The Astronomical Journal, 158:84 (20pp), 2019 August https://doi.org/10.3847/1538-3881/ab2629

The 21 cm Power Spectrum from the Cosmic Dawn: First Results from the OVRO-LWA

Michael W. Eastwood[1], Marin M. Anderson[1], Ryan M. Monroe[2], Gregg Hallinan[1], Morgan Catha[3], Jayce Dowell[4], Hugh Garsden[5], Lincoln J. Greenhill[5], Brian C. Hicks[6], Jonathon Kocz[1], Danny C. Price[7,8,5], Frank K. Schinzel[9], Harish Vedantham[1,10], and Yuankun Wang[1,11]
[1] Department of Astronomy, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA; mweastwood@astro.caltech.edu
[2] Department of Electrical Engineering, California Institute of Technology, 1200 East California Boulavard, Pasadena, CA 91125, USA
[3] California Institute of Technology, Owens Valley Radio Observatory, Big Pine, CA 93513, USA
[4] Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87131, USA
[5] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
[6] Naval Research Laboratory, Code 7213, Washington, DC 20375, USA
[7] Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Hawthorn, VIC 3122, Australia
[8] Department of Astronomy, University of California Berkeley, Berkeley, CA 94720, USA
[9] National Radio Astronomy Observatory, P.O. Box O, Socorro, NM 87801, USA
[10] Netherlands Institute for Radio Astronomy (ASTRON), Oude Hoogeveensedijk 4, 7991PD, Dwingeloo, The Netherlands
[11] Virtualitics, 225 S Lake Avenue, Pasadena, CA 91101, USA

Received 2018 November 4; revised 2019 May 20; accepted 2019 May 30; published 2019 July 29

Abstract

The 21 cm transition of neutral hydrogen is opening an observational window into the Cosmic Dawn of the universe—the epoch of first star formation. We use 28 hr of data from the Owens Valley Radio Observatory Long Wavelength Array to place upper limits on the spatial power spectrum of 21 cm emission at $z \approx 18.4$ ($\Delta_{21} \lesssim 10^4$ mK), and within the absorption feature reported by the EDGES experiment. In the process we demonstrate the first application of the double Karhunen–Loève transform for foreground filtering, and diagnose the systematic errors that are currently limiting the measurement. We also provide an updated model for the angular power spectrum of low-frequency foreground emission measured from the northern hemisphere, which can be used to refine sensitivity forecasts for next-generation experiments.

Key words: cosmology: observations – dark ages, reionization, first stars

1. Introduction

The Cosmic Dawn of star formation is one of the final unexplored epochs of the universe. During this time (approximately $25 \gtrsim z \gtrsim 15$) the first generation of stars and galaxies formed and brought an end to the Dark Ages. Lyα emission from this early star formation couples the excitation temperature of the 21 cm hyperfine structure transition (i.e., the spin temperature) to the local gas temperature of the intergalactic medium (IGM; Wouthuysen 1952; Field 1958). This allows the highly redshifted 21 cm transition to be used as a probe of the density, temperature, and ionization state of the IGM (e.g., Furlanetto et al. 2006; Pritchard & Loeb 2012).

The first possible detection of high-redshift ($z \sim 17$) atomic hydrogen in the globally averaged sky temperature was recently reported by the Experiment to Detect the Global EoR Signature (EDGES) experiment at 78 MHz (Bowman et al. 2018). This measurement was remarkable for its extreme amplitude (∼500 mK), and unusual width and shape. A plethora of new ideas have been proposed to explain the amplitude of the absorption trough. These new theories generally fall into two categories: those that invoke new physics to cool the IGM at a rate faster than pure adiabatic cooling would otherwise allow (e.g., Barkana 2018; Fialkov et al. 2018), and those that posit a new radio background originating from $z \gtrsim 20$ (e.g., Dowell & Taylor 2018; Ewall-Wice et al. 2018). Distinguishing between these alternatives and confirming the existing measurement (e.g., some concerns about the detection have been raised by Hills et al. 2018) now motivate a number of experiments.

The Large-Aperture Experiment to Detect the Dark Ages (LEDA; Price et al. 2018) and SARAS 2 (Singh et al. 2018) experiments are in the process of attempting to directly confirm the EDGES detection in the global sky temperature using radiometric dipole antennas. Notably, each of these experiments employs materially different antenna designs, which will, in principle, help address concerns regarding the role of the antenna beam and its potential to introduce spectral structure into the measurement. As noted by Bowman et al. (2018), these independent measurements with independent processing pipelines will be an important verification of an exceptionally difficult measurement. The primary observing challenge faced by global-detection experiments—such as EDGES, LEDA, and SARAS 2—is controlling systematic errors introduced by foreground radio emission, instrumental effects, and the interaction between them. Although these experiments calibrate their antennas and electronics with great care, they must rely on external models of the foreground radio emission to model many types of systematic errors.

