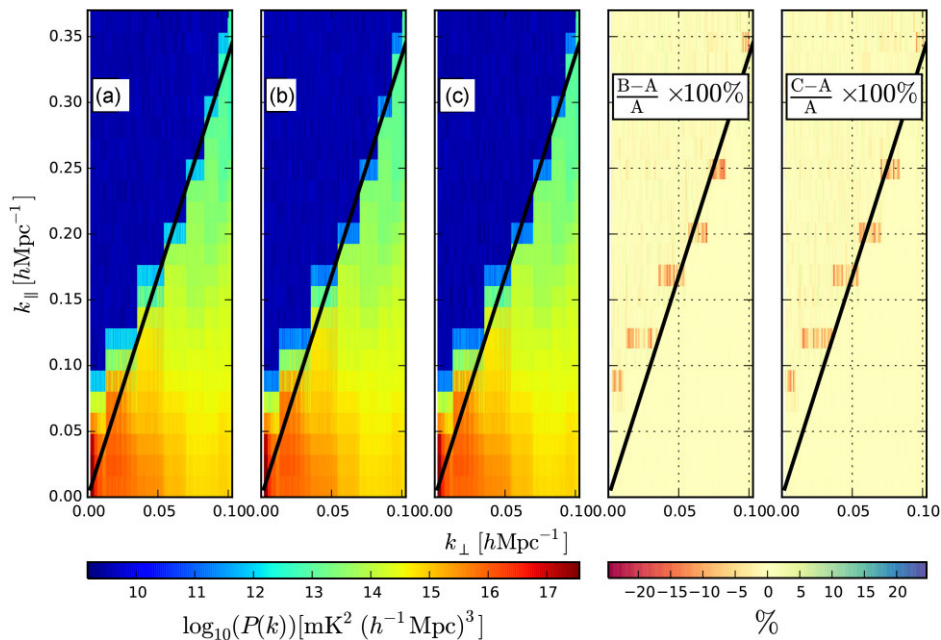


Figure 11. Same as Fig. 10 but averaged and folded over the delay centre to obtain the ‘wedge’-shaped delay power spectra. From left to right: (A) The power spectra generated from data already calibrated with OmniCal package; (B) the power spectra obtained after data have re-calibrated using without source information in CorrCal; (C) power spectra after re-calibration using source information in CorrCal. The right two panels show the linear scale percentage deviation without and with sources information in CorrCal formalism, respectively. The black diagonal line marks the horizon limit (τ_H).

and to data after CorrCal re-calibration. The middle panel in Fig. 11(c) shows power spectra after data have been calibrated using point sources information in CorrCal. For each case, the wedge extends to higher values of k_{\parallel} at a longer baseline (k_{\perp}) regime, demonstrating how the theoretical wedge limit increases as a function of baseline length. In addition, sources that are located far away from the pointing centre of the primary beam will also appear at higher delays, corresponding to higher k_{\parallel} . The relative deviation of power in percent between CorrCal and OmniCal calibration approaches is shown in the third (CorrCal without sky sources information) and fourth (CorrCal with sky sources information) panels of Fig. 11 from left to right, respectively. The negative percentage deviation at the bin right on the horizon in $[(B - A)/A] \times 100$ per cent and $[(C - A)/A] \times 100$ per cent of Fig. 11 indicates that the power spectra around the wedge limit have become slightly tighter after CorrCal re-calibration. This effect is not considerably saturated if we include the sources’ information in CorrCal formalism as shown in the right-hand panel of the figure.

The R.H.S panel of Fig. 12 depicts a portion of 2D power spectra for a given group of baselines whose lengths are about 150 m. We average over k_{\perp} axis over the range of $0.0733 h \text{ Mpc}^{-1} < k_{\perp} < 0.0736 h \text{ Mpc}^{-1}$ to generate the corresponding 1D power spectra shown on the L.H.S of Fig. 12. From this figure, we observe that there is a steep falloff of the power spectrum right away the horizon limit as expected, showing that the foreground power confined within the slice of the wedge is brighter than the 21-cm signal by 4–5 orders of magnitude. This sharp falloff follows the same trend for both calibration schemes. However, the inclusion of sky sources statistics as a prior during CorrCal re-calibration step does not show the significant change in the power spectra, as shown on the L.H.S of Fig. 12.

