FERMILAB-PUB-21-219-T

MNRAS **000**, 000–000 (0000) Preprint 18 May 2021 Compiled using MNRAS LATEX style file v3.0

AlgoSCR: An algorithm for Solar Contamination Removal from radio interferometric data

Anh Phan[1]*, Santanu Das[1,2]†, Albert Stebbins[2], Peter Timbie[1]‡, Reza Ansari[3], Shifan Zuo[4], Jixia Li[5], Trevor Oxholm[1], Fengquan Wu[5], Xuelei Chen[5,6,7], Shijie Sun[5,6], Yougang Wang[5], Jiao Zhang[8]

[1] *Department of Physics, University of Wisconsin Madison, 1150 University Ave, Madison WI 53703, USA*
[2] *Fermi National Accelerator Laboratory, P.O. Box 500, Batavia IL 60510-5011, USA*
[3] *Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405 Orsay, France*
[4] *Department of Astronomy and Tsinghua Center for Astrophysics, Tsinghua University, Beijing 100084, P.R.China*
[5] *National Astronomical Observatory, Chinese Academy of Science,20A Datun Road, Beijing 100101, P. R. China*
[6] *University of Chinese Academy of Sciences Beijing 100049, P. R. China*
[7] *Center of High Energy Physics, Peking University, Beijing 100871, P. R. China*
[8] *School of Physics and Electronics Engineering, Shanxi University, Taiyuan 030006, P. R. China*

arXiv:2105.07126v1 [astro-ph.IM] 15 May 2021

18 May 2021

ABSTRACT

Hydrogen intensity mapping is a new field in astronomy that promises to make three-dimensional maps of the matter distribution of the Universe using the redshifted 21 cm line of neutral hydrogen gas (HI). Several ongoing and upcoming radio interferometers, such as Tianlai, CHIME, HERA, HIRAX, etc. are using this technique. These instruments are designed to map large swaths of the sky by drift scanning over periods of many months. One of the challenges of the observations is that the daytime data is contaminated by strong radio signals from the Sun. In the case of Tianlai, this results in almost half of the measured data being unusable. We try to address this issue by developing an algorithm for solar contamination removal (AlgoSCR) from the radio data. The algorithm is based on an eigenvalue analysis of the visibility matrix, and hence is applicable only to interferometers. We apply AlgoSCR to simulated visibilities, as well as real daytime data from the Tianlai dish array. The algorithm can remove most of the solar contamination without seriously affecting other sky signals and thus makes the data usable for certain applications.

Key words: methods: analytical – methods: data analysis – instrumentation: interferometers – cosmology: observations – radio continuum: general

1 INTRODUCTION

Cosmologists study the Universe on the largest observable distance scales in order to understand its origin and evolution. In the past few decades, cosmic microwave background (CMB) instruments have mapped almost the entire sky with high sensitivity and fine angular resolution. These maps measure the intensity and polarization fluctuations at the last scattering surface and remain a primary tool for studying the Universe. However, for understanding the nature of dark matter and dark energy, it is essential to study the evolution of structure as a function of time. Galaxy redshift surveys have been extremely successful in mapping the large scale structure of the Universe by cataloging the distribution of luminous galaxies in redshift space. These maps can be used, for example, to observe the characteristic baryon-acoustic oscillation (BAO) signal, which can be used as a standard ruler to extract cosmological parameters. However, as we map larger and more distant volumes of the Universe, the method faces multiple challenges. For example, the galaxies become fainter and spectral lines are redshifted to wavelengths that are difficult to detect from the ground.

Intensity mapping, a radically different technique, creates 3D maps using the 21 cm emission of neutral hydrogen (HI) without resolving individual galaxies. This line is unique in cosmology as, for $\lambda > 21$ cm, it is the dominant astronomical line emission for all redshifts. Hence, to a good approximation the wavelength of a spectral feature can be converted to a redshift without having to first identify the atomic transitions. In principle, HI intensity-mapping could be used to make 3D maps of matter at all redshifts up into the "dark ages" ($z \approx 100$), even before galaxies have formed.

The first HI intensity mapping observations began over a decade ago (Abdalla and Rawlings 2005; Peterson et al. 2006; Morales 2008; Chang et al. 2008a; Mao et al. 2008) and inter-

*anh@wisc.edu
†sanjone@gmail.com
‡pttimbie@wisc.edu

© 0000 The Authors

est has continued to grow (Ansari et al. 2018; Liu and Shaw 2020; Slosar et al. 2019). A number of dedicated projects have been launched to detect the signal and turn the technique into a useful cosmological tool. These are mainly interferometers such as CHIME (Bandura et al. 2014; Newburgh et al. 2014), Tianlai (Chen 2011; Xu et al. 2014; Das et al. 2018; Li et al. 2020; Wu et al. 2020), MWA (Tingay et al. 2013), LWA Eastwood et al. (2018), HERA (DeBoer et al. 2017), HIRAX (Newburgh et al. 2016), and PUMA (Slosar et al. 2019), but also include single dishes with multiple feed antennas, such as BINGO (Battye et al. 2012; Dickinson 2014; Wuensche et al. 2019) and FAST(Hu et al. 2020). Intensity mapping instruments can address questions at a variety of redshift ranges. At $z \sim 10$ they probe the Epoch of Reionization (EoR), star formation, and galaxy assembly, while at lower redshifts they trace large scale structure for studies of dark energy, etc. (Peterson et al. 2006; Bull et al. 2015; Battye et al. 2004; Abdalla and Rawlings 2005; Chang et al. 2008b; Mao et al. 2008; Morales 2008).

So far, the HI signal has not been detected using intensity mapping by itself. Intensity mapping observations have set upper limits on HI from the EoR Ali et al. (2015) and, in the post-recombination epoch, have detected HI when cross-correlated with galaxy redshift surveys (Masui et al. 2010, 2013; Anderson et al. 2018). A number of challenging systematic effects must be overcome to allow autocorrelation detections. The foremost of these is separating the HI signal from Galactic and extra-Galactic astronomical foregrounds, which are $\sim 4-5$ orders of magnitude brighter (Liu and Shaw 2019). The Sun represents an astronomical foreground that is even brighter and of a different character.

The daytime data from radio interferometer arrays in general, and the Tianlai dish array in particular, are contaminated by the solar signal, making the data unusable for most astronomical analyses. The lost data have a significant impact on observing efficiency; reaching the required survey sensitivity means observing the sky for almost twice the number of days. This penalty is particularly problematic for HI intensity mapping, where long integration times (months or years) are necessary to detect the HI signal. Furthermore, this data loss prevents obtaining continuous, 24 hr data sets, which allow dense coverage of the $u-v$ plane and facilitate detection of periodic signals. Furthermore, not having 24 hours of continuous usable observations prevents the application of m-mode map making techniques Shaw et al. (2014). The (u,v) coverage can still be quite good with nighttime data, and one can recover full 24 hours RA coverage by combining nighttime data from observations about 6 months apart. In this paper we try to remove the solar contamination from the daytime data from a radio interferometer. We have used the data from the Tianlai dish array as our test sample. However, the problem is not unique to Tianlai; the same algorithm may be used for other radio interferometric observations. While daytime observations with single dish radio telescopes are also plagued by the Sun's signal, this algorithm is only applicable to interferometer arrays.

The Tianlai Project is led by the National Astronomical Observatory of China (NAOC). It consists of two pathfinder radio interferometers, an array of cylinder antennas and an array of dishes, at a radio-quiet site in Xinjiang, China (Chen 2012; Li et al. 2020; Wu et al. 2020). The objective is to obtain high fidelity 3D images of the northern sky using HI intensity mapping. Construction of the Pathfinder was completed in 2016 and the arrays now undergoing commissioning. So far, we have collected over 258 days of data with the dish array and 114 days with the cylinder array and have begun the process of calibrating, making maps, and removing foregrounds. Both arrays are operated in drift-scan mode. Near-term analyses include cross-correlating the radio maps with galaxy redshift surveys. If successful, the arrays can be expanded to increase sensitivity.

The analysis described in this paper concentrates on the dish array data. The Tianlai dish array consists of 16 steerable, 6 m diameter dishes; a schematic is shown in Fig. 1, which also shows the dish numbering scheme. We use these dish numbers for referring to different baselines in the paper. The feed antennas, amplifiers, and reflectors are designed to operate from 600 MHz to 1420 MHz. The instrument can be tuned to operate in an RF bandwidth of 100 MHz centered at any frequency in this range by adjusting the local oscillator frequency in the receivers and replacing the bandpass filters. The dish array currently operates between 685 MHz and 810 MHz, corresponding to redshift $0.75 < z < 1.07$, divided into 512 equally spaced frequency bins of width 244 kHz. The 16 dual polarization feeds yield 32 autocorrelation visibilities and $32 \times (32-1)/2 = 496$ cross-correlation visibilities, which are currently sampled every second. An additional electronic backend is being installed to search for transients (e.g. fast radio bursts) in parallel with the standard correlator used for HI mapping described here. The system noise temperatures for the dish antennas are $80-85$ K (Li et al. 2020; Zhang et al. 2016; Das et al. 2018; Wu et al. 2020).

The objective of this paper is to describe an eigenvalue-based approach for removing solar contamination from radio interferometric data. We propose an algorithm, AlgoSCR, that can remove most of the solar contamination. The paper is organized as follows. In the second section, we discuss the solar contamination problem in the Tianlai dish array in detail. The third and the fourth sections give the detailed algorithm for removing the solar contamination. We also show the results of our analysis on the real Tianlai data. To test what fraction of the Sun signal can be removed by our algorithm, and how much signal from other cosmic sources is removed by it, in section five we have applied it to simulated data where the amount of solar contamination and the external sources are known. In the discussion section we assess the efficacy of the method and the issues that we face when applying it. We also describe some future directions to pursue with this approach.

2 THE SOLAR CONTAMINATION PROBLEM

The Tianlai data show strong contamination from the solar signal during the daytime. In Fig. 2 we plot the sum of the absolute visibility from 10 frequency channels (out of 512 channels) at the center of the band, for 4 consecutive days (total 96 hours) for four different baselines. The horizontal axis shows the time in hours, starting at the beginning of the observations. The 4 different plots are for 4 representative visibilities. We can see a roughly smooth visibility amplitude for about 10 hours every day and then a sudden increase in the absolute visibility and a noisy pattern for about next 14 hours. The smooth part of the visibility comes from the data

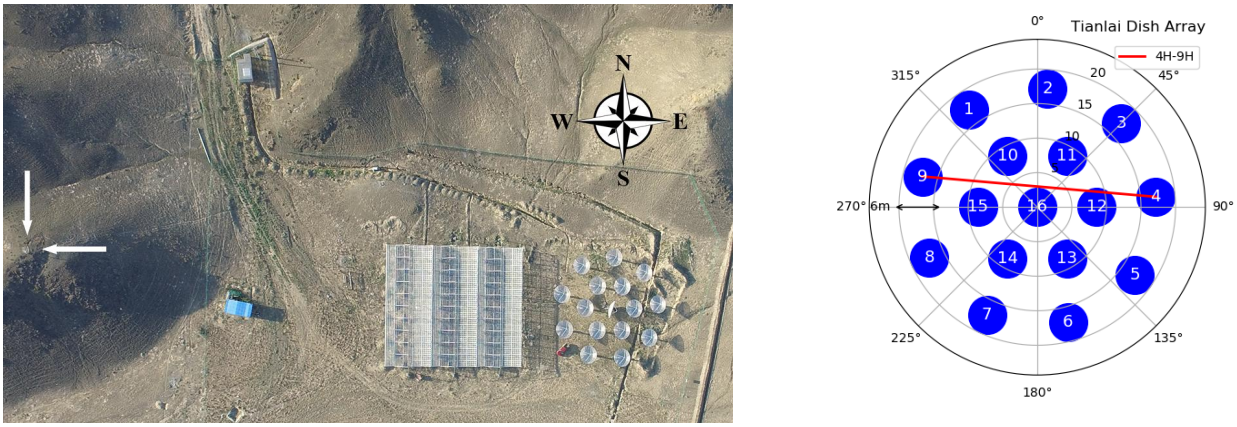

Figure 1. Left: A top view photograph of the Tianlai arrays, which consist of the Dish Array and the Cylinder Array. The photo was taken with a drone at a height of 280 m above the ground. The arrays saw first light in 2016. The position of the calibration noise source is indicated by the white arrows on the left. **Right:** A schematic diagram of the Tianlai dish array. The dishes are arranged in two concentric circles of radius 8.8 m and 17.6 m around a central dish. The dishes have dual-linear polarization feed antennas with one axis oriented parallel to the altitude axis (horizontal, H, parallel to the ground)) and the other orthogonal to that axis (vertical, V). For example, red line shows one of the baselines that is the H polarization of dish 4 correlated with the H polarization of dish 9: [4H 9H]. The above image is reproduced from (Wu et al. 2020).

taken during the night, whereas the noisy high amplitude section corresponds to daytime data.

The plots clearly show that the daytime signal is several times stronger than the night. The shape of the contamination pattern also varies with baseline. Some of the baselines show a bumpy feature with the strongest visibility occurring near noon, whereas for other baselines the signal is strongest during Sunrise and Sunset and shows a 'dip' feature during the daytime. The top 2 plots are the auto-polarization visibility corresponding to two horizontal feeds, whereas the bottom two plots show two auto-polarization visibility from two vertical feeds. The auto-polarization signals from similar feeds on other baselines show roughly similar types of patterns, except for a couple of baselines. The data are taken during observations of the North Celestial Pole (NCP) in April, 2019. During this observing period the path of the Sun is located at an angle of approximately 85° from the direction of the main beam. The plot gives an overview of the magnitude of the solar contamination problem in the Tianlai dish array.

An obvious conclusion of this strong daytime visibility is that the telescopes are responding to the Sun's illumination of their far sidelobes. For the baselines measuring correlations of the H polarization, the antenna sidelobes are aligned with the direction of Sun near noon, providing a strong visibility at mid-day, while for the V polarization the Sun falls between two side lobes at noon, producing stronger signal during Sunrise and the Sunset. These effects are consistent with the expected responses of the feed antennas, which are essentially orthogonally oriented crossed dipoles. In Fig. 3 we show part of the simulated beam pattern for a single dish measured using an electromagnetic simulation package (CST [1]). We also show the plane of the path of the Sun using a blue circle for both polarizations under assumptions that the beam shape will be the same for both feeds. Even though the real shape of the beam function may be different from the simulated one, the plot can provide us an understanding about the origin of the particular type of shape seen by the telescope. The figure shows that for some polarizations the beam sensitivity in the Sun's path is highest during Sunrise and Sunset, and the path does not go through any other strong sidelobes throughout the day, whereas for the other beam, the path of Sun intersects some of the side lobes during midday, causing a strong response at noon. In Fig. 4 we show a cut through Fig. 3 beams, corresponding to the sun track during daytime. We can see that for one of the polarization we are getting a low amplitude during the midday whereas for the other polarization (right plot) the amplitude is comparatively high during noon. The simulated patterns shown in Fig. 4 doesn't exactly replicate the observed pattern of Fig 3, because the sidelobes from these EM simulations don't exactly match that of the real beam. The sidelobes at this particular angle are also highly cluttered. A couple of degrees changes in the path gives rise to a very different shape in the sidelobes, making it difficult to reconstruct the exact pattern through such EM simulation.

This daily response to the Sun signal is relatively constant over a period of a year. Fig. 5 shows the directive gain of the dish antennas as computed by an electromagnetic simulation. The Sun enters the sidelobes of the antennas in a range of polar angles for which the beam patterns are relatively flat. The simulation is consistent with measurements of the daytime visibilities at different times of the year. Using the eigenvalue analysis described below, Fig. 6 shows the contribution by the Sun to the visibility for a typical baseline during January, 2018 and then again in April, 2019. As the paths of the Sun through the sidelobes of the antennas are different, at different time of the year, the visibilities are also slightly different. However, we can see that the amplitudes are within $\sim 30\%$ of each other.