In contrast, interferometers generally have the ability to self-calibrate and build self-consistent models for the sky emission (Readhead & Wilkinson 1978), even with the extremely wide fields of view common for low-frequency interferometers (e.g., Eastwood et al. 2018). However, interferometers are generally not used to measure the globally averaged sky brightness (Venumadhav et al. 2016), but instead measure the three-dimensional spatial power spectrum of the 21 cm brightness temperature fluctuations. The global average and the spatial power spectrum are both statistics of the same field and therefore a measurement of the spatial power spectrum can also provide evidence to support or reject a putative detection in the global average. A detection of the spatial power spectrum will—with some modeling—provide independent constraints on the temperature of the IGM and the timing of early star and galaxy formation (e.g., Greig & Mesinger 2017). The spatial power spectrum carries additional information about the scale of the brightness temperature fluctuations, which may be used to constrain, for example, the amplitude of Lyman–Werner feedback (Fialkov et al. 2013) and the spectral hardness of early X-ray sources (Fialkov et al. 2014) that heat the IGM.

At lower redshifts corresponding to the epoch of reionization (EoR), constraints on the 21 cm spatial power spectrum have been published by the Precision Array for Probing the Epoch of Reionization (PAPER) experiment (Ali et al. 2015), Low-frequency Array (LOFAR; Patil et al. 2017), the Murchison Widefield Array (MWA; Beardsley et al. 2016), and the Giant Metrewave Radio Telescope (GMRT; Paciga et al. 2013). At redshifts corresponding to the Cosmic Dawn, this measurement has been attempted by Ewall-Wice et al. (2016) using 6 hr of data from the MWA, and Gehlot et al. (2018) using 14 hr of data from the LOFAR-Low Band Antenna array. The under-construction Hydrogen Epoch of Reionization Array (HERA) experiment will aim to place the most sensitive limits to date on the 21 cm brightness temperature spatial power spectrum from both the EoR and Cosmic Dawn due to its large collecting area and design lessons inherited from the PAPER experiment (DeBoer et al. 2017). Similarly, the Square Kilometre Array (SKA) will build on the development of its pathfinder arrays to measure the spatial power spectrum and image the three-dimensional structure of the universe through the 21 cm transition (Mellema et al. 2013; Koopmans et al. 2015).

In this paper we present the first attempted measurement of the spatial power spectrum of 21 cm brightness temperature fluctuations with the Owens Valley Radio Observatory Long Wavelength Array (OVRO-LWA). In the process, we model, derive, and (where appropriate) measure the contribution of thermal noise, foreground emission, and the 21 cm signal to the full covariance matrix of the data. This is possible due to the application of m-mode analysis (Shaw et al. 2014, 2015), which introduces sparsity into the covariance matrices without which it would not be possible to store the full covariance matrix of the data. We also analyze the systematic sources of error that currently limit the measurement through simulation, which allows us to derive quantitative requirements for instrumental calibration errors.

In Section 2 we describe the observations, the calibration strategy, and point-source removal routines used in this work. In Section 3 we describe the m-mode analysis formalism and a new strategy for compressing the representation of the transfer matrix. In Section 4 we derive, model, and measure the contribution of noise, foreground emission, and the cosmological 21 cm signal to the full covariance matrix of the measured data. These covariance matrices are applied to filter the foreground emission in Section 5, where we also build physical intuition for the action of the foreground filters derived by Shaw et al. (2014, 2015). These foreground filters are applied to 28 hr of data from the OVRO-LWA to estimate the 21 cm power spectrum in Section 6, where we also analyze the limiting systematic errors in our measurement. Finally, in Section 7 we present our conclusions. Unless stated otherwise, we adopt the set of cosmological parameters measured by Planck Collaboration et al. (2016).

2. Observations

We collected 28 hr of continuous data using the OVRO-LWA beginning at 2017 February 17 12:00:00 UTC. The OVRO-LWA is a low-frequency radio interferometer with a bandpass covering 27–85 MHz ($50 \gtrsim z \gtrsim 16$), and is currently composed of 288 dual-polarization dipole antennas (G. Hallinan et al. 2019, in preparation). A total of 251 of these antennas are arranged within a dense 200 m diameter core in a configuration optimized for sidelobe levels in snapshot images; 32 additional expansion antennas are placed outside of the core, expanding the maximum baseline length to 1.5 km. The remaining five antennas are equipped with radiometric front-ends for total power measurements of the sky as part of LEDA (Price et al. 2018). The LEDA correlator (Kocz et al. 2015) serves as the back-end for the OVRO-LWA, and cross-correlates 512 inputs with 58 MHz instantaneous bandwidth. In this configuration the OVRO-LWA performs full cross-correlation of 256 antennas (512 signal paths), and 32 antennas (64 signal paths) are unused. We selected the correlator's integration time to be 13 s as this evenly divides the sidereal day to within 0.1 s. In snapshot images, the OVRO-LWA can capture the entire visible hemisphere at 10′ resolution (e.g., Anderson et al. 2018), and this same data set was used to generate maps of the sky north of $\delta = -30°$ (Eastwood et al. 2018).

At low radio frequencies, propagation effects through the ionosphere are important. During this observing period, however, geomagnetic and ionospheric conditions were mild. At 73 MHz, bright point sources were observed to refract by up to 4′ on 10 minutes timescales, and the apparent flux of point sources varied by up to 10% on 13 s timescales (averaging over 24 kHz bandwidth) due to ionospheric conditions.