6 CONCLUSIONS

In this work, we used a new calibration scheme (CorrCal) to re-calibrate the PAPER64 observations. Our new calibration scheme boils down to the covariance-based calculation of χ^2 , allowing one to include the relaxed array redundancy and sky information into its formalism. This formalism also provides space for cross-frequency bandpass calibration. Furthermore, the inclusion of known sources information in CorrCal breaks the statistical isotropy of the sky, which sets the overall phase gradient degenerate parameter across the array. We have compared the spectral smoothness of the observed data before and after CorrCal’s application. We observed that the spectral structure of the visibility data after CorrCal becomes slightly reduced, for instance, as shown in Fig. 9. The significant factor for this improvement could be the CorrCal natural formalism to carry out the cross-frequency bandpass calibration.

We implemented the delay-transform method (Parsons & Backer 2009; Parsons et al. 2012) to filter out smooth foregrounds from the spectrally structured weak 21-cm signal in 2D Fourier k -space. The comparison results in the delay space show that the power spectra around the foreground wedge limit have been reduced by about 6 per cent after the CorrCal re-calibration. Nevertheless, this effect does not significantly improve the power spectrum in any bin around the horizon limit of the foreground wedge and further investigation will be needed to test the algorithm effectiveness and its implementation methods. The current numerical tests lay the foundation of calibration radio instruments using correlated signals on the sky. Our future work will include possible beam shape errors, antenna pointing errors, and position errors, and improve the CorrCal method by using sky and antenna simulation (e.g. pyuvsim) and compare it with sky-based calibration and redundant baseline calibration. In conclusion, the CorrCal may be an alternative calibration method