The complex visibilities for four randomly chosen baselines are shown as 'waterfall plots' in Fig. 7 for a 24 hour period. We can see that the daytime data is dominated by bright

[1] https://www.3ds.com/products-services/simulia/products/cst-studio-suite/

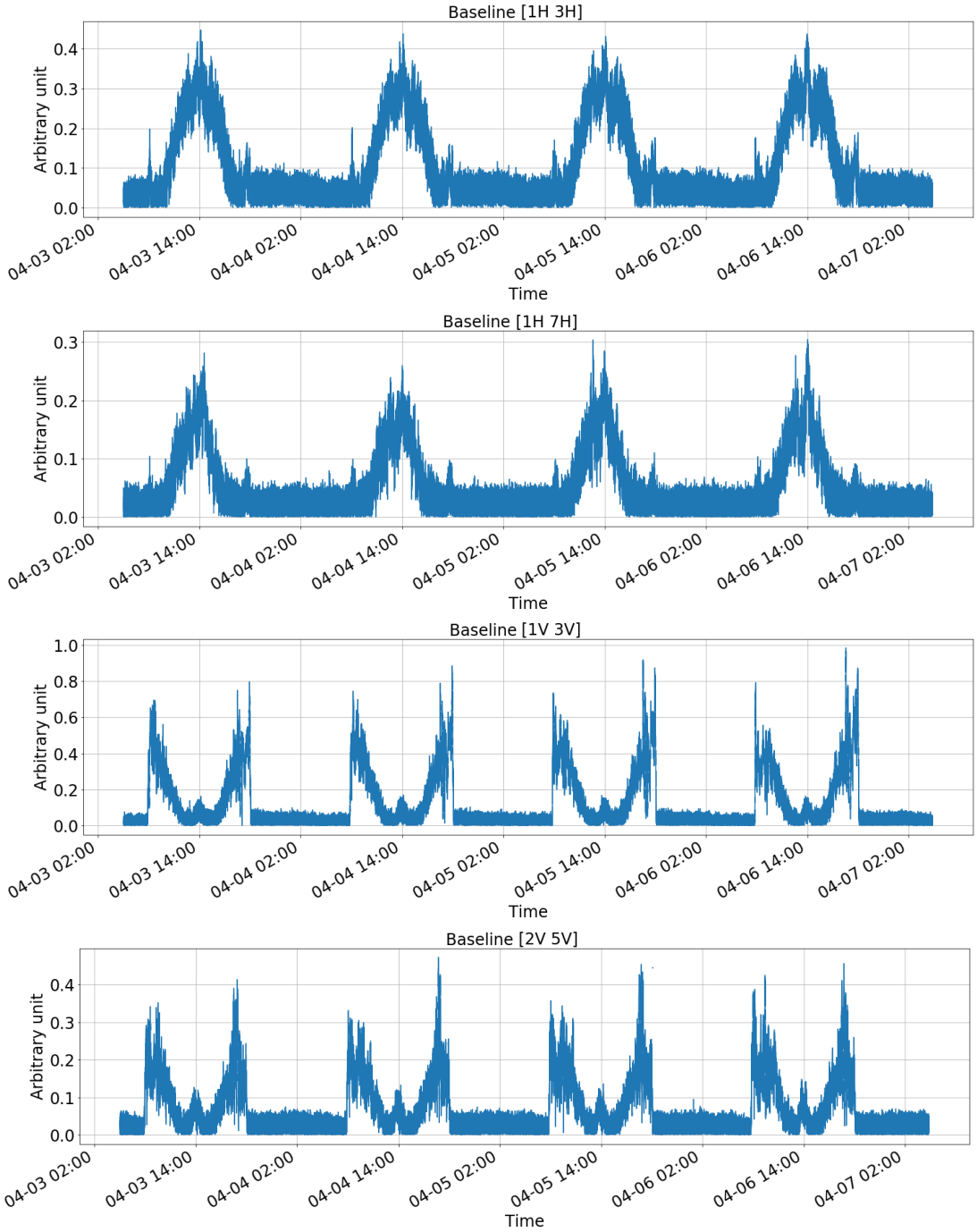

Figure 2. Value of the (uncalibrated) cross-correlation visibility amplitudes averaged over the 10 central frequency bands during 4 days of observations in April, 2019. Integration time is one second. Each plot corresponds to a different baseline, as indicated. The baseline numbering scheme appears in Fig. 1. Time is given in local sidereal time.

fringes caused by the Sun. On the other hand, the pattern in the nighttime data comes from the much dimmer radio sky and has a very different character. The dominant fringes in the nighttime data come from a combination of weak sources near the NCP and bright sources far from the NCP, particularly Cassiopeia A (Cas A) and Cygnus A (Cyg A).

3 REMOVING SUN CONTAMINATION USING EIGENVALUE ANALYSIS

We start by defining the notation used in this paper. The visibility matrix is given by

$$\mathbf{V} = [\mathbf{D}^{\mathbf{s}}\mathbf{G}]^{\dagger}\,[\mathbf{D}^{\mathbf{s}}\mathbf{G}] + \left\langle [\mathbf{N}]^{\dagger}[\mathbf{N}] \right\rangle , \tag{1}$$

where any individual component is given by

$$\mathbf{V}_{(i,j)} \quad = \quad \langle E_i^* E_j \rangle \, . \tag{2}$$

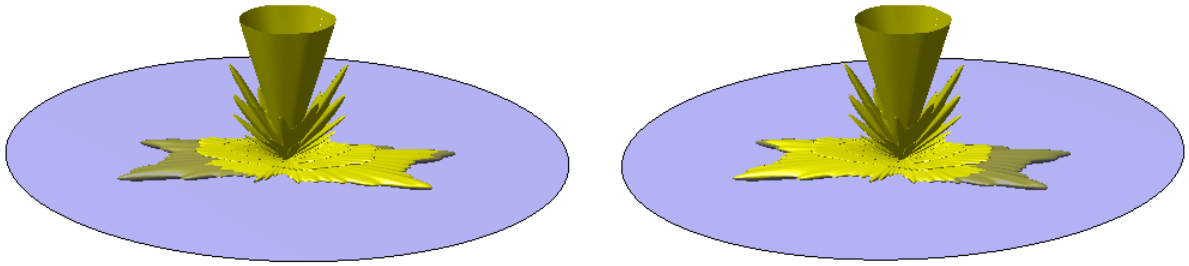

Figure 3. Plot of the beam pattern for a single dish in the Tianlai dish array, calculated using an electromagnetic simulation. The value of the antenna directivity is plotted in arbitrary linear units (not in dB). The main beam is oriented in the vertical direction and truncated in order to emphasize the sidelobes. The path of the Sun is indicated by the blue plane for the observations analyzed here. The simulated beam shape is the same for the V and H polarizations, but one is oriented 90° with respect to the other. The plots show that the sidelobes are highly structured and the paths cross through different sidelobes for the different polarizations.

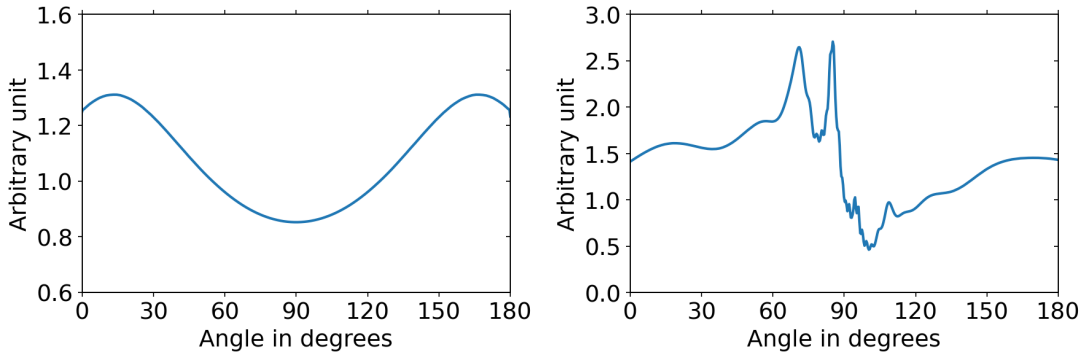

Figure 4. A cut through the Sun's track during daytime (180° corresponding to 12 hours) shown in Fig. 3. The value of the antenna directivity is plotted in arbitrary linear units (not in dB). The plot shows that for one of the polarizations, the power is low during the daytime, whereas for the other polarization there is high power during certain parts of the daytime. Even though the plot on the right does not show the exact feature that we are getting during daytime (Fig. 2), we must remember that the sidelobes from the electromagnetic simulations are not exactly the same as that of the real antenna.

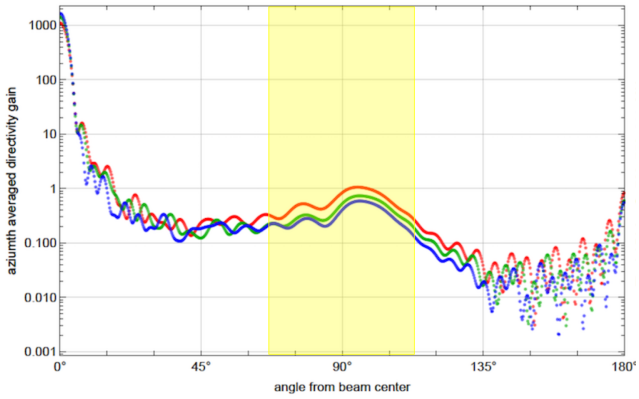

Figure 5. Simulated beam patterns as a function of beam angle θ from the beam center of the antennas for 3 different frequencies, 700 (red), 750 (green) and 800 MHz (blue). Each plot shows the absolute co-polar directive gain averaged over the azimuthal angle. The angle is the polar angle calculated from the center of the beam. The yellow shaded region shows the range of polar angles for which the Sun appears in the sidelobes of the beam, ranging from 66.55° at the Summer Solstice, to 113.45° at the Winter Solstice. The gain is relatively flat over this range of angles and causes the Sun signal to vary by only a factor of about 6 over the year.

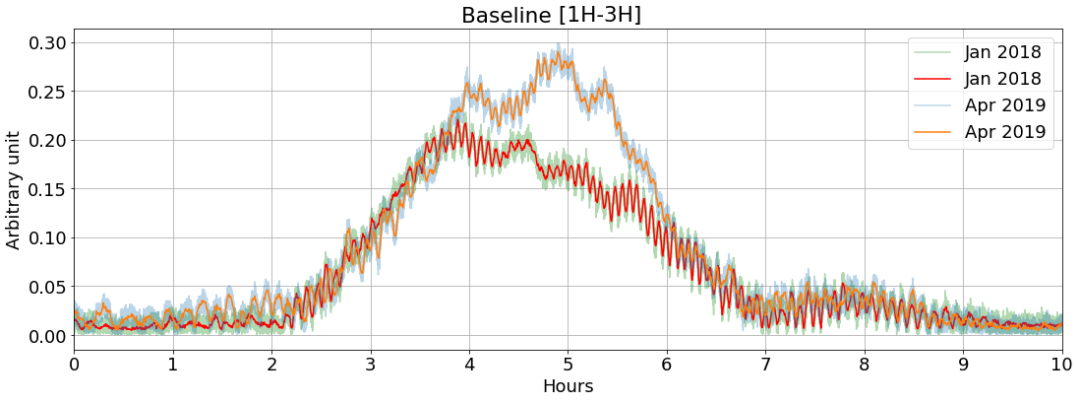

Figure 6. The amplitude of the daytime visibilities in January, 2018 and April, 2019. Due to the difference in the time of Sunrise in January and April, the 0 hr of each curve is adjusted so that the Sun signals from both data sets peak at about the same time. The fast oscillation fringes are seen in both observations. The green and blue curves are the amplitude of the visibility obtained from the telescope for January 2018 and April, 2019, respectively. The red and orange curves are corresponding spline fits to better highlight the fast oscillation fringes.

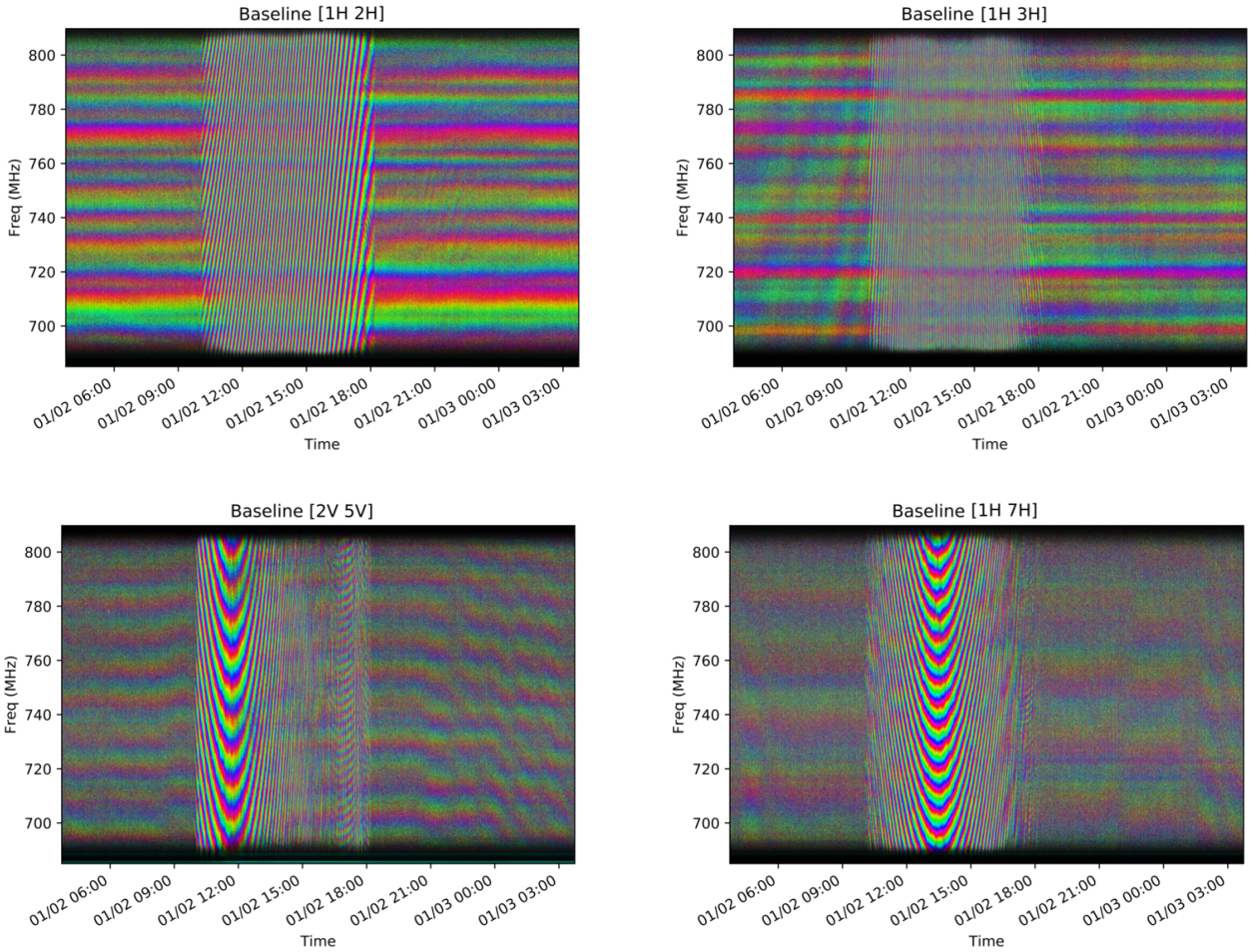

Figure 7. Left: The complex visibility for four typical baselines plotted over a 24 hour period. We represent the phase of the complex visibility by hue (color) and the amplitude by value (brightness) in a HSV (hue, saturation, value) display of the color model (see Fig. B2 for details). In each plot, the local sidereal time proceeds linearly from left to right, with a sampling interval of 1 s. The frequency increases linearly from bottom (685 MHz) to top (810 MHz) in 512 equally spaced frequency bins. The time interval from about 10:00 to 18:00 is dominated by the Sun.

E_i represents the voltage from receiver i and is given by

$$E_i = \left(\sum_s D_i \left(\vec{\omega}_s \right) e^{i\mathbf{k}\cdot\mathbf{r}_i} F_s \right) G_i + N_i, \quad (3)$$

where F_s is the electric field of the radio wave coming from a source on the celestial sphere, $D_i(\vec{\omega}_s)$ is the primary beam of antenna i, and this is a function of the direction vector $\vec{\omega}_s$. $\mathbf{k}$ is the 3-dimensional wavenumber Fourier dual to the position vector $\mathbf{r}_i$ of feed i. G_i is a direction-independent complex gain factor. N_i is the noise in receiver i.

The intensity of the source at any frequency ν, is given by

$$I_s(\nu) = |F_s(\nu)|^2 = F_s^*(\nu) F_s(\nu)\,. \quad (4)$$

For extended sources we need to integrate over different directions for calculating E_i:

$$E_i = \left(\int D_i \left(\vec{\omega}_s \right) e^{i\mathbf{k}\cdot\vec{r}_i} F_s d\omega_s \right) G_i + N_i\,. \quad (5)$$

The visibility is an ensemble average of the $E_i^* E_j$, i.e.

$$\begin{aligned} \mathbf{V}_{(i,j)} &= \langle\, E_i^*\, E_j \,\rangle_{\tau_{\rm int}} \\ &= \left[\frac{1}{\tau_{\rm int}} \int_0^{\tau_{\rm int}} E_i^*\, E_j \, {\rm d}t \right] \end{aligned} \quad (6)$$

where $\tau_{\rm int}$ is the integration time, which is constant for any time and frequency bin (t, ν). For the current Tianlai setup, the integration time is 1 s. The asterisk ($*$) represents the complex conjugate and the bracket $\langle\ \rangle$ represents the ensemble average.

Here we should note that the visibilities from different astrophysical sources are additive. Provided there is only one point source on the sky the visibility matrix, i.e. $\mathbf{V}_{(i,j)}$ at any time can be written as an outer product, of the electric field from the source measured at different feed antennas. Therefore, if the visibility matrix is decomposed into its corresponding eigenvalues and eigenvectors, there should be only one nonzero eigenvalue. In the presence of other weaker sources, the largest eigenvalue should correspond to the Sun signal and the eigenvector corresponding to the largest eigenvalue will roughly point toward the direction of that source in the eigenspace. The contributions from additional, weaker sources and noise may alter the direction slightly.

Comparing the visibility amplitudes between the daytime and the nighttime data in Fig. 2, we can infer that the largest contribution to the daytime signal is from the Sun, entering through the antenna sidelobes. Therefore, in the eigen decomposition of the visibility matrix the largest eigen value should represent the solar contamination.

3.1 Issues with the autocorrelation signal

The voltage from the feeds contain a contribution from the receiver noise. Therefore, the measured signal or voltage E_i for a given feed i is the sum of the sky signal, $E_{{\rm Sky}\,i}$ and the instrument noise, N_i, i.e. $E_i = E_{{\rm Sky}\,i} + N_i$.

Under the assumption that the noise terms from separate feeds are uncorrelated, we can say that the ensemble average of the noise from feed i and feed j is zero, i.e. $\langle N_i^* N_j \rangle \approx 0$. Therefore, the visibility for cross-correlated feed i and j, where $i \neq j$, is $\mathbf{V}_{(i,j)} \approx \langle E_{{\rm Sky}\,i}^* E_{{\rm Sky}\,j} \rangle$.