In this work we selected data from an instrumental subband centered at 73.152 MHz with 2.6 MHz bandwidth ($z = 18.4$, $\Delta z = 0.8$). This subband is contained within the absorption feature observed by Bowman et al. (2018), and contains the 73.0–74.6 MHz band allocated for radio astronomy in the United States. There is additionally a gap in television broadcasting between 72 MHz (the upper edge of channel 4) and 76 MHz (the lower edge of channel 5) that this observing band takes advantage of. Additionally, in previous work we published an updated low-frequency sky map at 73.152 MHz (Eastwood et al. 2018), which is available online at the Legacy Archive for Microwave Background Data Analysis.

When measuring the power spectrum of 21 cm fluctuations, it is common to make the implicit assumption that the 21 cm power spectrum is not evolving along the line-of-sight direction (see the Appendix). Greig & Mesinger (2018) simulated this effect and found that for volumes of equal comoving radial distance, this light-cone effect is more severe during the Cosmic Dawn than during the EoR. Near $z \sim 18$ and a volume with $\Delta z \sim 3$, the recovered spatial power spectrum is suppressed by a factor $\lesssim 2$. While the light-cone effect can limit the usable bandwidth for estimating the 21 cm spatial power spectrum, we conclude that 2.6 MHz of bandwidth is permissible for this initial analysis. Future studies of 21 cm fluctuations of the Cosmic Dawn, however, should instead consider estimating the multi-frequency angular power spectrum (Datta et al. 2007), which is a statistic that is less common in the literature, but can be measured without assuming that the statistics of the fluctuations are not evolving along the line of sight.

A summary of the analysis steps performed in this work—including the instrumental calibration and 21 cm power spectrum reduction—can be seen in Figure 1. In particular, a gain calibration was derived from a 45 minute track of data beginning at 2017 February 17 17:46:28 during which the two brightest point sources in the northern hemisphere (Cyg A and Cas A) are near the meridian. The sky model is initially