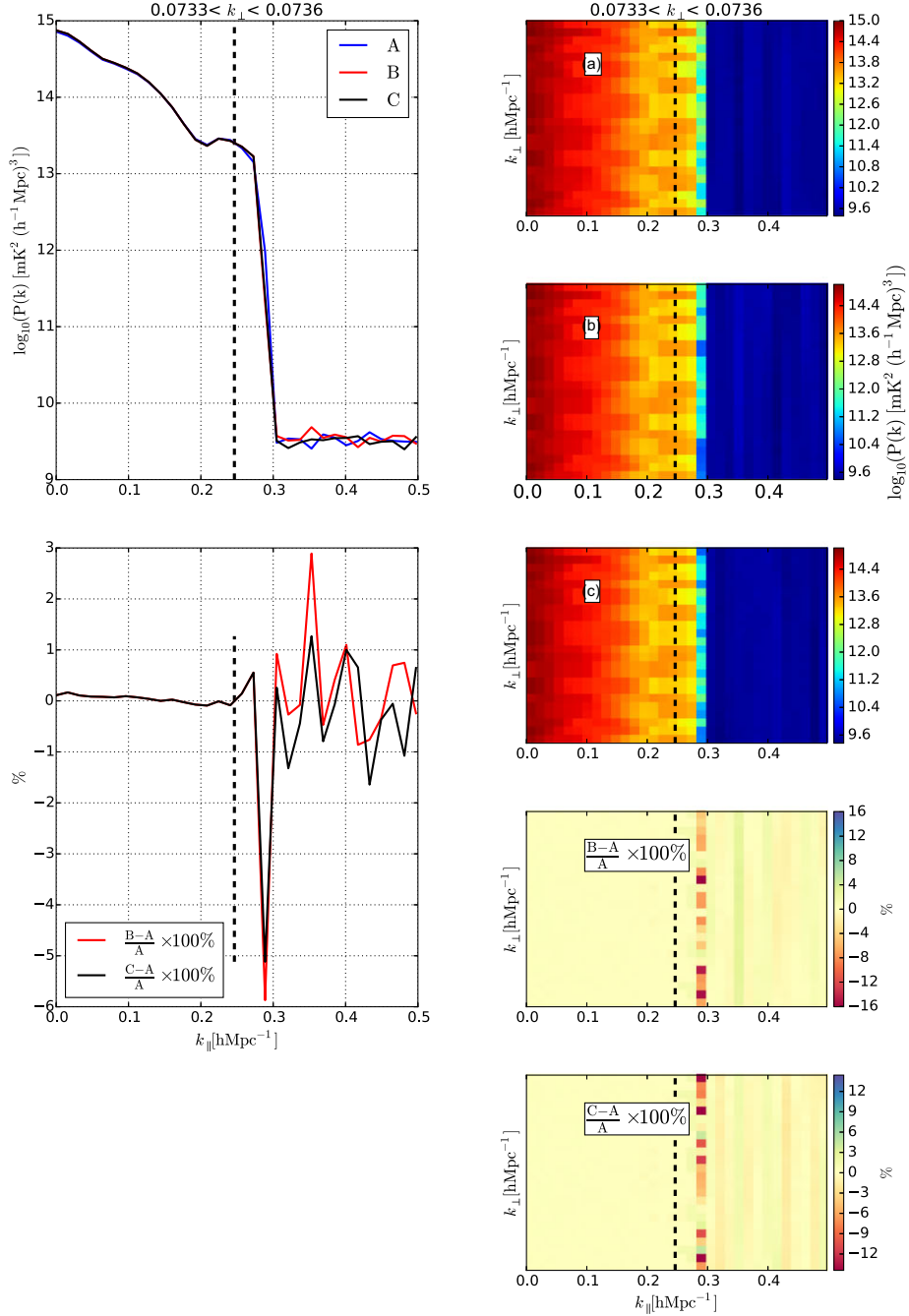


Figure 12. Right: Two-dimensional power spectra after (A) OmniCal, (B) CorrCal re-calibration without sources info, and (C) CorrCal re-calibration with sources information. The lower panels are the percentage deviation of B to A and C to A, respectively. The black dashed vertical line is representing the horizon (τ_H) limit. Left: The averaged slice of power spectra from $k_{\perp} = 0.0733$ h Mpc⁻¹ to $k_{\perp} = 0.0736$ h Mpc⁻¹ to both calibration schemes. The top panel is the one-dimensional average power spectrum as a function of k_{\parallel} for a portion of baselines (~ 150 m) to both calibration strategies. The bottom panel is the percentage deviation.

for radio interferometry, which potentially helps to probe the EoR using the 21-cm signal.

ACKNOWLEDGEMENTS

We would like to acknowledge the PAPER collaboration for providing us data that have been used in this work. YZM acknowledges the support of NRF-120385, NRF-120378, and NSFC-11828301. PK acknowledges funding support from the South African–China Collaboration program in Astronomy at the University of KwaZulu-Natal.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

- Ali Z. S. et al., 2015, *ApJ*, 809, 61
 Barkana R., Loeb A., 2001, *Phys. Rep.*, 349, 125
 Barry N., Hazelton B., Sullivan I., Morales M. F., Pober J. C., 2016, *MNRAS*, 461, 3135