However, for the autocorrelations, the visibilities, $V_{(i,i)}$ are dominated by the positive noise term $\langle N_i^* N_i \rangle$. The amplitudes of the autocorrelation signals are much higher than those of the cross-correlation signals. Therefore, in an eigen-decomposition of the visibility matrix, the eigenvectors are dominated by the noise signals from the autocorrelation, as the sky signals are typically much smaller than the noise.

It is not possible to ignore these auto-correlation signals or simply set them to 0 during the eigenvalue decomposition. To overcome this difficulty, we replace the corresponding terms in the visibility matrix by the following quantity as a proxy for the autocorrelation visibilities.

$$\mathbf{V}_{(i,i)} = \frac{1}{n} \sum_{k,j} {\rm abs} \left[\frac{\mathbf{V}_{(i,k)} \mathbf{V}_{(j,i)}}{\mathbf{V}_{(j,k)}} \right], \quad \forall i \neq j \neq k \quad (7)$$

Here, n is the number of values over which we are doing the sum, i.e. the number of (j, k) pairs. This brings the level of the amplitude of the autocorrelation to the order of the cross-correlation amplitude and we can do a meaningful eigenvalue decomposition.

3.2 DC offset in the visibility

If there is no strong source in the sky then the real and the imaginary parts of the visibility are expected to randomly fluctuate around 0. However, often in radio interferometers, there are some DC offsets in the real and imaginary components of the visibility. The offsets may originate from a variety of systematic effects, and cross-coupling of signals between the antennas is one of them. In the Tianlai data, we see it in multiple baselines as colored horizontal stripes in the waterfall plots of the complex visibility (see Fig. 7).

In the top-left plot of Fig. 8, we show the real and the imaginary parts of the visibility from a transit of Cas A observed by baseline [5H 7V]. On the top-right of the same plot, we show the amplitude of the visibility during a transit of Cas A, which is expected to form a Gaussian profile. However, the plot shows a wavy feature modulating the Gaussian. This pattern comes from the DC offsets in the real and the imaginary parts of the visibility, shown in the left plot. At the beginning of the plot, when there is no source, we can still see some DC signal in the real and the imaginary part and they do not fall on top of each other.

In the bottom panel of Fig. 8, we show the amplitude, as well as the real and the imaginary component of the visibility after subtracting the mean of the nighttime data from both the real and imaginary components of the visibility for each frequency channel and each baseline. The amplitude of the visibility, after DC offset removal, shows a Gaussian peak during the transit of Cyg A, as expected.

The presence of this DC offset may also introduce an error in the eigen-decomposition and it must be removed before running the Sun removal algorithm described below. We subtract the mean value of the real and imaginary parts of the visibility for each night of data. We do not include the daytime data when computing the mean, because it is contaminated by the Sun. However, the DC offset is very stable over each night and from night to night. So we remove the nightly mean from the entire 24 hours of data, including the daytime data. This removal also reduces the night-to-night variation

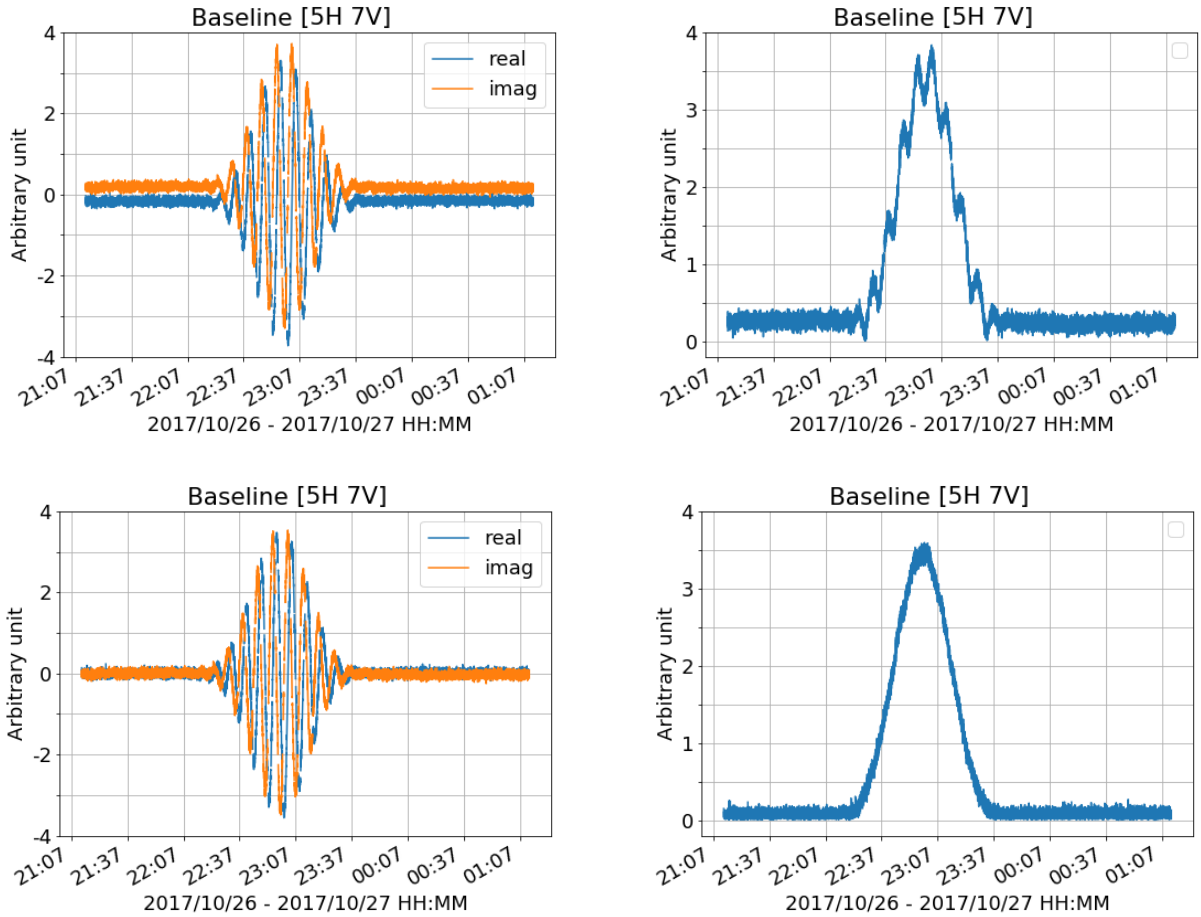

Figure 8. Top-Left: The real and imaginary components of the raw visibility of baseline [5H 7V] during transit of Cas A in October 2017. We can see that there is a small DC offset in both the real and imaginary components. Top-Right: Amplitude of the visibility of baseline [5H 7V] before removing the DC offset. We can see that there is an oscillatory pattern on top of the Gaussian transit peak.Bottom-Left: The real and imaginary components of the raw visibility after removing the offset form each of the components. Bottom-Right: Amplitude of the visibility of baseline [5H 7V] after removing the mean. We can see a perfectly Gaussian transit peak.

both in absolute terms and as a fraction of the remaining signal (Wu et al. 2020).

In Fig 9, we show the waterfall plots of the complex visibilities from four baselines after the nighttime mean removal. We can see the nighttime structures more prominently after the mean subtraction.

Here, we also like to point out that in this work, for simplicity we have only considered the auto-polarization signals. We set the cross polarization signal to 0, making the visibility matrix look like

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}^{(H)} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}^{(V)} \end{bmatrix}, \tag{8}$$

where the superscripts H and V refer to the horizontal or vertical polarization, respectively. $\mathbf{V}$ is the visibility matrix at each time and frequency (t, ν).

3.3 Understanding the eigenvalue decomposition

For removing the solar contamination from the real data, we first remove the DC offset from all the cross-correlation channels. We write the visibility matrix for each frequency and at every time bin in the form shown in Eq. (8). We replace the autocorrelation signals using the formula given in Eq. (7) and decompose the matrix into eigenvalues and eigenvectors as $\mathbf{V} = \boldsymbol{\mathcal{E}}\boldsymbol{\Lambda}\boldsymbol{\mathcal{E}}^{-1}$. At this point it should be noted that the eigenvalue decomposition is invariant under a U(1) transformation, i.e. if we multiply the full eigenvector matrix, $\boldsymbol{\mathcal{E}}$, by a factor of $e^{i\psi}$ for any real ψ, then the corresponding eigenvalue matrix $\boldsymbol{\Lambda}$ will remain invariant. Therefore, without loss of generality, we choose the first component of the eigenvector for each time and frequency component to be real and positive.

Also, in our case the visibility, $\mathbf{V}^{(X)}$ (where $X = \{H, V\}$) is a block diagonal matrix. Therefore, the eigenvalues and the eigenvectors of the matrix will be the eigenvalues and eigenvectors from each of the blocks, i.e. $\mathbf{V}^{(X)} = \boldsymbol{\mathcal{E}}^{(X)}\boldsymbol{\Lambda}^{(X)}\boldsymbol{\mathcal{E}}^{(X)\,-1}$. For each t and ν, $\boldsymbol{\mathcal{E}}$ is a $n \times n$ matrix whose whose i-th column is the complex normalized eigenvector, $\mathcal{E}_i^{(X)}$ of $\mathbf{V}$, and $\boldsymbol{\Lambda}$ is the diagonal matrix whose diagonal elements, $\boldsymbol{\Lambda}_{ii} = \lambda_i$, are the corresponding eigenvalues.

Fig. 10 shows 16 eigenvalues calculated from the horizontal polarization matrix ($\mathbf{V}^{(H)}$) as a function of time. The plot clearly shows that one eigenvalue is much higher then

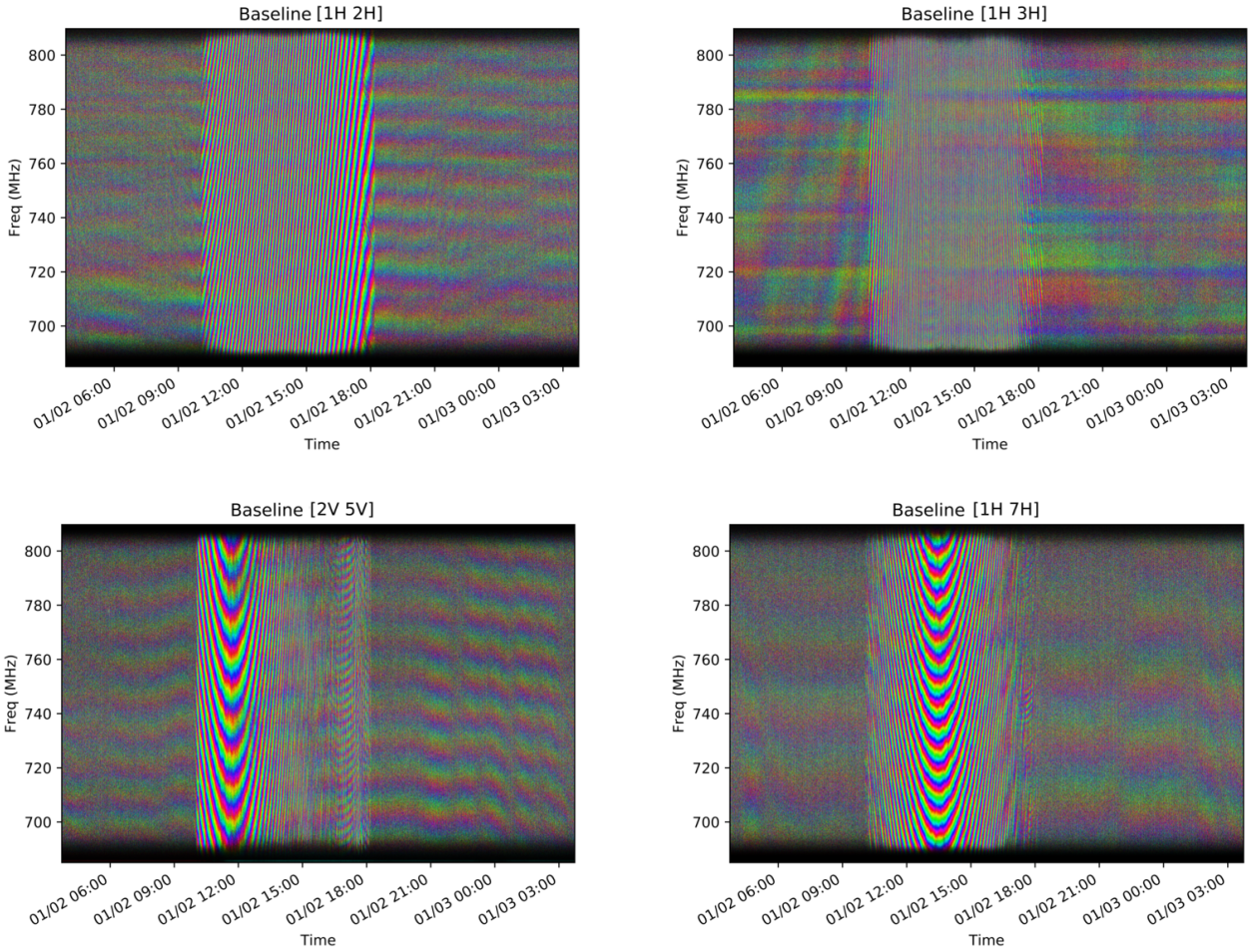

Figure 9. The waterfall plot of the complex visibility (same as Fig. 7) after the nightly mean subtraction. We can see that most of the horizontal stripes, which probably are caused by cross talk, are now gone from the waterfall plot. The structures from the sky are more prominently visible.

the other values during daytime. We can undoubtedly infer that the major contribution to the power in that particular eigenvalue comes from the solar contamination, as the Sun is by far the strongest source in the sky during day time.

Fig. 11 shows the phase of one of the components of the eigenvector that corresponds to the largest eigenvalue: $\boldsymbol{\mathcal{E}}^{(X)}_{(15,16)}$. Eigenvalues are sorted according to the daytime amplitude. The 16th eigenvalue is the largest and we have shown the 15th component of the corresponding eigenvector.

Clear fringes are visible in the daytime data, showing that the daytime signal in the eigenvector is coming from a single strong source. In the nighttime data we can see the phase varies completely randomly, proving the absence of any single strong source at the nighttime data.

In Fig. 12 we show the phase from one of the components of the 4th largest eigenvector, $\boldsymbol{\mathcal{E}}^{(X)}_{(15,13)}$. Unlike Fig. 11, no fringes are visible, indicating that there is no single strong source being detected by that particular baseline and the signal is coming from the background sky. The same thing is true for any of the other smaller eigenvalues. In Fig. 13 we have plotted the phase from one component of the eigenvector corresponding to the second largest eigenvalue. We can find weak fringes during the daytime, indicating that some of the Sun signal has 'leaked' into this eigenvector. Ideally, the second eigenvalue represents the second strongest sources in the sky, and this leakage may be due to the presence of other sources and the background noise. In addition, the re-normalization of the autocorrelation signal using Eq. 7 is another possible cause of this leakage. Finally, it may also be that some of the solar radiation is being reflected from the ground and illuminating the feeds from a different direction from the main Sun signal.

If we plot the phase from any component of the eigenvector corresponding to the third largest eigenvalue, we can still see some fringe pattern in the daytime data. However, these fringes are much weaker in comparison to $\boldsymbol{\mathcal{E}}^{(X)}_{(15,x)}$ showing that the leakage of solar power is mostly restricted to the second largest eigenvalue.

3.4 A first attempt to subtract the solar contamination signal

Because the solar signal is contributing mostly to the largest eigenvalue, as a first step in removing the Sun signal we can

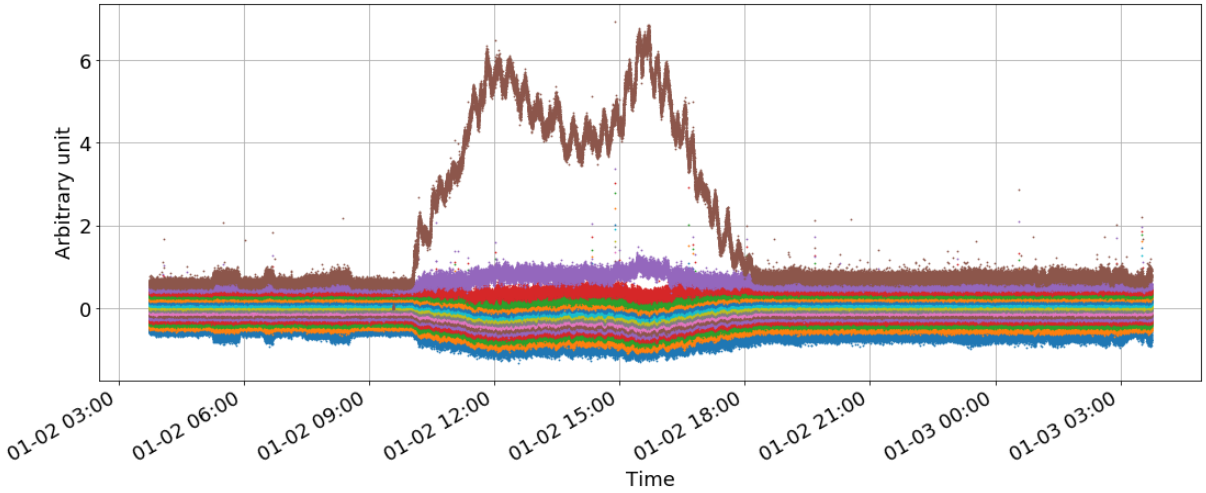

Figure 10. Plot of all the 16 eigenvalues in the eigen-decomposition of the horizontal polarization visibility $\mathbf{V}^{(H)}$. We can see that one of the eigenvalue is much larger than the other eigenvalues during daytime. This particular eigenvalue is coming from the solar contamination of the daytime data. We can see that the other eigenvalues are also affected during daytime. This happens due to the change in the eigenvectors, one of which (the eigenvector corresponding to the largest eigenvalue) is oriented towards the Sun during the daytime. The plot here is shown at the central frequency (747.5 MHz) of the observed Tianlai Dish Array band. All other one dimensional plots also use this frequency.