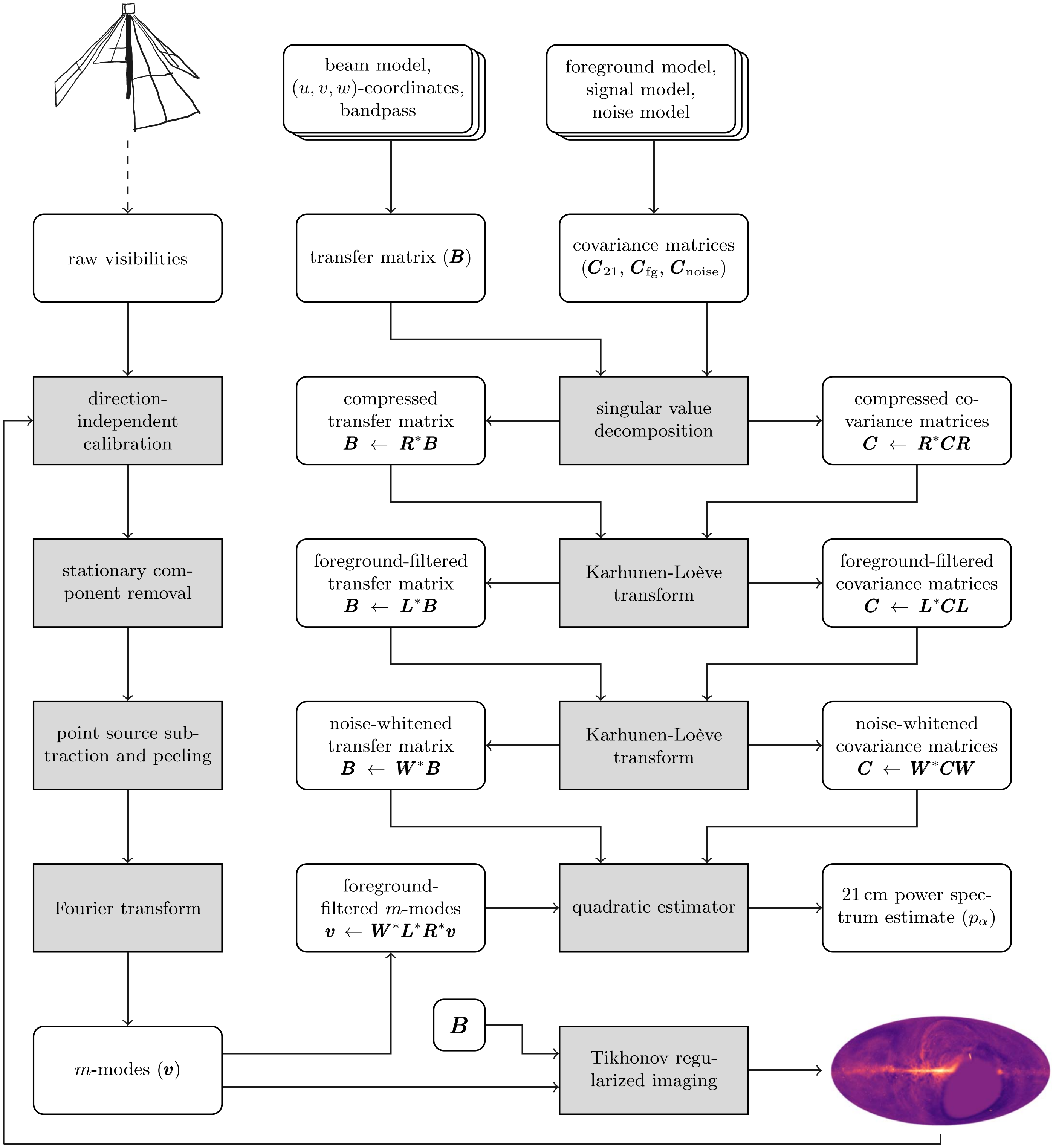


Figure 1. Flow chart describing the data analysis steps performed in this paper. Shaded boxes represent processing steps, whereas unshaded boxes represent data. Radio waves are received by antennas (depicted in the upper-left corner), which are correlated to produce raw visibilities. These visibilities are then flagged, calibrated, and bright point sources are removed. After a full sidereal day's worth of data has been collected, these visibilities can be Fourier transformed to compute the measured m-modes. Separately, an empirical beam model is used to calculate the transfer matrix elements that describe the interferometer's sensitivity to the sky. Full covariance matrices are computed for the foreground emission, 21 cm signal, and thermal noise. These matrices are used to compress and filter foreground emission, and whiten the noise covariance. Finally, the resulting filtered m-modes are used to estimate the spatial power spectrum of 21 cm emission. Images of the sky can be constructed through the use of Tikhonov-regularized imaging (Eastwood et al. 2018), which are useful for diagnosing errors in the analysis.

composed of Cyg A and Cas A where the absolute spectrum of Cyg A is given by Baars et al. (1977), and the spectrum of Cas A is adjusted for its secular decrease of 0.77% per year (Helmboldt & Kassim 2009). Because this initial sky model is incomplete on large angular scales, baselines shorter than 15 wavelengths are excluded from the calibration routine. The gains are optimized using a variant of alternating least squares independently described by Mitchell et al. (2008) and Salvini & Wijnholds (2014). The bandpass amplitude is fit with a fifth-order polynomial, and the phase is fit with a term for the delay and a term for dispersion through the ionosphere. Smoothing the gain calibration in this way helps to avoid modeling errors during calibration propagating into bandpass errors that can limit the sensitivity of the interferometer to the 21 cm power spectrum (Barry et al. 2016; Ewall-Wice et al. 2017). After this initial calibration and source removal, a model of the diffuse Galactic emission is constructed using Tikhonov-regularized m-mode analysis imaging. This model is then used to recalibrate the data with a more complete model of the sky.

The OVRO-LWA analog signal path is susceptible to additive common-mode radio frequency interference (RFI). A model for the common-mode RFI is constructed from the gain-calibrated visibilities after averaging over the entire 28 hr observing period with the phase center left at zenith. Averaging the visibilities in this way smears out the contribution of the sky along characteristic sidereal tracks. We then select the dominant components of the averaged visibilities to be used as templates for the RFI. The templates are manually inspected for residual sky emission by imaging each component with WSCLEAN (Offringa et al. 2014), and checking for features that are swept along sidereal tracks. These templates are scaled and subtracted from each integration to suppress the contamination of the common-mode RFI.

The top panel of Figure 2 is a dirty image of the sky constructed from this data set prior to any point source removal. A handful of bright point sources occupy the northern sky—namely Cas A, Cyg A, Her A, Hya A, Tau A, Vir A, 3C 123, 3C 353, and the Sun. Each of these sources is removed from the visibilities by employing a combination of peeling (direction-dependent calibration) for the brightest sources, and source fitting and subtraction for the fainter sources. This source removal strategy is described in greater detail by Eastwood et al. (2018).

Finally, in order to reduce the data volume and computational cost of further reductions, we selected only baselines representable with spherical harmonics with multipole number $l \leqslant 300$. This effectively selects only baselines from the core of the OVRO-LWA (23,947 baselines in total), which contain the majority of the brightness temperature sensitivity. The data were additionally averaged down to channel widths of 240 kHz. At 73 MHz, this averaging effectively smears out the spatial power spectrum on $k_\parallel \approx 1\ \mathrm{Mpc}^{-1}$ scales, but is permissible because the expected cosmological signal is small at these scales.

3. Formalism

3.1. m-*mode Analysis*

In this paper we apply the m-mode analysis formalism developed by Shaw et al. (2014, 2015). The interested reader should consult the aforementioned references for additional details, but m-mode analysis is briefly summarized below.

The measured quantity in a drift-scanning telescope is a periodic function of sidereal time. The Fourier transform with respect to sidereal time of this measured quantity is called an m-mode, where the value of m indicates how rapidly this mode varies over the course of a sidereal day. $m = 0$ corresponds to the mean value of the measurement over a sidereal day. $m = \pm 1$ corresponds to the components that vary once over a sidereal day. Larger absolute values of m represent contributions to the measurement that vary on increasingly rapid timescales.