Beardsley A. P. et al., 2016, *ApJ*, 833, 102
 Bernardi G. et al., 2009, *A&A*, 500, 965
 Bowman J. D., Morales M. F., Hewitt J. N., 2009, *ApJ*, 695, 183
 Byrne R. et al., 2019, *ApJ*, 875, 70
 Chapman E., Jelić V., 2019, preprint (arXiv:1909.12369)
 Cheng C. et al., 2018, *ApJ*, 868, 26
 Datta A., Bowman J. D., Carilli C. L., 2010, *ApJ*, 724, 526
 DeBoer D. R. et al., 2017, *PASP*, 129, 045001
 Dillon J. S. et al., 2014, *Phys. Rev. D*, 89, 023002
 Dillon J. S. et al., 2018, *MNRAS*, 477, 5670
 Dillon J. S. et al., 2020, *MNRAS*, 499, 5840
 Dillon J. S., Liu A., Tegmark M., 2013, *Phys. Rev. D*, 87, 043005
 Edler H. W., de Gasperin F., Rafferty D., 2021, *A&A*, 652, A37
 Ewall-Wice A., Dillon J. S., Liu A., Hewitt J., 2017, *MNRAS*, 470, 1849
 Furlanetto S. R., Oh S. P., Briggs F. H., 2006, *Phys. Rep.*, 433, 181
 Hothi I. et al., 2021, *MNRAS*, 500, 2264
 Hurley-Walker N. et al., 2017, *MNRAS*, 464, 1146
 Jacobs D. C. et al., 2013, *ApJ*, 776, 108
 Kohn S. A. et al., 2016, *ApJ*, 823, 88
 Kolopanis M. et al., 2019, *ApJ*, 883, 133
 Li W. et al., 2018, *ApJ*, 863, 170
 Li W. et al., 2019, *ApJ*, 887, 141
 Liu A., Parsons A. R., 2016, *MNRAS*, 457, 1864
 Liu A., Tegmark M., 2011, *Phys. Rev. D*, 83, 103006
 Liu A., Tegmark M., Bowman J., Hewitt J., Zaldarriaga M., 2009, *MNRAS*, 398, 401
 Liu A., Tegmark M., Scott M., Lutomirski A., Zaldarriaga M., 2010, *MNRAS*, 408, 1029
 Liu A., Parsons A. R., Trott C. M., 2014a, *Phys. Rev. D*, 90, 023018
 Liu A., Parsons A. R., Trott C. M., 2014b, *Phys. Rev. D*, 90, 023019
 Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, *MNRAS*, 283, 1388
 Mertens F. G. et al., 2020, *MNRAS*, 493, 1662
 Moore D. F., Aguirre J. E., Parsons A. R., Jacobs D. C., Pober J. C., 2013, *ApJ*, 769, 154
 Morales M. F., Hewitt J., 2004, *ApJ*, 615, 7
 Morales M. F., Wyithe J. S. B., 2010, *ARA&A*, 48, 127
 Morales M. F., Hazelton B., Sullivan I., Beardsley A., 2012, *ApJ*, 752, 137
 Oh S. P., 2001, *ApJ*, 553, 499
 Orosz N., Dillon J. S., Ewall-Wice A., Parsons A. R., Thyagarajan N., 2018, *MNRAS*, 487, 537
 Paciga G. et al., 2011, *MNRAS*, 413, 1174
 Parsons A. R. et al., 2010, *AJ*, 139, 1468
 Parsons A. R. et al., 2014, *ApJ*, 788, 106
 Parsons A. R., Backer D. C., 2009, *AJ*, 138, 219
 Parsons A. R., Pober J. C., Aguirre J. E., Carilli C. L., Jacobs D. C., Moore D. F., 2012, *ApJ*, 756, 165
 Patil A. H. et al., 2017, *ApJ*, 838, 65
 Pober J. C. et al., 2013, *ApJ*, 768, L36
 Pober J. C. et al., 2014, *ApJ*, 782, 66
 Pober J. C. et al., 2016, *ApJ*, 819, 8
 Pritchard J. R., Loeb A., 2012, *Rep. Prog. Phys.*, 75, 086901
 Rahimi M. et al., 2021, *MNRAS*, 508, 5954
 Santos M. G., Cooray A., Knox L., 2005, *ApJ*, 625, 575
 Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, *A&A*, 345, 380
 Sievers J. L., 2017, preprint (arXiv:1701.01860)
 Thyagarajan N. et al., 2013, *ApJ*, 776, 6
 Tingay S. J. et al., 2013, *PASA*, 30, e007
 Trott C. M. et al., 2020, *MNRAS*, 493, 4711
 van Haarlem M. P. et al., 2013, *A&A*, 556, A2
 Vedantham H., Shankar N. U., Subrahmanyan R., 2012, *ApJ*, 745, 176
 Wang X., Tegmark M., Santos M. G., Knox L., 2006, *ApJ*, 650, 529
 Wayth R. B. et al., 2018, *PASA*, 35, 33
 Wieringa M. H., 1992, *Exp. Astron.*, 2, 203
 Woodbury M. A., 1950, *Memorandum Rep.*, 42, 336
 Zhang Z. et al., 2020, *PASA*, 37, e045
 Zheng H. et al., 2014, *MNRAS*, 445, 1084

APPENDIX A: DELAY SPECTRUM

After data with XX and YY linear polarization being re-calibrated using `CORRCAL` separately, we form the Stokes I parameter by combining each linear polarization following Moore et al. (2013):

$$V_I = \frac{1}{2}(V_{XX} + V_{YY}),$$

where V_{XX} and V_{YY} , respectively, are visibilities from XX and YY linear polarization. We then use V_I (Stokes I polarization visibility) to form the power spectra using a delay-transform approach following that of Ali et al. (2015).