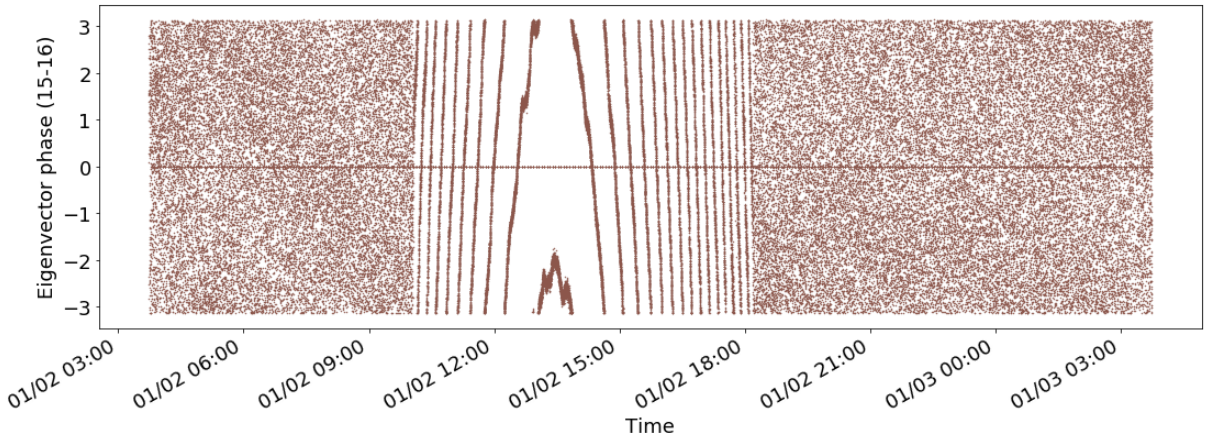

Figure 11. Phase plot of the 15th component of the eigenvector corresponding to the largest eigenvalue of the horizontal polarization visibility, $\mathbf{V}^{(H)}$. We can see the strong fringes during daytime, which confirms that the eigenvalue is coming from a single strong source, which is the Sun. During night, as there is no single strong source, the phase is varying randomly. The horizontal line in the center is caused by the calibration noise source, which is turned on and off periodically.

set the largest eigenvalue during the daytime data to 0, and then reconstruct the visibility. In Fig. 14 we show the largest eigenvalue as a function of time (in blue during the daytime). The red curve shows the value after setting the largest eigenvalue during daytime to be 0. All the other eigenvalues are kept fixed. In Fig. 15, we show the waterfall plot of complex visibilities from 4 baselines that we recover after this step. The plots show that most of the contamination is removed. However, some solar contamination signal is still discernible in the visibility plot. We can see clear, faint fringes for each of the baselines plotted in Fig. 15.

4 IMPROVING THE SUN SUBTRACTION

As we can see from Fig. 15, some of the solar signal still remains in the visibility matrix, mainly the signal that leaked to the second largest or even to the third largest eigenvalue. We attempt to remove this residual signal through the following 2 steps.

4.1 Smoothing the eigenvalues and eigenvectors

The problem with the direct eigenvalue removal method described above is that it is based on the assumption that there is a single source on the sky. This is not true in this case, as the visibility matrix contains signals from other sources as well as instrument noise. In Fig. 16, one component of the

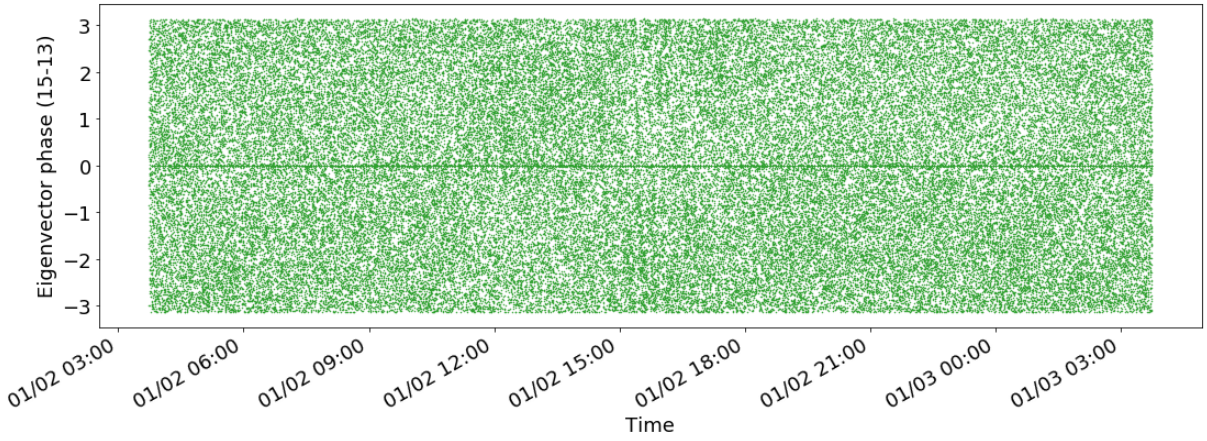

Figure 12. Phase plot of the 15th component of the eigenvector corresponding to the 4th largest eigenvalue in the horizontal polarization visibility $\mathbf{V}^{(H)}$. Here we don't see any fringes, showing that no individual strong source is contributing to this particular eigenvalue.

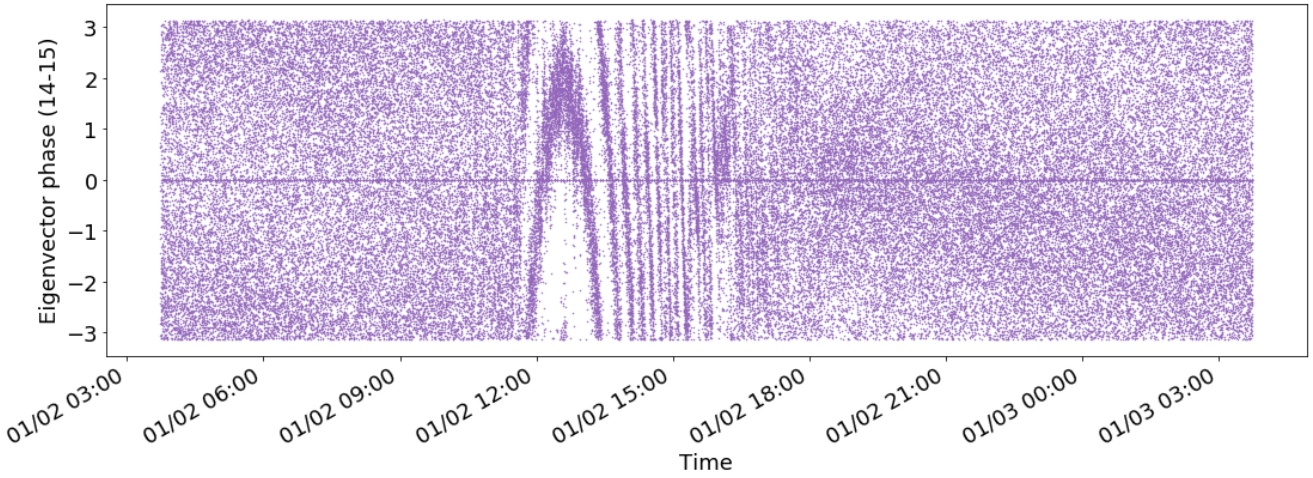

Figure 13. The phase of one component of the eigenvector corresponding to the 2nd largest eigenvalue. We can see weak fringes, indicating that some of the solar contribution is present in the second largest eigenvalue.

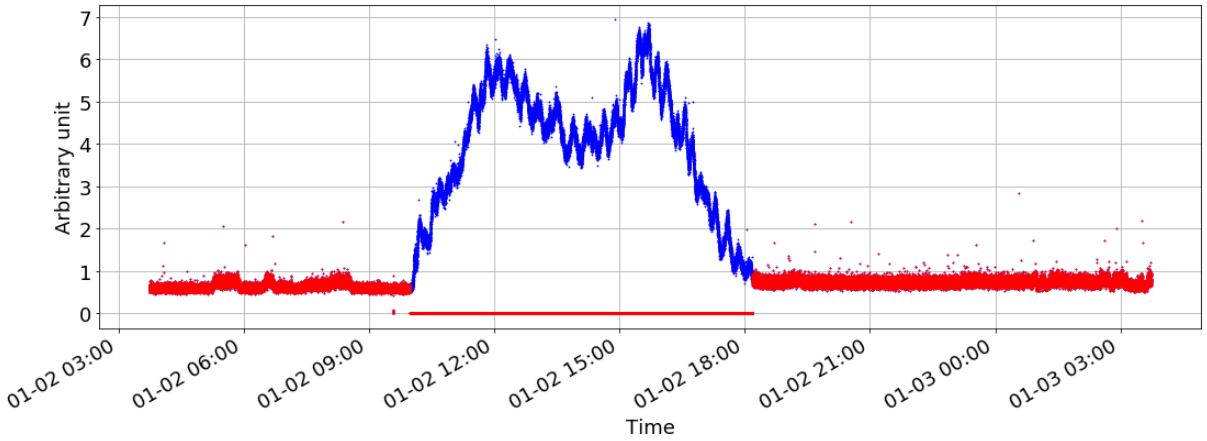

Figure 14. Blue: The original largest eigenvalue from the eigen-decomposition, during daytime. Red: The same largest eigenvalue but with zero value during the daytime. This step simulates removing the Sun signal in Eq. 11, since the largest eigenvector during the daytime points in the direction of the Sun in the eigenspace.

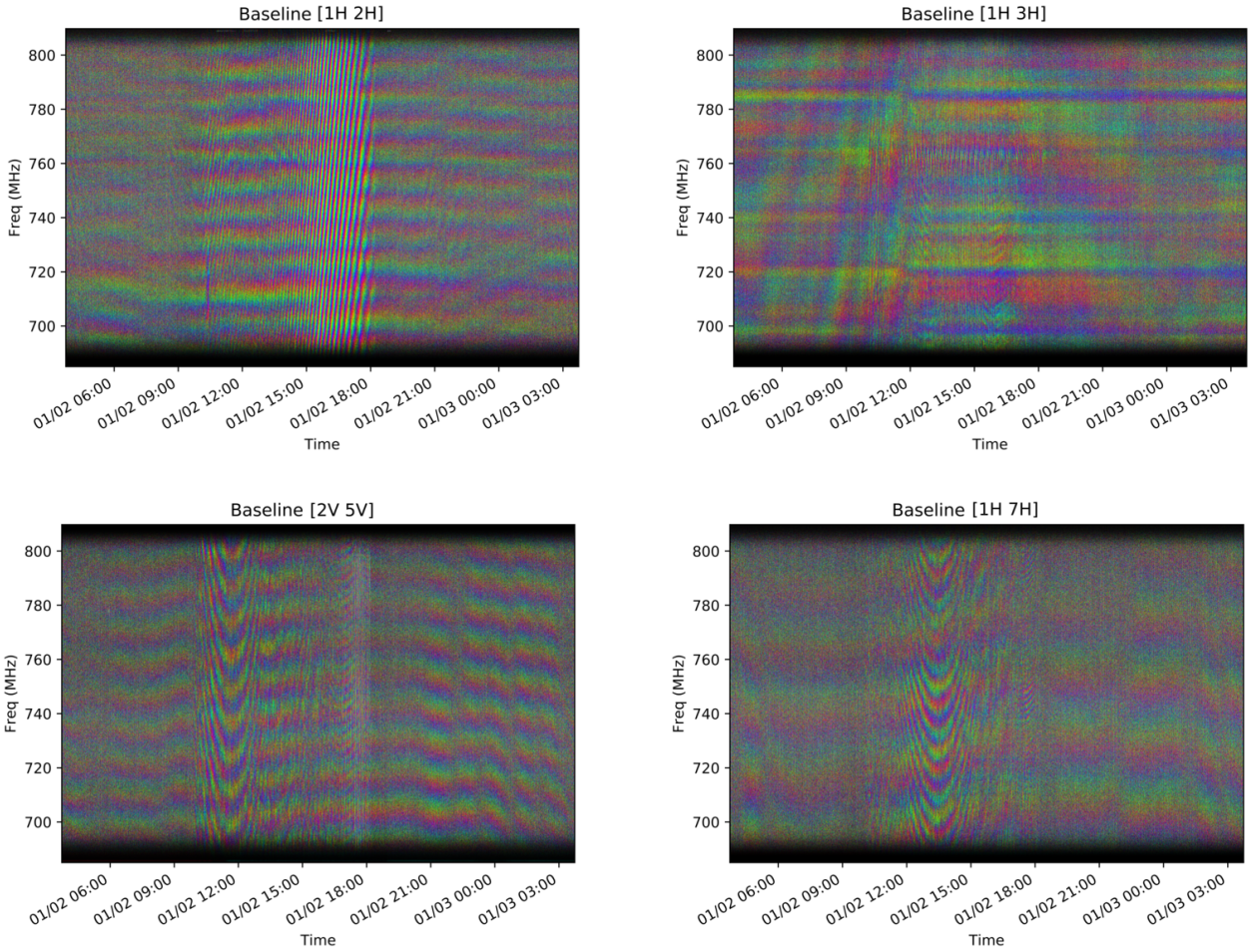

Figure 15. The waterfall plot of the complex visibility after zeroing the largest eigenvalue during daytime for four typical baselines. Nightly mean subtraction has been applied. There is still some residual solar contamination signal, which is causing the weak fringes during the daytime.

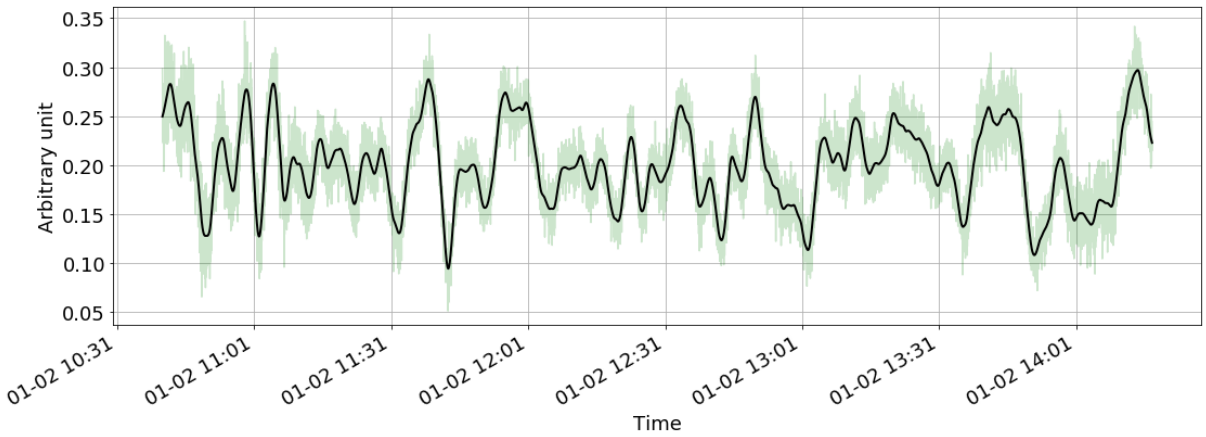

Figure 16. A 4-hour segment of the amplitude of one component of the eigenvector corresponding to the largest eigenvalue is shown in the light green curve. The sampling interval is 1 s. The random fluctuations in the data come from the noise. The black curve shows this component after smoothing the data, as described in section 4.1.