The primary advantage of making this transformation to m-modes is that it can be shown that the set of measured m-modes with a given value for m are a linear combination of the spherical harmonic coefficients with the same value of m. This allows the data to be partitioned by m, and each partition can be manipulated independently of the remaining data set. Typically this leads to a large reduction in the processing time, which allows for the application of otherwise infeasible data analysis techniques that make use of the full covariance matrix of the data set.

We will adopt the convention that the measured m-modes are contained in a vector $\boldsymbol{v}$, and the spherical harmonic coefficients of the sky brightness are contained in a vector $\boldsymbol{a}$. The transfer matrix $\boldsymbol{B}$ describes the interferometer's response to the sky and is block-diagonal when both $\boldsymbol{v}$ and $\boldsymbol{a}$ are sorted by the absolute value of m. If we explicitly decompose the sky in terms of the high-redshift 21 cm contribution $\boldsymbol{a}_{21}$, and the foreground radio emission $\boldsymbol{a}_{\mathrm{fg}}$, then

$$\boldsymbol{v} = \boldsymbol{B}\boldsymbol{a}_{21} + \boldsymbol{B}\boldsymbol{a}_{\mathrm{fg}} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{n}$ is the contribution of thermal noise to the measurement.

The rows of the transfer matrix $\boldsymbol{B}$ fundamentally describe the response of each baseline to the sky represented by $\boldsymbol{a}$. The individual elements of the matrix are computed from spherical harmonic transforms of each baseline's fringe pattern (including the response of the antenna beams and bandpass). Eastwood et al. (2018) demonstrated all-sky imaging in a single synthesis imaging step through inverting Equation (1). However, that demonstration was restricted to single-channel imaging due to, in part, the computational and storage requirements associated with computing $\boldsymbol{B}$.

3.2. *Hierarchical Transfer Matrices*

Modern interferometers are composed of large numbers of antennas ($N \gg 10$) arranged in configurations that have both long and short baselines. For instance, the OVRO-LWA has over 30,000 baselines. The shortest baseline is 5 m, and the longest is 1.5 km. Consequently the OVRO-LWA measures a large range of angular scales. We can exploit this fact to reduce the computer time and disk space required to compute and store the transfer matrix $\boldsymbol{B}$.

The sensitivity of a baseline of length b to spherical harmonic coefficients with multipole moment l is $\propto j_l(2\pi b/\lambda)$, where j_l is the spherical Bessel function of the first kind, and λ is the wavelength. When $l \gtrsim 2\pi b/\lambda$, the spherical Bessel functions rapidly drop to zero (see the Appendix for more details about spherical Bessel functions). Consequently, even though the transfer matrix is block-diagonal, each diagonal block of the transfer matrix can also contain a large number of zero elements.

Therefore, when the columns and rows of each transfer matrix block $\boldsymbol{B}_m$ are sorted by the multipole number l and baseline length respectively, each block has the following

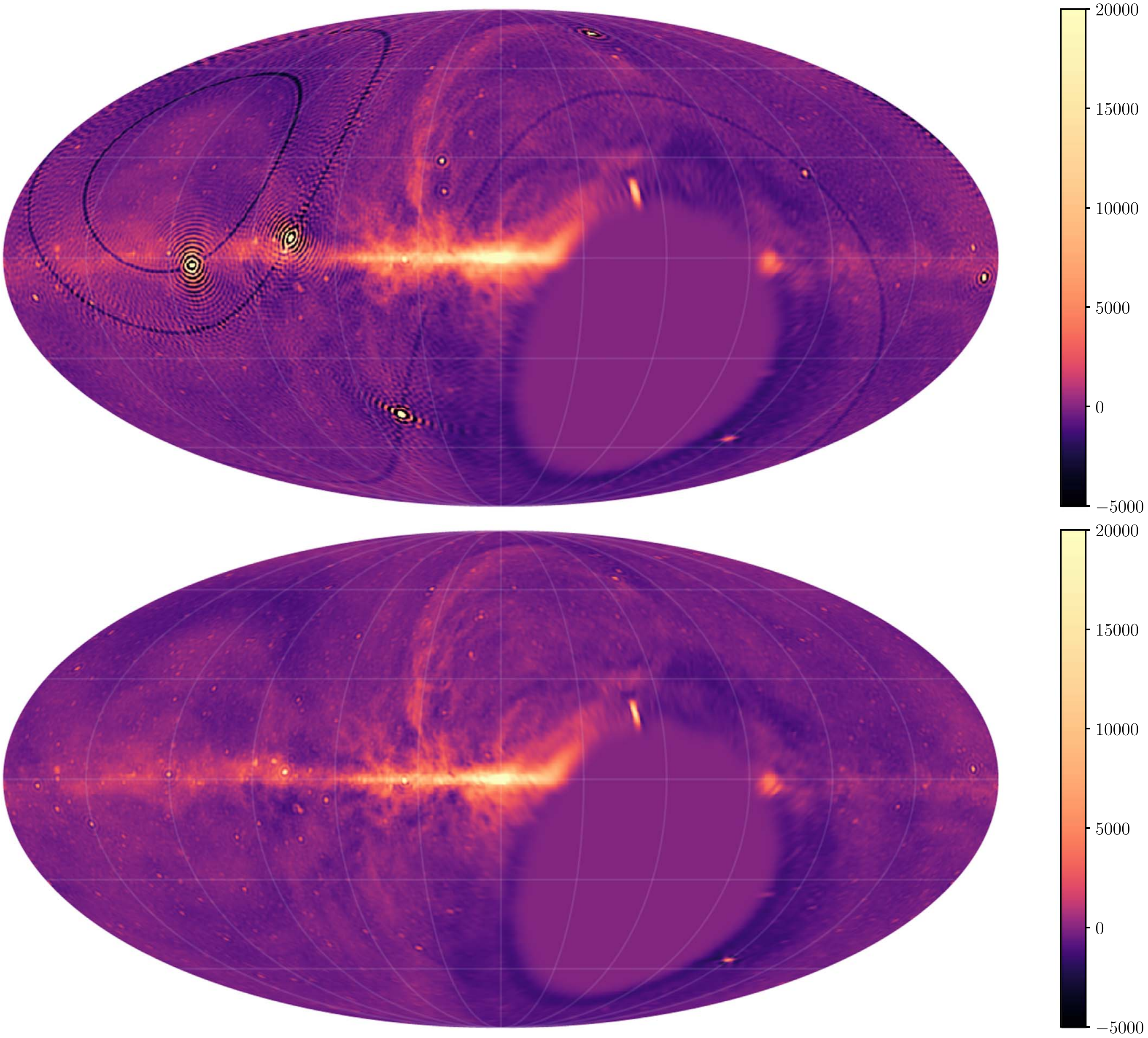
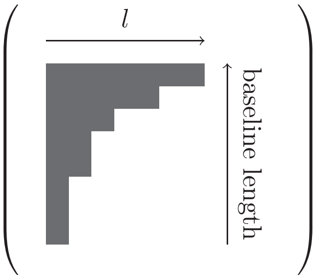