Approximating the interferometric visibility equation that is defined in equation (25) over the flat-sky model, we have

$$d(u, v, f) = \int dl dm I(l, m, f) A(l, m, f) e^{-2\pi i(ul+vm)}, \quad (\text{A1})$$

where (l, m) are direction cosines of a unit vector \hat{r} , pointing to the source on the sky. These sky-plane direction cosines have the corresponding antenna-plane Fourier dual u and v , respectively, that is defined in Section 3.1. Following the representation of Parsons & Backer (2009), we can rewrite equation (A1) in terms of geometric time delay τ_g for a single baseline \mathbf{b} as

$$d_b(f) = \int dl dm I(l, m, f) A(l, m, f) e^{-2\pi i f \tau_g}, \quad (\text{A2})$$

where

$$\tau_g = \frac{1}{c} \mathbf{b} \cdot \hat{r} = \frac{1}{c} (b_x l + b_y m), \quad (\text{A3})$$

with $\mathbf{b} = (b_x, b_y)$, $(u, v) = f\mathbf{b}/c$ and $\hat{r} = (l, m)$. Here, b_x and b_y are projected baseline length in metre along east and north directions in the antenna plane, respectively. Note that τ_g in equation (A2) shows the time shifting of signal arriving one antenna relative to other due to source location on the sky with respect to the baseline vector orientation in the antenna plane.

Delay transform technique begins by converting the frequency spectrum of the visibility into the delay spectrum through application of Fourier transform on the visibility along its frequency direction. Applying the windowing function $W(f)$ over equation (A2) and then Fourier transforming it along its frequency axis, the delay transformed visibility takes a form

$$\begin{aligned} \tilde{V}_b(\tau) &= \int_B df W(f) v_b^{\text{mea}}(f) e^{2\pi i f \tau} \\ &= \int_B df W(f) \left[\int dl dm I(l, m, f) A(l, m, f) e^{-2\pi i f \tau_g} \right] \\ &\quad \times e^{2\pi i f \tau} \\ &= \int_B dl dm \tilde{W}(\tau) * \tilde{I}(l, m, \tau) * \tilde{A}(l, m, \tau) * \delta_D(\tau - \tau_g), \end{aligned} \quad (\text{A4})$$

where the window function $W(f)$ chosen by the data analyst is used to control the quality of delay spectrum as suggested in Vedantham et al. (2012) and Thyagarajan et al. (2013), B is the bandwidth of observation over which the integration has to be carried out, the delay parameter τ is the Fourier dual of f in unit of time, $\tilde{V}_b(\tau)$ is the delay transformed visibility observed by a baseline b , and $*$ stands for convolution. Hence, from equation (A4), one can clearly see that there is mapping between point sources on the sky in \tilde{V}_b space and the Dirac-delta function δ_D , convolved by Fourier transform of window function $\tilde{W}(\tau)$, sky flux density $\tilde{I}(l, m, \tau)$, and the antenna beam response $\tilde{A}(l, m, \tau)$, as shown in Parsons et al. (2012). Analytically, the delay formalism that relates

the spatial power spectrum $P_{21}(k_{\parallel}, k_{\perp})$ of 21-cm signal from the EoR fluctuations to the delay-transformed visibility \tilde{V} has been derived in Parsons et al. (2012) as

$$P_{21}(k_{\parallel}, k_{\perp}) \approx \tilde{V}_{21}^2 \left(\frac{\lambda^2}{2k_{\text{B}}} \right)^2 \frac{X^2 Y}{\Omega B}, \quad (\text{A5})$$

where λ is the observing wavelength, k_{B} is the Boltzmann constant, B is the observing bandwidth, Ω is the solid angle of the primary beam of the antenna, and X and Y are cosmological scalars that convert observed angles and frequencies into $h \text{ Mpc}^{-1}$ as calculated in Parsons et al. (2012).