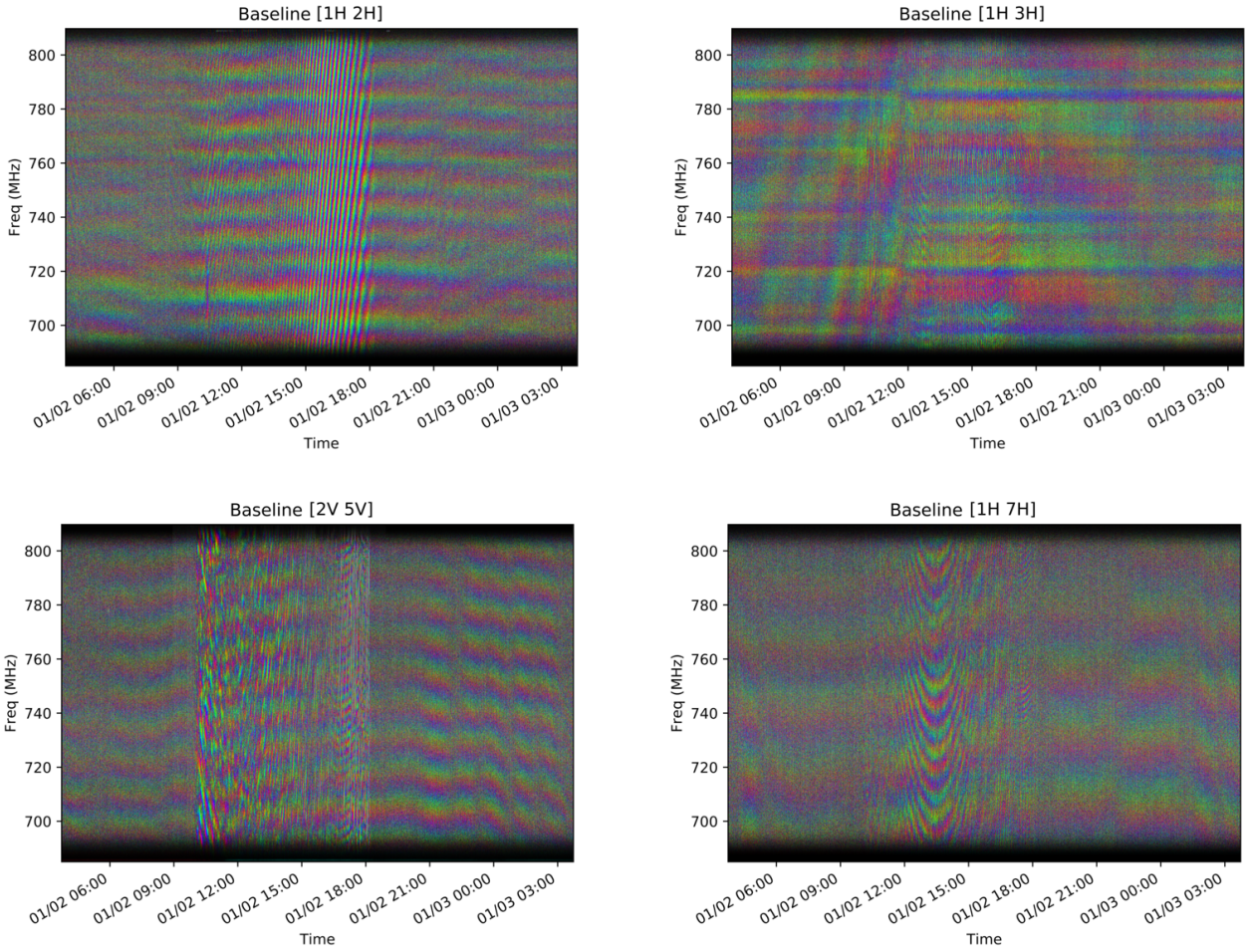

Figure 17. The complex visibility after smoothing the components of the largest eigenvector in spherical coordinates (described in Sec. 4.1, before scaling by a gain factor, described in Sec. 4.2) for four different typical baselines. Nightly mean subtraction has been applied. There is still some signal from the Sun that has not been removed.

eigenvector corresponding to the largest eigenvalue is shown in light green over a period of 4 hours. The data, sampled every second, are noisy. However, as the Sun moves smoothly and the beam is not expected to be structured on small scales, we expect the eigenvector to vary smoothly with time. These fluctuations in the eigenvalue probably come from noise. The long term (minute level and longer) fluctuations originate in the structure of the sidelobes of the telescopes.

As the noise in the visibility matrix may cause the Sun signal to leak from the largest eigenvalue to other eigenvalues, in this section we try to reduce the effect of noise. For doing that, we fit a smooth curve (black line) through the eigenvectors corresponding to the Sun signal. This smoothed signal from the largest eigenvector is then subtracted from the original visibility to construct the Sun-removed visibility.

The cleaning routine can be summarized as follows. The visibility matrix $\mathbf{V}^{(X)}$ is first decomposed into the eigenvalue and the eigenvectors for each time and each frequency bin,

$$\mathbf{V}^{(X)} = \boldsymbol{\mathcal{E}}^{(X)}\boldsymbol{\Lambda}^{(X)}\boldsymbol{\mathcal{E}}^{(X)\ -1}, \qquad X = \{H, V\}. \tag{9}$$

Suppose $\mathcal{E}_S$ is the eigenvector corresponding to the largest eigenvalue, λ_S. As shown in Fig. 16, the direction of $\mathcal{E}_S$ will vary in every second. The n-dimensional complex eigenvectors have only $2n-1$ degrees of freedom as we have already set the first component to be real and positive. As the eigenvectors are unit vectors, the total number of independent components becomes $2n-2$. If we fit a smooth line through each of the $2n-1$ components, then we will overfit and the amplitude of the eigenvectors will not be 1. To keep the eigenvector normalized while doing the fitting, we express each (complex) component of the eigenvector in n-dimensional spherical coordinates and then fit a smooth line through the tangents of the angles in spherical coordinates and convert back to Cartesian space. This gives the black line, shown in Fig. 16. Smoothing in spherical coordinates ensures that the normalization of the eigenvector is preserved during the smoothing procedure. (See Appendix. A for details.)

Let the smoothed components of the largest eigenvectors be $\tilde{\mathcal{E}}_S$. If $\tilde{\lambda}_S$ is the contribution to the visibility from the direction of the eigenvector $\tilde{\mathcal{E}}_S$, then we can write, $\tilde{\lambda}_S = \tilde{\mathcal{E}}_S^T \mathbf{V} \tilde{\mathcal{E}}_S$. If we assume that this smoothed component comes from the Sun signal, then the contribution to the visibility from the Sun is given by

$$\tilde{\mathbf{V}}_S = \tilde{\lambda}_S \left[\tilde{E}_S \otimes \tilde{E}_S\right]. \tag{10}$$

After subtracting the Sun signal, the contribution to the visibility from the rest of the radio sky and noise is given by

$$\mathbf{V}_{\rm Sky} = \mathbf{V} - \tilde{\mathbf{V}}_S \,. \tag{11}$$

In Fig. 17 we show the complex visibility after removing the Sun signal using this particular algorithm. In comparison to the simplest algorithm, of just removing the largest eigenvalue, this new algorithm works better. However, we can see that some of the Sun signal is still present in the visibility.

4.2 Scaling the Sun signal from eigenvalue analysis

The above eigenvalue analysis is based on the assumption that the signal coming from Sun is contained in the largest eigenvalue. As discussed before, this assumption is not correct because of the leakage of power into other eigenvalues.

To overcome this issue, we consider that during the daytime the signals from the sky are much smaller than the solar signal. Therefore, the Sun signal, $\mathbf{V}_S(t,\nu)$, calculated from our analysis should roughly match with the visibility $\mathbf{V}(t,\nu)$ during the daytime as the other signals are negligible in comparison to the $\mathbf{V}_S(t,\nu)$, provided that there are no other strong sources during the day. To do that we introduce a scaling (gain) factor, $g = Ae^{i\phi}$, for each 1000 seconds (about 15 min) of daytime data and minimize

$$\begin{aligned}\chi^2 &= \sum_{t,\nu} [\Re(\mathbf{V} - g\mathbf{V}_S(t,\nu))]^2 \\ &+ \sum_{t,\nu} [\Im(\mathbf{V} - g\mathbf{V}_S(t,\nu))]^2 \,. \end{aligned} \tag{12}$$

Here $\Re(\,)$ and $\Im(\,)$ are the real and imaginary parts of the quantity inside the bracket. We get 36 gain factors (g), calculated from 10 hours (36,000 seconds) of daytime data. In Fig. 18, we show the plot of g over 10 hours of daytime, with circular dots. The smooth lines show the interpolated data. We can see that g varies smoothly throughout the day. The expectation is that the $|g|$ should be very close to 1 and very smooth, and the phase variation should be very small. This is because the Sun and the sky move smoothly through the beams over the day. As long as the Sun signal is strong enough in comparison to the background sky we can expect that the power leakage will vary smoothly and the gain variation should also be smooth. As the signal in the largest eigenvector and leaked power both are coming from Sun we can expect the phase variation to be minimun. Fig. 18 shows that the assumption is a good one in this case. However, near sunrise and sunset the amplitude and the phase change rapidly, possibly because the Sun signal is weaker at those times.

The interpolated g is used as a multiplication factor to determine the solar contribution $g \times \tilde{\mathbf{V}}_S(t,\nu)$, which is finally subtracted from $\mathbf{V}(t,\nu)$. This gives our final Sun-removed signal from the daytime data, i.e.

$$\mathbf{V}_{\rm Sky}(t,\nu) = \mathbf{V} - g_{\rm int}(t,\nu)\tilde{\mathbf{V}}_S(t,\nu) \,. \tag{13}$$

In Fig. 19 we show the complex visibility after the solar contamination removal using Eq. 13. We can see by visual inspection that most of the contamination signal is removed and the fringes from the weaker sources in the background sky are visible. However, for some of the baselines, a significant fraction of the Sun in the form of weak fringes still remains, as we can see in the bottom left plot. In the next section, we will make a first estimate of the performance of our solar signal subtraction and its effect on the signals from the fainter sources.

5 UNDERSTANDING THE EFFICIENCY OF THE ALGORITHM

To test the efficiency of the above algorithm (AlgoSCR), we apply it to simulated data sets. This allows us to check the fraction of the solar contaminant signal that is removed and how much of the sky signal we are erroneously removing by the analysis.

5.1 Construction of simulated data

For constructing a simulated visibility signal $\mathbf{V}_{\rm sim}$, we assume that the electric field at each feed antenna contains contributions from Sun, the sky, and noise. We have assumed that the noise variance is the same throughout the analysis.

The receiver noise is modeled as Gaussian noise in the electric field, $E_{{\rm noise}\,i}$, at the feed antenna. We consider the noise contribution to the electric field to be Gaussian in each sample.

In the Tianlai dish array the integration time in the correlator is 1-sec. The correlator takes in the data that are collected every few microseconds and averages them in an interval of 1-sec. To simulate this process, we add Gaussian random noise in the electric field with a sampling interval of 10 ms. We then calculate the noise contribution to the visibility as $V_{{\rm noise}\,(i,j)} \equiv \langle E^*_{{\rm noise}\,i}\, E_{{\rm noise}\,j}\rangle_{\tau_{\rm int}}$, where $\langle\quad\rangle_{\tau_{\rm int}}$ represents the ensemble average over integration period, $\tau_{\rm int} = 1$ second. Here we have 100 data points for every second on which the average is carried out. This method also ensures that the autocorrelation visibilities follow a χ^2 distribution and the cross-correlation visibilities follow a product normal distribution. The mean and variance of $E_{{\rm noise}\,i}$ are chosen empirically so that the simulated visibility, $\mathbf{V}_{\rm sim}$, matches the observed visibility. The mathematical details on how to calculate the visibilities from artificial point sources in the sky are shown in Appendix B.

To create the simulated Sun signal, we have taken the largest eigenvalue and correponding eigenvector from Eq. A6 from the Tianlai dish array data and treated it as the solar signal. The electric field for the Sun, thus calculated, is added to the simulated noise.

For the simulated artificial sources, we assume the telescope array is pointed at the NCP. The artificial sources are three made-up sources near the NCP. All the artificial sources are visible within the main beam, which is assumed to be Gaussian. Their brightnesses are chosen so that the amplitude of their combined visibility is about 10 times smaller than the noise. (An analysis with different source strengths is presented in the next section.) The artificial source visibilities are frequency- and baseline-dependent, just as visibilities from real sources on the sky. We also assume that the visibilities for the Sun and artificial sources are uncorrelated, i.e., there is no cross-term between the Sun and the artificial sources. This makes the visibilities for the Sun and

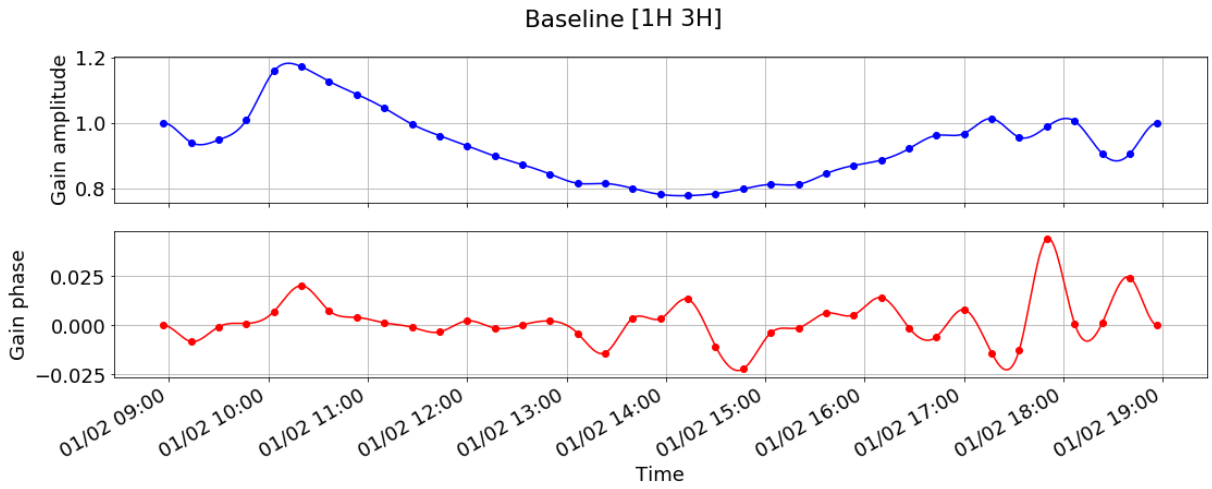

Figure 18. Plot of the gain $g = Ae^{i\phi}$ after χ^2 minimization for 10 hours during the daytime. Blue: Plot of the gain amplitude A. Red: Plot of the gain phase ϕ in radians. The dots represent the points of χ^2 minimization that occur every 1000 seconds. These gain values are extrapolated to the intervening points for a total of 36,000 seconds.

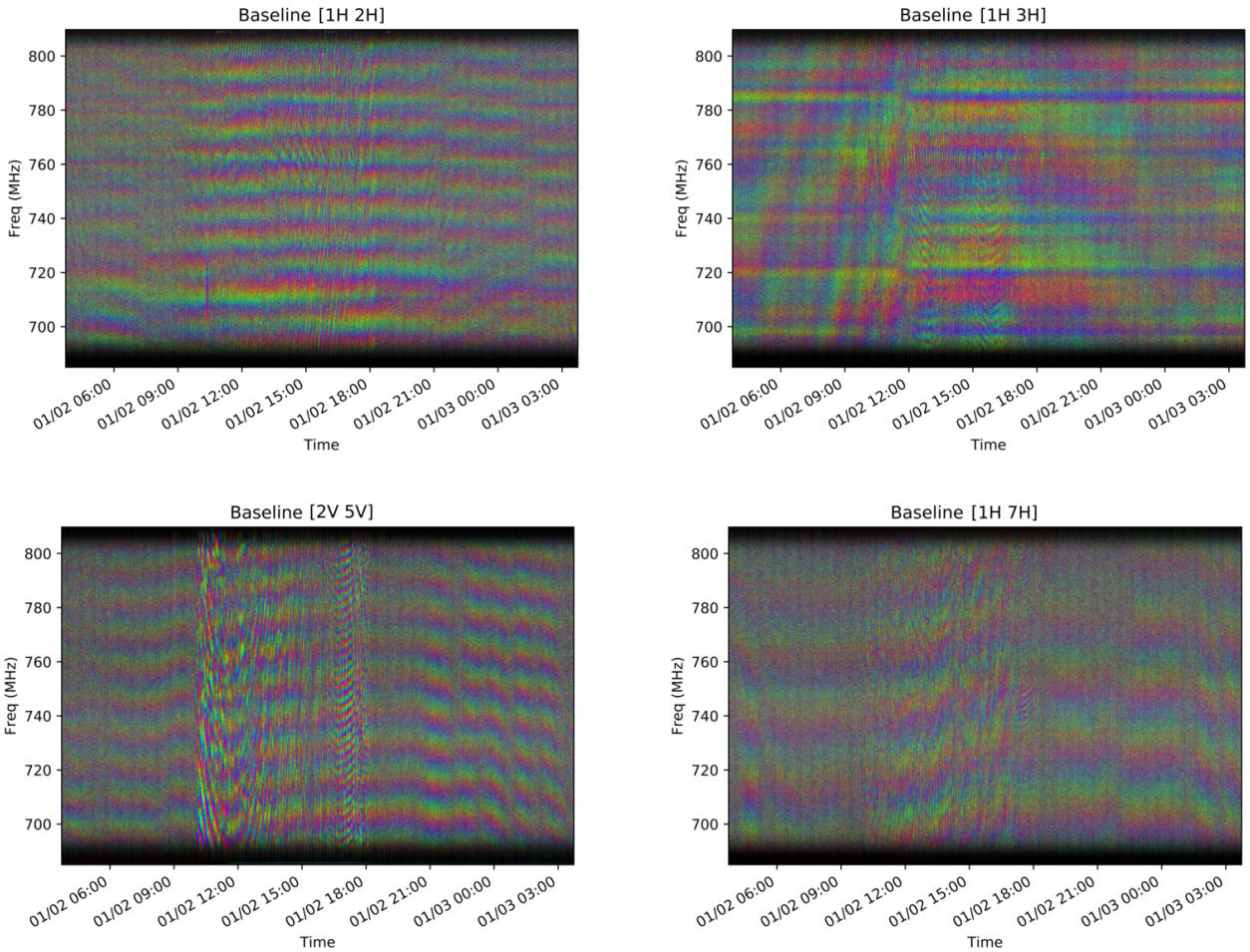

Figure 19. The waterfall plot of the complex visibility after χ^2-optimization scaling the Sun signal by a gain factor (Section 4.2) for four typical baselines. We can see that most of the sun signal contribution has been removed. However, very weak noisy fringes are still visible during daytime, specially in some baselines, such as in [2V 5V], shown in the bottom left panel.