Figure 2. Molleweide projection of a Tikhonov-regularized image of the sky constructed from all baselines representable with $l_{\rm max} \leqslant 200$, and 2.6 MHz of bandwidth centered on 73.2 MHz. The color scale is linear between −1000 K and +1000 K, and logarithmic outside of this range. No cleaning has been performed, so all point sources are convolved with a point-spread function, and no masking of low declinations has been performed. The resolution of the maps naturally degrades at low declinations and the regularization scheme naturally encourages the map to be zero below the horizon. Negative rings at the decl. of bright point sources are an artifact of the fact that $m = 0$ modes are filtered from the data set due to their susceptibility to RFI and common-mode pickup. Top: before bright point sources are removed from the data set; bottom: after point source removal.

structure:

$$\boldsymbol{B}_m = \qquad (2)$$



Shaded regions represent elements with nonzero value, whereas unshaded regions represent elements with approximately zero value due to the fact that $l \gtrsim 2\pi b/\lambda$. This structure makes it apparent that it is not necessary to store every element of each transfer matrix block. In fact, by partitioning the array into sets of baselines with similar length, one can achieve significant cost savings when computing and storing the transfer matrix elements.

Ultimately, for the OVRO-LWA we achieve a 58% compression of the transfer matrix by not storing elements that are approximately zero.

3.3. Data Compression

Further data compression is desirable because it reduces the computational costs of all subsequent analysis steps. We implement the singular value decomposition (SVD) compression described by Shaw et al. (2014, 2015). The SVD factorizes a matrix into a unitary matrix $\boldsymbol{U}$, a diagonal matrix $\boldsymbol{\Sigma}$, and another unitary matrix $\boldsymbol{V}$ such that

$$\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*. \tag{3}$$

The diagonal elements of $\boldsymbol{\Sigma}$ are called singular values and, in this case, represent the amplitude of the response of the interferometer to the corresponding singular vectors (i.e., the columns of $\boldsymbol{U}$). The data can therefore be compressed by selecting all singular values above a given threshold and computing

$$\boldsymbol{R} = \begin{pmatrix} & \vdots & \vdots & \\ \cdots & \boldsymbol{u}_i & \boldsymbol{u}_{i+1} & \cdots \\ & \vdots & \vdots & \end{pmatrix} \tag{4}$$

$$\boldsymbol{v}_{\text{compressed}} = \boldsymbol{R}^*\boldsymbol{v}, \tag{5}$$

where $\boldsymbol{u}_i$ is a column of $\boldsymbol{U}$ whose singular value passes the threshold, and $\boldsymbol{v}$ is the vector of measured m-modes. The transfer matrix is similarly transformed $\boldsymbol{B}_{\text{compressed}} = \boldsymbol{R}^*\boldsymbol{B}$, and covariance matrices become $\boldsymbol{C}_{\text{compressed}} = \boldsymbol{R}^*\boldsymbol{C}\boldsymbol{R}$.

This compression is especially effective for the OVRO-LWA because the compactness of the interferometer leads to many partial redundancies between similar baselines. This is simply a statement that the number of baselines used in the calculation $N_{\text{baselines}}$ is larger than the number of unknowns in each transfer matrix block. In this paper, we adopted $l_{\max} = 300$ as the maximum value of the multipole number. For the OVRO-LWA $N_{\text{baselines}} \gg 300$, so there are many redundancies in the data set even though no pair of baselines is individually redundant. In total this compression reduces the volume of data to a mere 0.6% of its original size (before discarding any singular values).