Following the method in Pober et al. (2013), the successive timestamp in delay-transformed visibilities has been cross-multiplied to avoid the noise bias. Using the formalism in Kohn et al. (2016), it can be written as

$$\tilde{V}_{21}^2 \approx |\tilde{V}_{21}(\tau, t) \times \tilde{V}_{21}(\tau, t + \Delta t) e^{i\theta_{\text{zen}}(\Delta t)}|^2. \quad (\text{A6})$$

Here, $\Delta t \approx 42.9 \text{ s}$ (a time resolution of PAPER-64 data) and $\theta_{\text{zen}}(\Delta t)$ is the zenith re-phasing factor. The line-of-sight k_{\parallel} , and transverse co-moving k_{\perp} components of k in terms of the instrumental and cosmological parameters have been shown in Morales & Hewitt (2004),

$$k_{\perp} = \frac{2\pi |\mathbf{u}|}{D_{\text{m}}(z)},$$

$$k_{\parallel} \approx \eta \frac{2\pi H_0 f_{10} E(z)}{c(1+z)^2},$$

where $|\mathbf{u}| = \sqrt{u^2 + v^2}$, η is the Fourier conjugate of f , which used to denote the spatial frequency along the line of sight and it has the unit of time, $f_{10} = 1420 \text{ MHz}$ is the rest frequency of 21-cm emission from cosmic re-ionization, c is the speed of light in free space, $D_{\text{m}}(z)$ is the transverse co-moving distance at the reference redshift z , the present-day Hubble constant $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and the function $E(z) = \sqrt{\Omega_{\text{m}}(1+z)^3 + \Omega_{\text{k}}(1+z)^2 + \Omega_{\Lambda}}$. Here, the cosmological density parameters Ω_{m} , Ω_{Λ} , and Ω_{k} are total matter density, the dark energy density, and curvature, respectively. The geometric delay τ_{g} in equation (A3) sets the maximum delay limit beyond which no emissions enter the interferometry involving that baseline. This limit will be attained when the angle between \mathbf{b} and $\hat{\mathbf{r}}$ is set to be zero. Such an alignment to both vectors is achieved, the maximum delay is commonly referred to ‘horizon limit (τ_{H})’.

Theoretically, over the $(k_{\perp}, k_{\parallel})$ cylindrical space, it can be determined from

$$k_{\parallel} = \frac{H_0 E(z) D_{\text{m}}(z)}{c(1+z)} k_{\perp}. \quad (\text{A7})$$

As exhaustively discussed in Parsons et al. (2012), any strictly smooth foreground emissions with power-law spectra are confined within the narrow range of horizon limits such that $-\tau_{\text{H}} \leq \tau \leq \tau_{\text{H}}$. However, intrinsic sky components and instrumental effects that are generally non-smooth can scatter beyond the horizon limit. This is due to the sidelobes of the convolving kernel of sky and beam, i.e. $\tilde{I} * \tilde{A}$, which broadens the footprint of k_{\parallel} modes in cosmological k space. This kernel is narrow for smooth spectrum foreground sources and confined within the wedge. However, flux beyond the horizon limit is representing the power spectrum of unsmooth emissions such as 21-cm signal. Therefore, regions in a wide range of k_{\parallel} modes are considerably affected by the 21-cm signal if one keeps the frequency response of the instrument smoother. More of these modes are free from foreground contamination on shortest baseline, i.e. at smallest k_{\perp} regions (Datta, Bowman & Carilli 2010; Parsons et al. 2012; Vedantham et al. 2012; Pober et al. 2013).