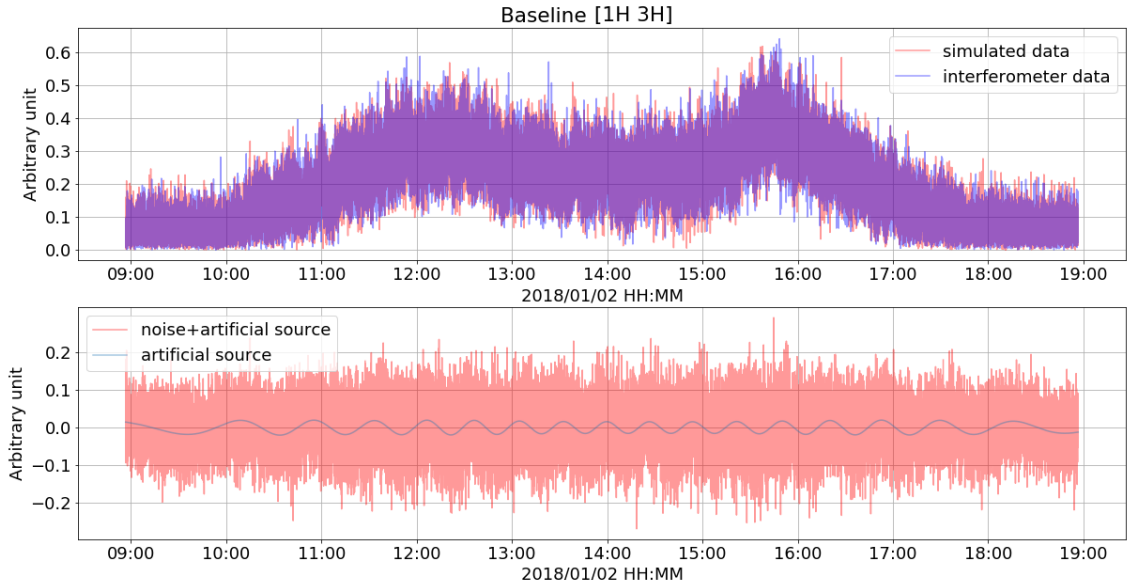

Figure 20. Top: The amplitude of the visibility for baseline [1H 3H]. The actual data from the Tianlai dish array is colored blue, while the simulated data are shown in red. The actual data are very similar to the simulation; regions of overlap appear purple. Bottom: In red, the real part of the noise plus the artificial sources. The blue line shows the real part of the signal from the artificial sources that is added to the data. The imaginary part is similar to the real part and is not shown.

artificial sources additive, as shown in Eq. 14. Please check Appendix B for details.
We can also assume that the visibilities for the Sun and artificial sources are uncorrelated, i.e., there is no cross-term between the Sun and the artificial sources. This makes the visibilities for the Sun and artificial sources additive, as shown in Eq. 14. Please check App. B for details.

$$\mathbf{V}_{\rm sim} = \mathbf{V}_{\rm noise} + \mathbf{V}_{\rm S} + \mathbf{V}_{\rm artificial\ sources} \tag{14}$$

5.2 Results from the simulated data

We generated simulated data as shown in Fig. 20. The top panel of Fig. 20 shows the amplitude of the visibility for baseline [1H 3H] for simulated and Tianlai dish array data: The plot in blue shows the actual complex visibilities from the Tianlai dish array, and the red plot shows the simulated data in our simulation (see Equation 14). The bottom panel shows the real part of the signal from the artificial sources (in blue). The real part of the combined signal (simulated noise and the visibility of the artificial sources that is added to the Sun) is shown in red.

We apply AlgoSCR to the simulated data. The top panel in Fig. 21 shows the real part of the visibility (in red) after applying the Sun removal algorithm, along with the real part of the visibility of the artificial sources (in blue) for baseline 40 [1H 3H]. As the nature of the imaginary part will be similar, we have not explicitly shown it in the plot. In the bottom three panels, the visibilities of the Sun-removed simulated data and artificial sources are binned in 60 second bins. The 60 second binning is used to increase the signal-to-noise ratio. We see that the real and imaginary parts of the Sun-removed signal closely resemble those from the artificial sources and that the phase is not affected.

The ratio of the difference between $V_{\rm sim} - V_{\rm org}$, and the noise standard deviation, σ, is plotted in Fig. 22. We can see that the ratio is within 3. As the injected noise is Gaussian, we can expect that most of the visibility should also fall within 3σ. Therefore, Fig. 22 ensures that the recovery of the signal using AlgoSCR does not introduce additional noise.

The red plot in Fig. 23, shows the difference between the amplitude of the simulated visibility (including the Sun), and the visibility that we are getting after applying AlgoSCR. So this give the contribution from Sun in our simulated data. The blue curve is showing the Sun signal that we introduced for generating the simulated data. We can see that the plots match very well. Top plot is constructed using the data from each second and the bottom plot is after averaging the data over a minute.

In the next set of plots, Fig. 24, we show the complex visibilities for some of the baselines, before and after the solar contamination removal by AlgoSCR. The visibility data show that the artificial sources that we had introduced are clearly visible after the solar signal removal, even though the source strength was much smaller than the Sun signal and the noise. This simulation shows the potential of the Sun removal algorithm.

5.3 Comparing efficiency of the method for different external source strengths

Here we compare the efficiency of AlgoSCR in recovering the artificial sources for different source strengths. For this analysis we use the real daytime visibility data taken by the Tianlai dish array as the base visibility. To this data we add the artificial visibility signal with different source strengths. We

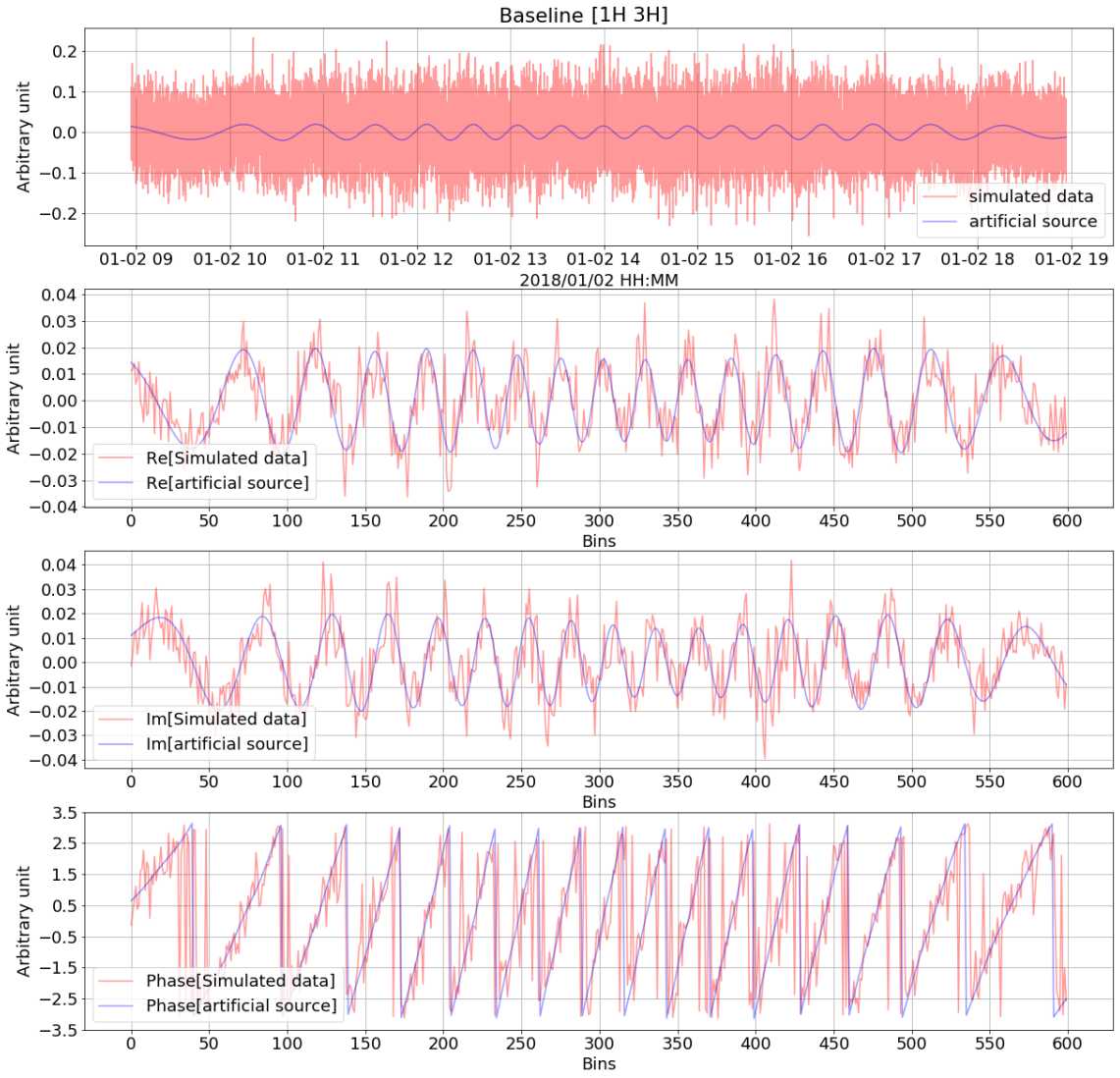

Figure 21. Top: the real part of the Sun-removed visibility from simulated data for baseline [1H 3H]. The signal from the artificial sources is shown in blue. The visibilities are sampled every second. The lower three plots show the real part, imaginary part, and phase of the Sun-removed visibility, after 60-second averaging.

assume that sources are not correlated with the visibility data and the visibilities are additive.

After running AlgoSCR to remove the Sun, we using a χ^2 statistic to compare the signal with the source visibility that was originally inserted. The plot of the reduced χ^2 for baseline [1H 3H] is shown in Fig. 25 against the source strength. Here χ^2 is defined as $\chi^2 = \frac{1}{n_t \sigma^2} \sum_{t,\nu} |\mathbf{V}_{\rm org}(t,\nu) - \mathbf{V}_{\rm sim}(t,\nu)|^2$ during 10 hours of daytime. Here σ^2 is the noise variance and n_t is the number of time-steps, which is the number of degrees of freedom in this case. As we are taking sampling each second for total of 10 hours, the number of degrees of freedom $n_t = 36000$.

We can see the χ^2 value is small for the cases in which the artificial source amplitudes are small compared to the Sun signal amplitude. As the artificial source strengths increase, the fit gets worse. This is because our analysis is based on the assumption that the solar signal is the only dominant signal. As the strength of the artificial sources increases, the assumption slowly breaks down. In such cases, the largest eigenvalue starts to capture signal from the artificial sources. When the artificial sources are larger than the Sun, the largest eigenvalue provides the contribution from the artificial sources and not the Sun. In such cases we are essentially removing the artificial sources and thus the χ^2 grows quadratically.

In Fig. 26 we have plotted the same χ^2, where instead of dividing by σ^2 we have divided by $|V_{org}|$. Here we can see that the χ^2 is lowest when the strength of the artificial source is about 40% of the Sun signal. When the source strength is small, the χ^2 is dominated by the noise and the χ^2 is high. On the other hand, when the strength of the artificial sources is high compared to the Sun contamination as described before, the recovery gets worse.

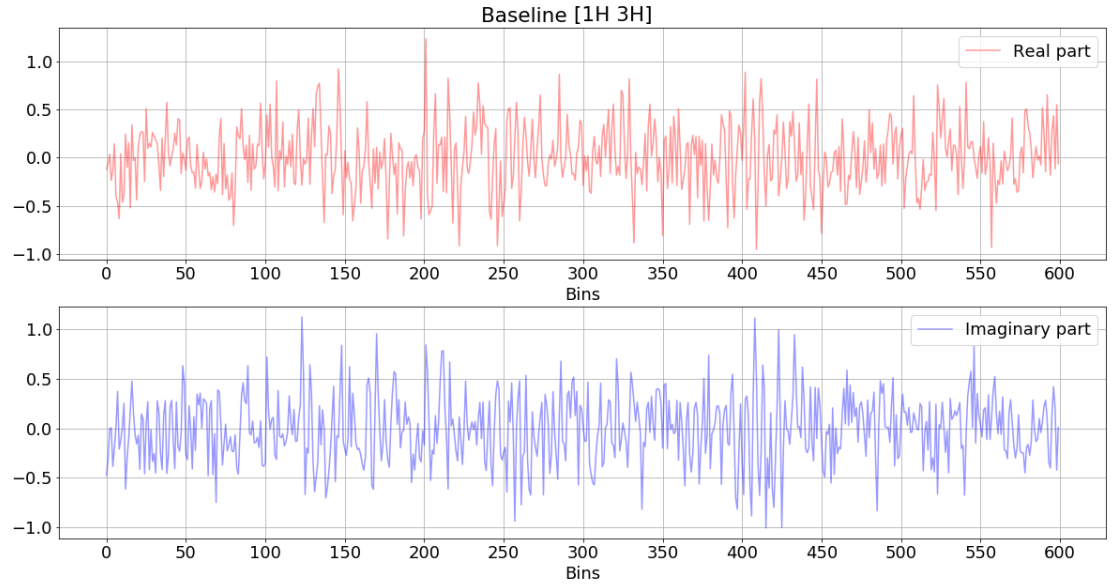

Figure 22. Plot of the real and imaginary part of $(V_{\rm sim} - V_{\rm org})/\sigma$, where σ^2 is the variance of the added noise. We can see that the ratios for both the real and imaginary parts are roughly within 3. As the noise is Gaussian, we can expect that the noise signal should be with 3σ. we conclude that the Sun removal algorithm does not introduce any significant additional noise in this analysis.

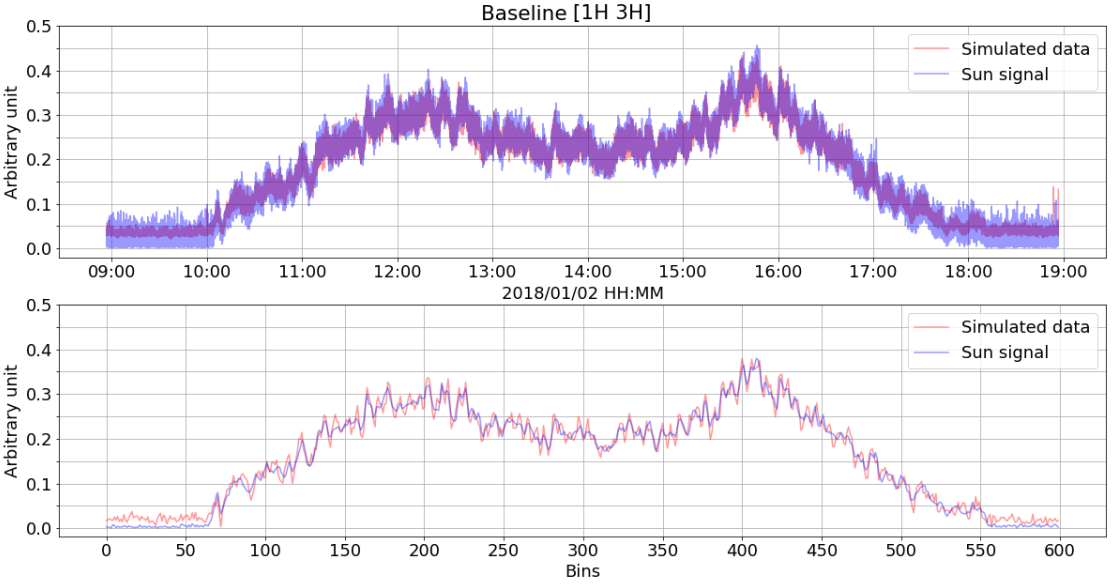

Figure 23. Top: The difference between the simulated visibility (shown in Fig. 20) and Sun-removed visibility (shown in Fig. 21) is shown in red. The signal from the largest eigenvector, which is used as the Sun signal during the daytime is shown in blue. Bottom: Same plot over the same time interval but the data are averaged in 60 second time bins to reduce the noise.

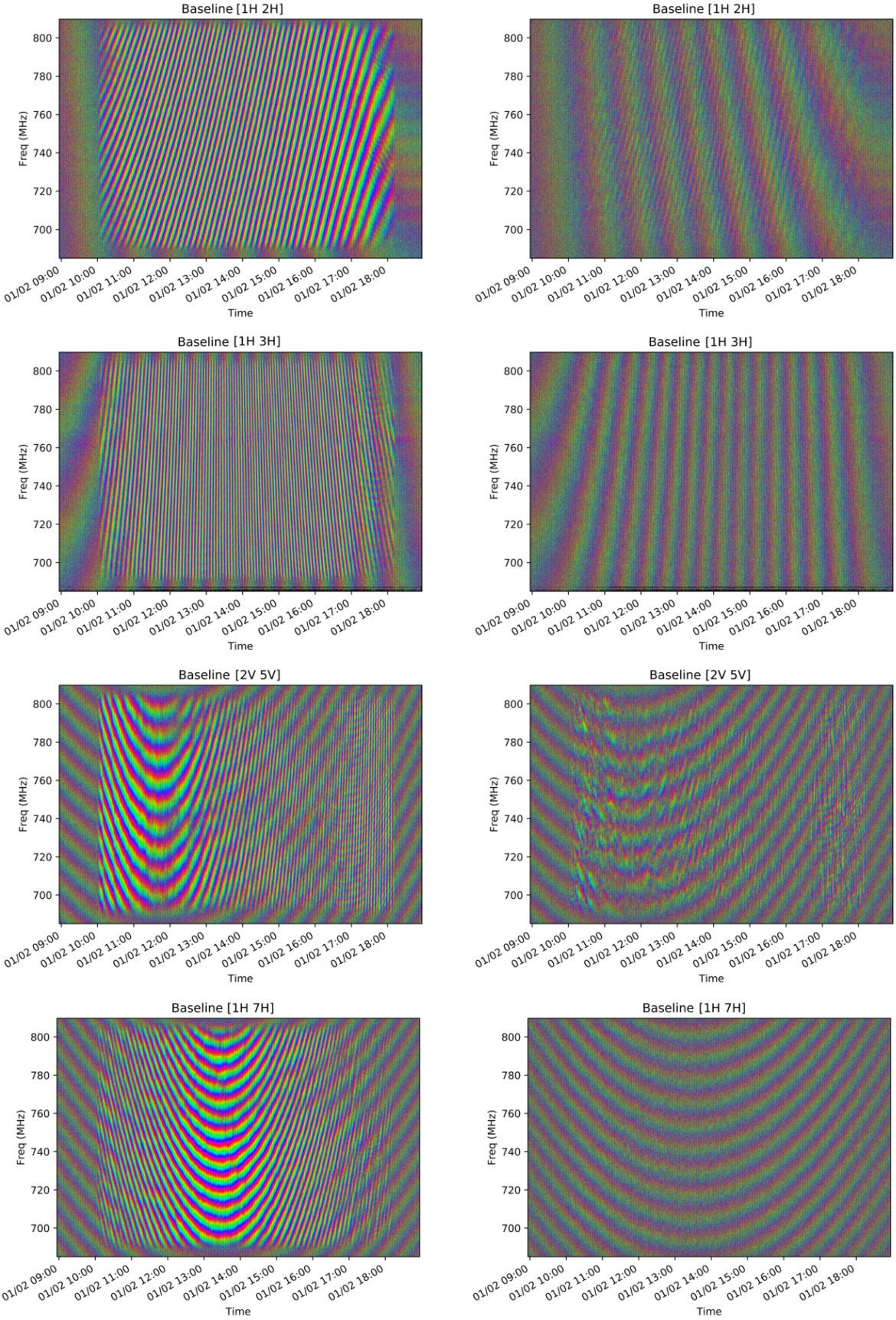

Figure 24. Left column: Simulated daytime visibilities for several baselines. The bright fringes are from the Sun and the weaker fringes are from the artificial sources. Right column: Sun-removed visibility using AlgoSCR. The fringes from the artificial sources are clearly visible. We can see that most of the solar contamination is removed by AlgoSCR.