4. Covariance Matrices

We model the covariance of the observations $\boldsymbol{C} = \langle \boldsymbol{v}\boldsymbol{v}^* \rangle$ with contributions from thermal noise $\boldsymbol{C}_{\text{noise}}$, foreground emission $\boldsymbol{C}_{\text{fg}}$, and the cosmological 21 cm signal itself $\boldsymbol{C}_{21}$:

$$\langle \boldsymbol{v}\boldsymbol{v}^* \rangle = \boldsymbol{C} = \boldsymbol{C}_{21} + \boldsymbol{C}_{\text{fg}} + \boldsymbol{C}_{\text{noise}}, \tag{6}$$

where this expression implicitly assumes that the sky is an isotropic Gaussian-random field, and that the sky covariance should be understood as an average over realizations of the sky.

We will begin with a detailed description of the models, measurements, and calculations used to compute each of these covariance matrices.

4.1. Thermal Noise Covariance

The 21 cm signal is expected to be unpolarized,[12] so we form Stokes-I visibilities from the mean of the xx and yy visibilities. Under this convention, the covariance of the complex-valued Stokes-I visibilities is (see Chapter 9 of Taylor et al. 1999):

$$\boldsymbol{C}_{\text{noise}} = \left(\frac{2k_{\text{B}}T_{\text{sys}}}{\eta A_{\text{eff}}\sqrt{2\Delta\nu\tau}}\right)^2 \boldsymbol{I}, \tag{7}$$

where k_{B} is the Boltzmann constant, T_{sys} is the system temperature, η is the antenna efficiency, A_{eff} is the effective collecting area (each assumed to be the same for all antennas), $\Delta\nu$ is the bandwidth, and τ is the total integration time. The effective collecting area of the antenna is related to the solid angle of the primary beam Ω through $A_{\text{eff}} = \lambda^2/\Omega$. At 73 MHz, the OVRO-LWA dipoles have primary beams with $\Omega \sim 2.4$ sr or $A_{\text{eff}} \sim 7\ \text{m}^2$.

OVRO-LWA dipoles are designed to be sky-noise dominated ($\geqslant$6 dB between 20 and 80 MHz; Hicks et al. 2012). More precisely, the system temperature is given by

$$T_{\text{sys}} \approx \eta T_{\text{sky}} + T_{\text{pre-amp}}, \tag{8}$$

where η is the antenna efficiency, T_{sky} is the averaged brightness temperature of the sky (primarily the Galactic synchrotron emission) weighted by the primary beam pattern, and $T_{\text{pre-amp}}$ is the noise temperature of the first amplifier in the analog signal path. We expect $T_{\text{pre-amp}} \approx 250$ K and $\eta \lesssim 0.5$ (Hicks et al. 2012).

The LEDA experiment hosted at the OVRO-LWA measured the brightness temperature of the diffuse Galactic emission in the northern hemisphere using the five radiometric antennas (Price et al. 2018). At 70 MHz, the brightness temperature varies between 1700 K and 3200 K with a relatively flat spectral index that varies between -2.28 and -2.38. In the southern hemisphere, the EDGES experiment measured that the brightness temperature of the sky at 150 MHz varies between 257 K and 842 K with a spectral index that varies between -2.50 and -2.62 (Mozdzen et al. 2017). Extrapolating to 70 MHz, we expect the beam-weighted sky brightness temperature in the southern hemisphere to vary between 1700 K and 6200 K. The maximum brightness temperature corresponds to sidereal time when the Galactic center transits.

We measured the system temperature as a function of frequency and sidereal time using a five-point stencil to suppress the contribution of the sky emission to the measured visibilities:

$$\begin{aligned}\Delta(\nu, t) = 4V(\nu, t) &- V(\nu - 24\ \text{kHz}, t) - V(\nu, t - 13\ \text{s}) \\ &- V(\nu + 24\ \text{kHz}, t) - V(\nu, t + 13\ \text{s}),\end{aligned}$$

where $\Delta(\nu, t)$ is a quantity whose variance is 20 times larger than that of the measured visibilities $V(\nu, t)$ at the given frequency ν and time t. Note that 24 kHz is the native frequency resolution of the OVRO-LWA and 13 s is the integration time. Therefore this stencil takes the difference between each measured visibility and the bilinear interpolation from adjacent frequency channels and time integrations. We then estimated the system temperature from the variance of Δ. The measured system temperature is shown in Figure 3 compared to the sky temperature measured by LEDA and extrapolated from EDGES. As expected, the system temperature increases at lower frequencies due to the increasing sky brightness temperature, and varies sidereally reaching a maximum as the Galactic center transits the meridian. These measurements suggest that the antenna efficiency $\eta \sim 0.25$. Although the system temperature varies with time and frequency, we adopt a constant system temperature of 3500η K when computing the sensitivity of the OVRO-LWA. We expect this approximation to potentially introduce errors of ~10% to the

---

[12] Venumadhav et al. (2017) find that circular polarization may be used to measure primordial magnetic fields, but the amplitude of this effect is too small to consider measuring with existing low-frequency telescopes.