APPENDIX B: DETERMINING THE EOR SIGNAL LOSS IN CORR CAL ANALYSIS

To determine whether our analysis methods are lossless or not in the EoR power spectrum estimate, first, we simulate visibility data using HERA simulation pipeline (`hera_sim`⁸). For this simulation, we use the analytic airy beam model (the same for all antenna elements), the simulated antenna gains, the GaLactic and Extragalactic All-sky MWA (GLEAM) sky sources (Hurley-Walker et al. 2017), and the simulated EoR signal using `hera_sim`. Then, to quantify our new calibration approach does not lead to the EoR power spectrum signal suppression, we propagate simulated data into the `CORR CAL` for calibration. We then compute the 2D power spectrum for both simulated and recovered data after `CORR CAL` calibration using the delay spectrum approach described in Appendix A. Fig. B1 shows the delay space power spectra for both simulated and recovered data. From the figure, it is apparent that the recovered power spectra after `CORR CAL` are comparable to the expected one, confirming that `CORR CAL` calibration method does not lead to significant signal loss.

⁸https://github.com/HERA-Team/hera_sim.git

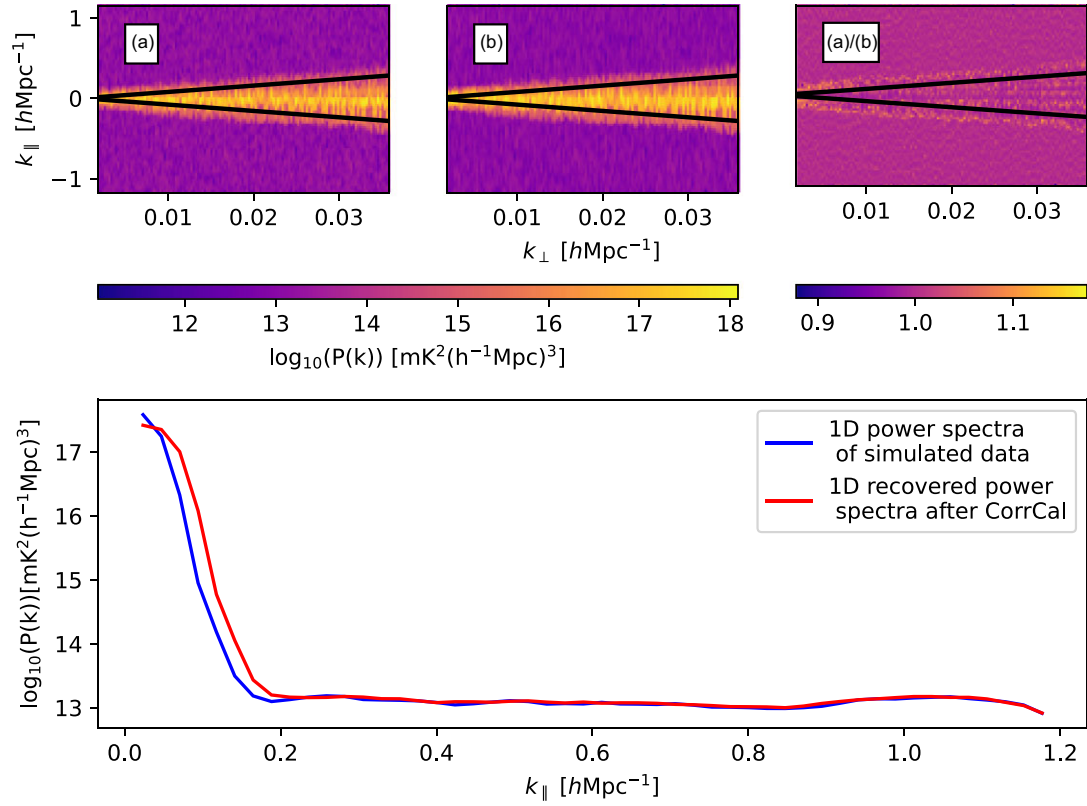


Figure B1. Top: (A) Expected delay power spectra, (B) recovered power spectra after `CorrCal` calibration, and (C) the power spectra ratio of B to A. The solid black dashed line is the horizon limit, τ_H . Bottom: Averaged 1D power spectra as a function of k_{\parallel} for a portion of baselines in the ~ 30 m group.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.