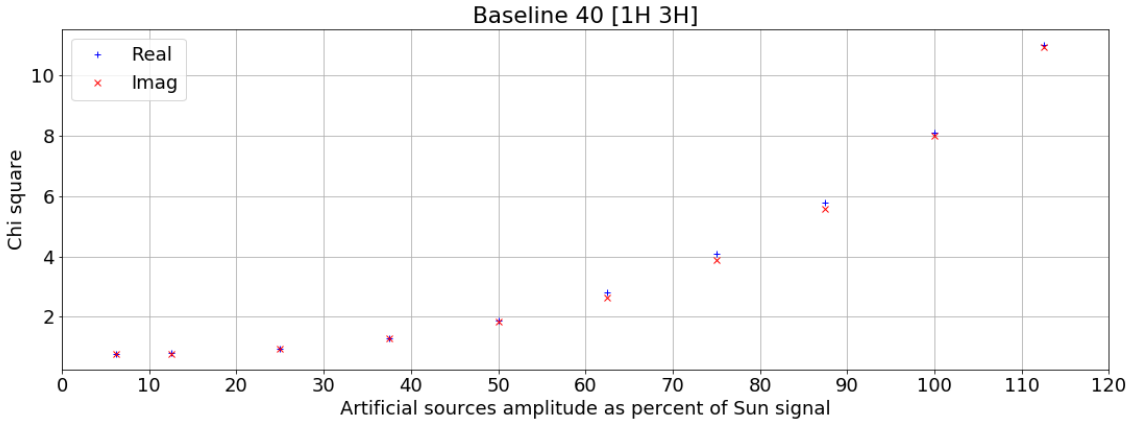

Figure 25. Plots of $\chi^2 = \frac{1}{n_t \sigma^2} \sum_{t,\nu} (V_{\rm org} - V_{\rm sim})^2$ from the real and imaginary parts of the visibility for different artificial source amplitudes. n_t is the number of sample points in the time direction and σ^2 is the noise variance. The amplitude of the original visibility $V_{\rm org}$ and the simulated visibility $V_{\rm sim}$ are shown in Fig. 21. We can see that the χ^2 is increasing as we increase the amplitude of the artificial source.

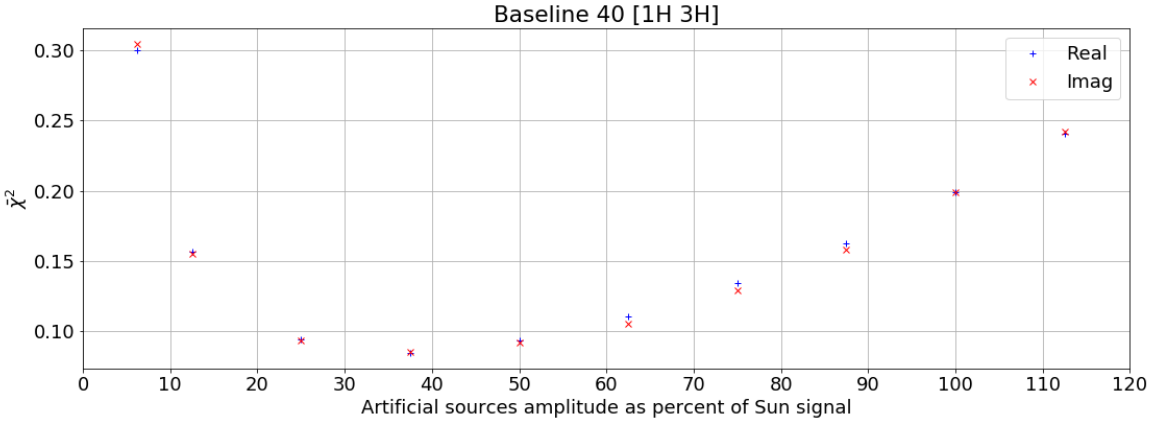

Figure 26. Plots of $\bar{\chi}^2 = \frac{1}{n_t |V_{org}|} \sum_{t,\nu} (V_{\rm org} - V_{\rm sim})^2$ from the real and imaginary parts of the visibility for different artificial source amplitudes. Here, the plots are normalized by $|V_{org}|$ instead of σ^2. We can see that the $\bar{\chi}^2$ is lowest at about 40, indicating that the recovery is best when the amplitude of the source is about 40% of the Sun signal.

6 DISCUSSION

While developing AlgoSCR we explored multiple techniques and came across various issues. Here we discuss some of the points that are important in the context of optimizing AlgoSCR.

In Sec. 4, we address the issue of removing the residual Sun signal after subtraction of the largest eigenvalue. Here we introduce the concept of the multiplication factor g. This procedure raises the question of what will happen if instead of filtering out just the largest eigenvalue, we filter out a smooth component from the two largest eigenvalues. We find that removing the two largest components after smoothing, then the signal from the second largest component, which includes some radio sources, also gets removed, i.e. we will be removing the components from other radio sources and hence the method will not work.

In Sec. 4.2, while choosing the gain values, g, we calculate the gains at intervals of 15 min and then interpolate. We find that the results are fairly insensitive to the choice of time interval (say, 10 min or 30 min), as is expected because the gain varies smoothly throughout the day. However, if we choose a long time interval (several hours) for setting the gains, then we expect the results to worsen as the gain may change significantly in that time. However, we have not simulated these cases.

Another important fact that came up during our analysis is that the Sun removal algorithm works better with more baselines, i.e. if we use all 16 dishes from Tianlai instead of, say, 10 dishes, then the effectiveness of the algorithm increases. Analysis with fewer dishes increases the power leakage to other eigenvalues. The exact reason behind this is not known, but it may happen as more baselines reduce the effect of the noise in the eigen-decomposition.

In addition, if we increase the integration time from 1 sec-

ond to a larger value, the results get worse, which may be due to the fact that the Sun is not a point source. This is different from co-adding the signal from multiple days, which is eventually what the Tianlai array is designed to do. However, we have not tested the algorithm on co-added signals yet.

Our analysis shows that AlgoSCR removed most of the solar contamination during the day. However, it is just a first step. We have not yet tested its effect on map-making and power-spectrum estimation. A critical next step is to make sure that AlgoSCR does not affect the statistics of the maps. This can be checked by comparing the HI power spectra and other statistical quantities from the maps produced using only nighttime data and the maps produced using the full day data after solar contamination removal. Such an analysis requires foreground subtraction and mapmaking and is outside the scope of the present work.

7 CONCLUSION

In this paper, we present a way to separate out the solar contamination from the daytime data observed by an interferometric radio array using eigen-decomposition techniques. The technique is primarily based on the assumption that if the Sun signal is the dominant signal in the sky, along with other weaker sources, and if the signals from the different sources are not correlated, then in the eigen-decomposition of the visibility matrix, the largest eigenvalue is from the strongest source, i.e. the Sun. The eigenvector corresponding to the largest eigenvalue points in the direction of that source in the eigenspace. The technique should filter out this largest eigenvalue while retaining the signals from other sources in the sky.

However, antenna gain fluctuations, noise, sidelobe gain patterns, ground reflection, thermal effects on the instruments and cables, and cross-talk between antennas introduce mixing between the largest eigenvalue and other smaller eigenvalues. For these reasons singling out and removing the Sun signal is not straightforward, and there is some residual contamination from the Sun. Therefore, we apply some novel techniques to remove the leftover Sun signal.

We show that our algorithm is able to remove the solar contamination without removing other, weaker sources in the sky. We have also presented the application of our algorithm to simulated signals, where the background radio sky is known. We show that AlogSCR can recover the background signal pretty well, after removing the solar contamination.

To the best of our knowledge, this is the first published method for removing solar contamination from radio interferometer data. AlgoSCR can contribute to other ongoing and upcoming radio interferometers for solar contamination removal.

8 ACKNOWLEDGEMENT

The authors also wish to thank Richard Shaw, Jeff Peterson and John Marriner for several fruitful discussions about the project during the Tianlai meeting in Guizhou. We wish to thank Kevin Gayley for allowing us to use the beam pattern from his CST simulation. We thank Juyong Zhang and his team for letting us use the beam mapping data from their drone survey.

Work at UW-Madison and Fermilab is partially supported by NSF Award AST-1616554.

This research was performed using the compute resources and assistance of the UW-Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science.

The Tianlai array is operated with the support of NAOC Astronomical Technology Center. The work at NAOC is supported by the Ministry of Science and Technology of China under grants 2018YFE0120800, 2016YFE0100300 and 2012AA121701, the National Natural Science Foundation of China under grants 11633004, 11473044, 11761141012, 11653003, 11773031, Chinese Academy of Science grants QYZDJ-SSW-SLH017, XDA15020200, ZDKYYQ20200008.

This document prepared by the Tianlai Collaboration includes personnel and uses resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract No. DE-AC02-07CH11359.

APPENDIX A: SUMMARY OF ALGOSCR

Here, we review the step-by-step procedure for Sun removal using the algorithm described above.

- For this procedure to work, first we separate the visibility $\mathbf{V}$ into the horizontal and vertical polarizations, $\mathbf{V}^{(H)}$ and $\mathbf{V}^{(V)}$, respectively. If we don't separate the polarizations, the noise and crosstalk in the same dish will give an additional large eigenvalue. The dimension of $\mathbf{V}$ is 32×32, since we have 16 dual-polarization feeds. The dimension of $\mathbf{V}^{(H)}$ and $\mathbf{V}^{(V)}$ will be 16×16.
- Remove the night-time mean from the visibility matrix $\mathbf{V}^{(X)}$: $\mathbf{V}^{(X)} = \mathbf{V}^{(X)} - \langle\mathbf{V}^{(X)}\rangle_{\rm night}$. Here the average is over the time direction for different frequency channels. This will remove the cross-talk between the antennas.
- Replace the auto-correlations by Eq. 7. In practice, if the denominator, $\mathbf{V}^{(X)}_{(i,j)}$, is zero, we replace the term inside the sum by a small number, such as 0.0001.
- Perform an eigen-decomposition of $\mathbf{V}^{(X)}$:

$$\mathbf{V}^{(X)} = \boldsymbol{\mathcal{E}}^{(X)}\boldsymbol{\Lambda}^{(X)}(\boldsymbol{\mathcal{E}}^{(X)})^{-1}. \tag{A1}$$

- For each second of integration time, let the largest (normalized) eigenvector corresponding to the largest eigenvalue, $\lambda^{(X)}_{S\,(t,\nu)}$, be $\mathcal{E}^{(X)}_{S\,(t,\nu)}$. Now $\mathcal{E}^{(X)}_{S\,(t,\nu)}$ is a vector containing $n = 16$

complex numbers. For fitting the smooth line through these vectors we calculate the tangents, $T^{(X)}_{S\,(t,\nu)}$ as:

$$T^{(X)}_{S\,(t,\nu)}(i) = \frac{\|\mathcal{E}^{(X)}_{S\,(t,\nu)}(i)\|}{\sqrt{\sum_{j=i+1}^{n}\left(\|\mathcal{E}^{(X)}_{S\,(t,\nu)}(j)\|\right)^2}} \qquad \forall i \in [1, n-1]. \quad \text{(A2)}$$

- For smoothing the tangents, $T^{(X)}_{S\,(t,\nu)}$, along the time direction, we apply a Butterworth low-pass filter to remove the high frequency signal. For our dataset, the filter order is 2 and the -3 dB cut-off frequency is 0.01 Hz. A 0 phase filtering is done by `scipy`'s `filtfilt` function. Let the filtered (smoothed) tangents be $\tilde{T}^{(X)}_{S\,(t,\nu)}$.
- Convert the eigenvectors back to Cartesian coordinates. For each second of integration time,

$$\|\tilde{\mathcal{E}}^{(X)}_{S\,(t,\nu)}(i)\| = \sin\left(\tan^{-1}(\tilde{T}^{(X)}_{S\,(t,\nu)}(i))\right) \times \prod_{j=1}^{i}\cos\left(\tan^{-1}(\tilde{T}^{(X)}_{S\,(t,\nu)}(j))\right), \qquad \forall i \in [1, n-1] \quad \text{(A3)}$$

and

$$\|\tilde{\mathcal{E}}^{(X)}_{S\,(t,\nu)}(n)\| = \prod_{j=1}^{n}\cos\left(\tan^{-1}(\tilde{T}^{(X)}_{S\,(t,\nu)}(j))\right). \quad \text{(A4)}$$

We then calculate the real and the imaginary parts of the of the eigenvectors

$$\tilde{\mathcal{E}}^{(X)}_{S\,(t,\nu)}(i) = \|\tilde{\mathcal{E}}^{(X)}_{S\,(t,\nu)}(i)\|\left[\cos\left(\theta^{(X)}_{S\,(t,\nu)}(i)\right) + i\sin\left(\theta^{(X)}_{S\,(t,\nu)}(i)\right)\right] \quad \text{(A5)}$$

where $\theta^{(X)}_{S\,(t,\nu)}(i) = \tan^{-1}\left(\Im(\mathcal{E}^{(X)}_{S\,(t,\nu)})/\Re(\mathcal{E}^{(H)}_{S\,(t,\nu)})\right)$.

- The contribution to the visibility from the Sun is given by

$$\mathbf{V}^{(X)}_{S\,(t,\nu)} = \lambda^{(X)}_{S\,(t,\nu)}\left(\tilde{\mathcal{E}}^{(X)}_{S\,(t,\nu)} \otimes \tilde{\mathcal{E}}^{(X)}_{S\,(t,\nu)}\right) \quad \text{(A6)}$$

where $\otimes$ denotes the outer product between eigenvector $\mathcal{E}^{(X)}_S$ and itself.

- After removing the Sun contribution, the sky contribution to the visibility is

$$\begin{aligned}\mathbf{V}^{(X)}_{Sky} &= \mathbf{V}^{(X)} - \mathbf{V}^{(X)}_S \\ &= \boldsymbol{\mathcal{E}}^{(X)}\boldsymbol{\Lambda}^{(X)}(\boldsymbol{\mathcal{E}}^{(X)})^{-1} - \mathbf{V}^{(X)}_S. \end{aligned} \quad \text{(A7)}$$

However, the above steps still leave some Sun signal contamination, as shown in Fig. 15. To better remove this leftover Sun contamination, we multiply the Sun signal in Equation A6 by a complex factor of $g = Ae^{i\phi}$:

$$\mathbf{V}^{(X)}_{Sky} = \mathbf{V}^{(X)} - Ae^{i\phi}\mathbf{V}^{(X)}_S. \quad \text{(A8)}$$

- To find A and ϕ, we divide the 10 hours of daytime data into 36 intervals (1000 s each) and minimize the χ^2 for each interval. Here we define the χ^2 as

$$\chi^2 = \sum_{t,\nu}\left(\Re\left[\mathbf{V}^{(H)} - Ae^{i\phi}\mathbf{V}^{(H)}_{S\,(t,\nu)}\right]\right)^2 + \sum_{t,\nu}\left(\Im\left[\mathbf{V}^{(H)} - Ae^{i\phi}\mathbf{V}^{(H)}_{S\,(t,\nu)}\right]\right)^2 \quad \text{(A9)}$$

The sum is done over all seconds in the chosen interval and frequency 700.625 MHz to 794.375 MHz. We don't sum the frequency channels before 700 MHz and after 800 MHz, because we don't want to include the edge of the band-pass filter. At the end of this process, we have 36 $[A, \phi]$ pairs.

- We have 36 $[A, \phi]$ pairs corresponding to thirty-six 1000 sec intervals in 10 hours of daytime data. We use a cubic spline to interpolate a $[A, \phi]$ pair for each second in 10 hours of daytime data.
- Subtract the corrected Sun signal:

$$\mathbf{V}^{(H)}_{\mathrm{Sky}\,(t,\nu)} = \mathbf{V}^{(H)} - A_{\mathrm{int}\,(t,\nu)}e^{i\phi_{\mathrm{int}\,(t,\nu)}}\mathbf{V}^{(H)}_{\mathrm{S}\,(t,\nu)}. \quad \text{(A10)}$$

This gives us the final solar contamination removed result, and the results are shown in Fig. 18.

APPENDIX B: CALCULATING THE ARTIFICIAL SOURCE VISIBILITIES

The visibilities of the artificial sources come from three made up sources near the NCP. The three simulated sources are randomly chosen to be at (RA, DEC) = (75.75,81.25), (79.5, 80.5) and (245.0,79,75) with constant brightness temperatures across all observed frequencies. We calculated the visibilty for each frequency in Tianlai's 512 frequency bins (equally spaced between 685 MHz and 810 MHz).

Each astronomical source exhibits a linearly varying phase with time and frequency, since the visibility is proportional to $e^{-i\varphi}$, where φ is the fringe phase and is defined as

$$\varphi = 2\pi\nu\tau_g(\nu, t) = \frac{2\pi\nu\mathbf{b}\cdot\mathbf{s}}{c}. \quad \text{(B1)}$$

$\tau_g(\nu, t)$ is the frequency independent geometric delay and is equal to

$$\begin{aligned}\tau_g(\nu,t) \equiv \frac{\mathbf{b}\cdot\mathbf{s}}{c} &= \frac{b_x}{c}\cos\delta\cos H(t) \\ &\quad -\frac{b_y}{c}\cos\delta\sin H(t) + \frac{b_z}{c}\sin\delta, \end{aligned} \quad \text{(B2)}$$

where δ is the source declination, $H(t)$ is the source hour angle as a function of sidereal time, $\mathbf{b} = (b_x, b_y, b_z)$ are the baseline components with units of length in the radial, eastern and northern polar directions and $\mathbf{s}$ is the source vector. c is the speed of light. We can also calculate the fringe rates as follows:

$$\frac{\partial\varphi}{\partial t} = \frac{2\pi\nu}{c}[-b_x\sin H(t) + b_y\cos H(t)]\cos\delta, \quad \text{(B3)}$$

For each second of integration time and each frequency, the visiblity for dish i and j is calculated as follows

$$\mathbf{V}_{(i,j)} = \sum_{k=1}^{n}\mathcal{F}_k A(\mathbf{s})e^{-i\varphi} \quad \text{(B4)}$$

where $\mathcal{F}_k$ is the flux of source k. In our simulations, we used three sources with $\mathcal{F} = (546, 170, 128)$ Jansky. $A(\mathbf{s})$ is the gain of the antenna in the direction of the source vector $\mathbf{s}$. For simplicity, $A(\mathbf{s})$, the simulated main beam gain, is taken as a Gaussian distribution with a standard deviation of 3° (FWHM = 7°), and we assume that all three artificial sources fall within the main beam. Therefore we did not model the beam sidelobe gain. $A(s)$ is also assumed to be independent of frequency. In the real Tianlai Dish beam pattern, the mainbeam (excluding the sidelobe) FWHM is about 5° (Wu

et al. (2020)). The simulated baselines are identical to the real Tianlai dish array, and the procedure for calculating the visibility is repeated for every baseline. The waterfall plots of the simulated visibilities for a few typical baselines are shown in Fig. B1. As expected, longer baselines give higher fringe rates, and for a given baseline, we see a faster fringe rate at lower frequency.

REFERENCES

F. B. Abdalla and S. Rawlings. Probing dark energy with baryonic oscillations and future radio surveys of neutral hydrogen. MNRAS, 360:27–40, June 2005. doi:10.1111/j.1365-2966.2005.08650.x.

Filipe B Abdalla and S Rawlings. Probing dark energy with baryonic oscillations and future radio surveys of neutral hydrogen. Monthly Notices of the Royal Astronomical Society, 360(1): 27–40, 2005.

Zaki S Ali, Aaron R Parsons, Haoxuan Zheng, Jonathan C Pober, Adrian Liu, James E Aguirre, Richard F Bradley, Gianni Bernardi, Chris L Carilli, Carina Cheng, et al. Paper-64 constraints on reionization: The 21cm power spectrum at z= 8.4. arxiv. org, 2015.

CJ Anderson, NJ Luciw, Y-C Li, CY Kuo, J Yadav, KW Masui, TC Chang, X Chen, N Oppermann, YW Liao, et al. Low-amplitude clustering in low-redshift 21-cm intensity maps cross-correlated with 2df galaxy densities. Monthly Notices of the Royal Astronomical Society, 476(3):3382–3392, 2018.

Réza Ansari, Evan J Arena, Kevin Bandura, Philip Bull, Emanuele Castorina, Tzu-Ching Chang, Shi-Fan Chen, Liam Connor, Simon Foreman, Josef Frisch, et al. Inflation and early dark energy with a stage ii hydrogen intensity mapping experiment. arXiv preprint arXiv:1810.09572, 2018.

Kevin Bandura, Graeme E Addison, Mandana Amiri, J Richard Bond, Duncan Campbell-Wilson, Liam Connor, Jean-François Cliche, Greg Davis, Meiling Deng, Nolan Denman, et al. Canadian hydrogen intensity mapping experiment (chime) pathfinder. In Ground-based and Airborne Telescopes V, volume 9145, page 914522. International Society for Optics and Photonics, 2014.

RA Battye, ML Brown, IWA Browne, RJ Davis, P Dewdney, C Dickinson, G Heron, B Maffei, A Pourtsidou, and PN Wilkinson. Bingo: a single dish approach to 21cm intensity mapping. arXiv preprint arXiv:1209.1041, 2012.

Richard A Battye, Rod D Davies, and Jochen Weller. Neutral hydrogen surveys for high-redshift galaxy clusters and protoclusters. Monthly Notices of the Royal Astronomical Society, 355(4):1339–1347, 2004.

Philip Bull, Pedro G Ferreira, Prina Patel, and Mario G Santos. Late-time cosmology with 21 cm intensity mapping experiments. The Astrophysical Journal, 803(1):21, 2015.

Tzu-Ching Chang, Ue-Li Pen, Jeffery B. Peterson, and Patrick McDonald. Baryon acoustic oscillation intensity mapping of dark energy. Phys. Rev. Lett., 100(091303), 2008a.

Tzu-Ching Chang, Ue-Li Pen, Jeffrey B Peterson, and Patrick McDonald. Baryon acoustic oscillation intensity mapping of dark energy. Physical Review Letters, 100(9):091303, 2008b.

Xuelei Chen. Radio detection of dark energy—the tianlai project. Scientia Sinica Physica, Mechanica & Astronomica, 41(12): 1358–1366, 2011.

Xuelei Chen. The tianlai project: a 21cm cosmology experiment. In International Journal of Modern Physics: Conference Series, volume 12, pages 256–263. World Scientific, 2012.

Santanu Das, Christopher J Anderson, Reza Ansari, Jean-Eric Campagne, Daniel Charlet, Xuelei Chen, Zhiping Chen, Aleksander J Cianciara, Pierre Colom, Yanping Cong, et al. Progress in the construction and testing of the tianlai radio interferometers. In Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy IX, volume 10708, page 1070836. International Society for Optics and Photonics, 2018.

David R DeBoer, Aaron R Parsons, James E Aguirre, Paul Alexander, Zaki S Ali, Adam P Beardsley, Gianni Bernardi, Judd D Bowman, Richard F Bradley, Chris L Carilli, et al. Hydrogen epoch of reionization array (hera). Publications of the Astronomical Society of the Pacific, 129(974):045001, 2017.

Clive Dickinson. Bingo-a novel method to detect baos using a total-power radio telescope. arXiv preprint arXiv:1405.7936, 2014.

Michael W Eastwood, Marin M Anderson, Ryan M Monroe, Gregg Hallinan, Benjamin R Barsdell, Stephen A Bourke, MA Clark, Steven W Ellingson, Jayce Dowell, Hugh Garsden, et al. The radio sky at meter wavelengths: m-mode analysis imaging with the ovro-lwa. The Astronomical Journal, 156(1):32, 2018.

Wenkai Hu, Xin Wang, Fengquan Wu, Yougang Wang, Pengjie Zhang, and Xuelei Chen. Forecast for fast: from galaxies survey to intensity mapping. Monthly Notices of the Royal Astronomical Society, 493(4):5854–5870, Mar 2020. ISSN 1365-2966. doi:10.1093/mnras/staa650. URL http://dx.doi.org/10.1093/mnras/staa650.

JiXia Li, ShiFan Zuo, FengQuan Wu, YouGang Wang, JuYong Zhang, ShiJie Sun, YiDong Xu, ZiJie Yu, Reza Ansari, YiChao Li, Albert Stebbins, Peter Timbie, YanPing Cong, JingChao Geng, Jie Hao, QiZhi Huang, JianBin Li, Rui Li, DongHao Liu, YingFeng Liu, Tao Liu, John P. Marriner, ChenHui Niu, Ue-Li Pen, Jeffery B. Peterson, HuLi Shi, Lin Shu, YaFang Song, HaiJun Tian, GuiSong Wang, QunXiong Wang, RongLi Wang, WeiXia Wang, Xin Wang, KaiFeng Yu, Jiao Zhang, BoQin Zhu, JiaLu Zhu, and XueLei Chen. The Tianlai Cylinder Pathfinder array: System functions and basic performance analysis. Science China Physics, Mechanics, and Astronomy, 63(12):129862, September 2020. doi:10.1007/s11433-020-1594-8.

JiXia Li, ShiFan Zuo, FengQuan Wu, YouGang Wang, JuYong Zhang, ShiJie Sun, YiDong Xu, ZiJie Yu, Reza Ansari, YiChao Li, et al. The tianlai cylinder pathfinder array: System functions and basic performance analysis. SCIENCE CHINA Physics, Mechanics & Astronomy, 63(12):1–27, 2020.

Adrian Liu and J. Richard Shaw. Data Analysis for Precision 21 cm Cosmology. arXiv e-prints, art. arXiv:1907.08211, Jul 2019.

Adrian Liu and J Richard Shaw. Data analysis for precision 21 cm cosmology. Publications of the Astronomical Society of the Pacific, 132(1012):062001, 2020.

Y. Mao, M. Tegmark, M. McQuinn, M. Zaldarriaga, and O. Zahn. How accurately can 21cm tomography constrain cosmology? PRD, 78(2):023529, July 2008. doi:10.1103/PhysRevD.78.023529.

K. W. Masui, E. R. Switzer, N. Banavar, K. Bandura, C. Blake, L.-M. Calin, T.-C. Chang, X. Chen, Y.-C. Li, Y.-W. Liao, A. Natarajan, U.-L. Pen, J. B. Peterson, J. R. Shaw, and T. C. Voytek. Measurement of 21 cm brightness fluctuations at z 0.8 in cross-correlation. The Astrophysical Journal Letters, 763(1):L20, 2013. URL http://stacks.iop.org/2041-8205/763/i=1/a=L20.

Kiyoshi Wesley Masui, Patrick McDonald, and Ue-Li Pen. Near-term measurements with 21 cm intensity mapping: Neutral hydrogen fraction and bao at z< 2. Physical Review D, 81(10): 103527, 2010.

M. F. Morales. HI Structure Observations of Reionization and Dark Energy. In R. Minchin and E. Momjian, editors, The Evolution of Galaxies Through the Neutral Hydrogen Window, volume 1035 of American Institute of Physics Conference Series, pages 82–86, August 2008. doi:10.1063/1.2973619.

Laura B Newburgh, Graeme E Addison, Mandana Amiri, Kevin

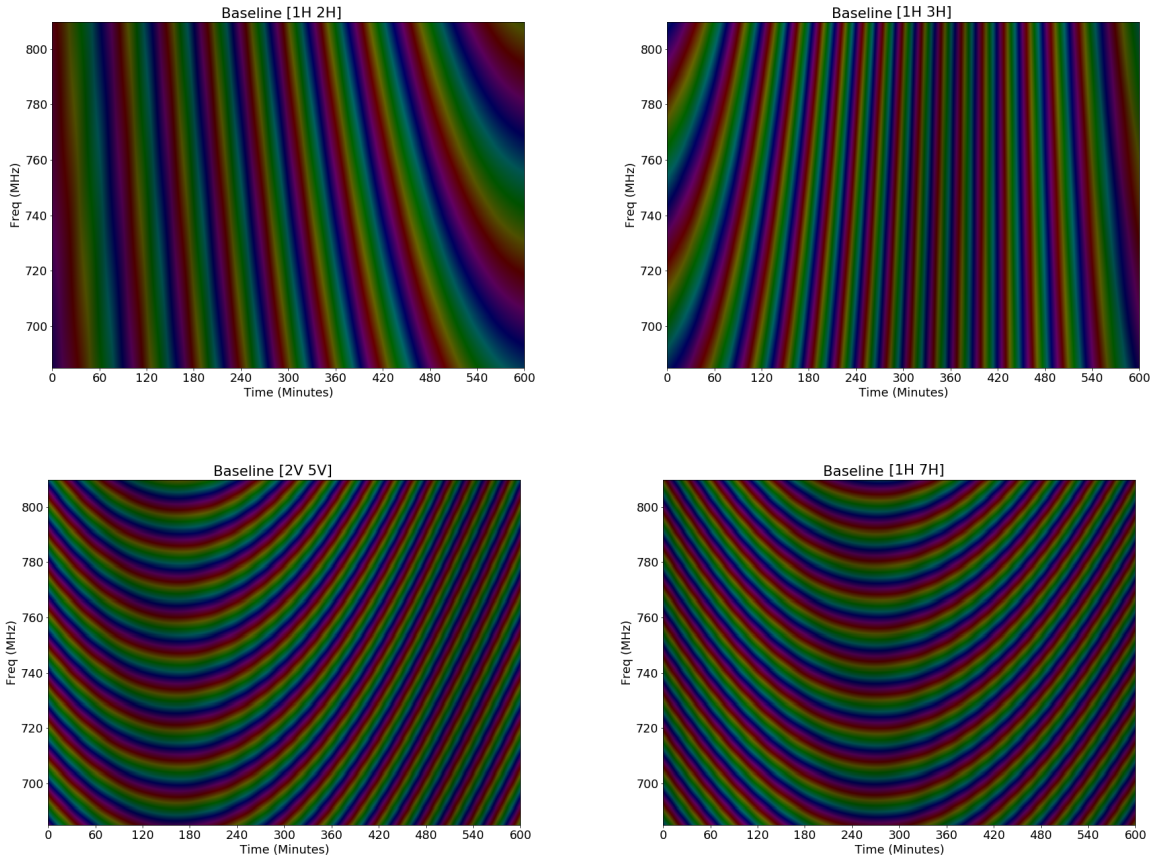

Figure B1. Combined visibilities of three artificial sources for different baselines, shown here for 10 hours. As expected, longer baselines give higher fringe rates, and for a given baseline, we see a faster fringe rate at lower frequency.

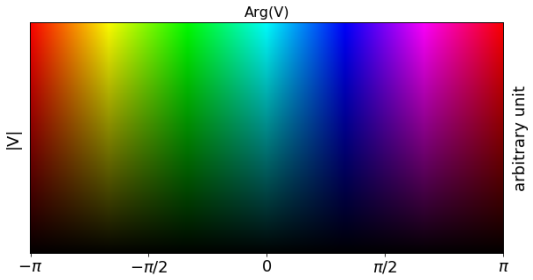

Figure B2. The color palette used to represent complex visibilities in this paper is shown in this plot. The phase of the complex visibility is represented by the color and the amplitude of the visibilities are represented by the hue of the color.

Bandura, J Richard Bond, Liam Connor, Jean-François Cliche, Greg Davis, Meiling Deng, Nolan Denman, et al. Calibrating chime: a new radio interferometer to probe dark energy. In Ground-based and Airborne Telescopes V, volume 9145, page 91454V. International Society for Optics and Photonics, 2014.

L.B. Newburgh et al. HIRAX: A Probe of Dark Energy and Radio Transients. Proc. SPIE Int. Soc. Opt. Eng., 9906:99065X, 2016. doi:10.1117/12.2234286.

J. B. Peterson, K. Bandura, and U. L. Pen. The Hubble Sphere Hydrogen Survey. ArXiv Astrophysics e-prints, June 2006.

Jeffrey B Peterson, Kevin Bandura, and Ue Li Pen. The hubble sphere hydrogen survey. arXiv preprint astro-ph/0606104, 2006.

J. Richard Shaw, Kris Sigurdson, Ue-Li Pen, Albert Stebbins, and Michael Sitwell. All-Sky Interferometry with Spherical Harmonic Transit Telescopes. Astrophys. J., 781:57, 2014. doi:10.1088/0004-637X/781/2/57.

Anže Slosar et al. Packed ultra-wideband mapping array (puma): A radio telescope for cosmology and transients. Technical report, Brookhaven National Lab.(BNL), Upton, NY (United States); SLAC National . . . , 2019.

Steven John Tingay, Robert Goeke, Judd D Bowman, David Emrich, Stephen M Ord, Daniel A Mitchell, Miguel F Morales, Tom Booler, Brian Crosse, Randall B Wayth, et al. The murchison widefield array: The square kilometre array precursor at low radio frequencies. Publications of the Astronomical

Society of Australia, 30, 2013.

Fengquan Wu, Jixia Li, Shifan Zuo, Xuelei Chen, Santanu Das, John P Marriner, Trevor M Oxholm, Anh Phan, Albert Stebbins, Peter T Timbie, et al. The tianlai dish pathfinder array: design, operation and performance of a prototype transit radio interferometer. arXiv preprint arXiv:2011.05946, 2020.

CA Wuensche, BINGO Collaboration, et al. The bingo telescope: a new instrument exploring the new 21 cm cosmology window. In Journal of Physics: Conference Series, volume 1269, page 012002. IOP Publishing, 2019.

Yidong Xu, Xin Wang, and Xuelei Chen. Forecasts on the dark energy and primordial non-gaussianity observations with the tianlai cylinder array. The Astrophysical Journal, 798(1):40, 2014.

Jiao Zhang, Reza Ansari, Xuelei Chen, Jean-Eric Campagne, Christophe Magneville, and Fengquan Wu. Sky reconstruction from transit visibilities: Paon-4 and tianlai dish array. Monthly Notices of the Royal Astronomical Society, 461(2):1950–1966, 2016.