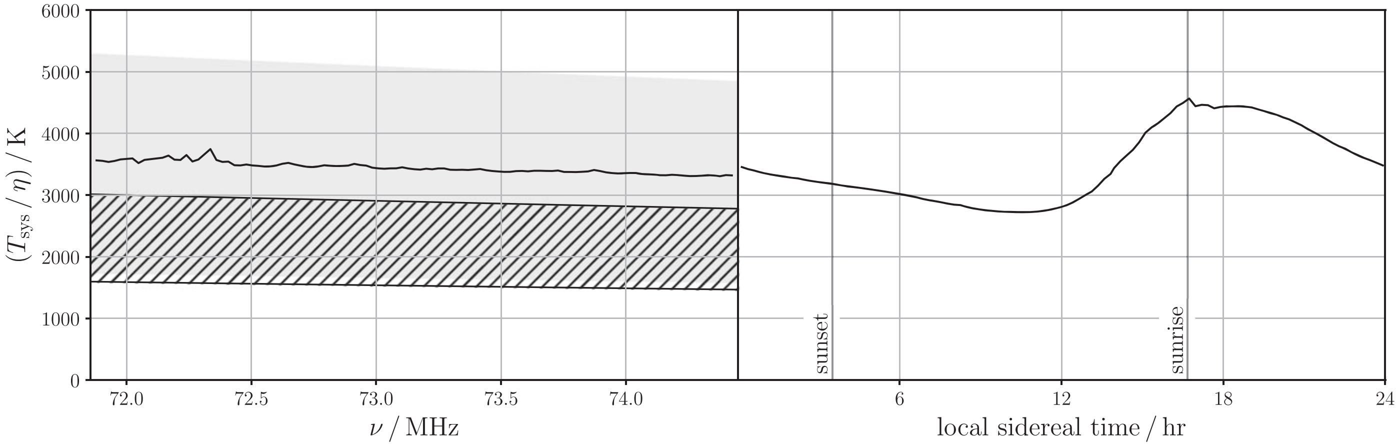


Figure 3. System temperature $T_{\rm sys}$ (scaled by the antenna efficiency η) measured as a function of frequency (left panel, solid black line), and local sidereal time (right panel, solid black line). The hatched region denotes the range of sky temperatures measured by the LEDA experiment (Price et al. 2018). The shaded region denotes the range of sky temperatures measured by the EDGES experiment in the southern hemisphere (Mozdzen et al. 2017).

computed sensitivity and error bars, which does not materially impact the results presented in this paper. In Section 6.5 we additionally simulate the impact of the implicit assumption of stationary thermal noise and conclude that it also does not bias our results.

4.2. Foreground Covariance

Under the assumption of a Gaussian random field, the covariance contributed by the sky can be computed from the multi-frequency angular power spectrum:

$$\langle a_{lm}(\nu)\, a^*_{l'm'}(\nu')\rangle = C_l(\nu,\, \nu')\, \delta_{ll'}\, \delta_{mm'}, \tag{9}$$

where the angled brackets denote an ensemble average over realizations of the sky, $a_{lm}(\nu)$ is the spherical harmonic coefficient of the sky brightness at frequency ν, $C_l(\nu,\, \nu')$ is the multi-frequency angular power spectrum at the multipole moment l, and between the frequencies ν and ν'. The Kronecker delta is represented by δ. The transfer matrix $\boldsymbol{B}$ describes how to relate the covariance of the spherical harmonic coefficients to the covariance of the measurements themselves, such that

$$\boldsymbol{C}_{\rm sky} = \boldsymbol{B}\boldsymbol{C}'_{\rm sky}\boldsymbol{B}^*, \tag{10}$$

where $\boldsymbol{C}_{\rm sky}$ is a term in Equation (6), and $\boldsymbol{C}'_{\rm sky}$ is a matrix whose elements are specified by Equation (9).

A common parameterization of $C_l(\nu_1,\, \nu_2)$ for foreground radio emission is (Santos et al. 2005)

$$C_l^{\rm fg}(\nu,\, \nu') = \sum_i A_i \left(\frac{l}{l_0}\right)^{-\alpha_i} \left(\frac{\nu\nu'}{\nu_0^2}\right)^{-\beta_i} \times \exp\left(-\frac{(\log\nu - \log\nu')^2}{2\zeta_i^2}\right), \tag{11}$$

where A_i represents the overall amplitude of a foreground component; α_i determines its angular spectrum, and β_i determines its frequency spectrum. Finally, ζ_i controls the degree to which nearby frequency channels are correlated. The statement that foreground emission is spectrally smooth here implies $\zeta_i^2 \gg \log^2(\nu/\nu')$ for each component. This parameterization allows for multiple power-law foreground components and ensures that the covariance matrix is positive definite. Because the fractional bandwidth is small, in this paper we assume $\zeta_i^2 \gg \log^2(\nu/\nu')$. For simplicity when measuring the foreground covaraiance, we will additionally assume a single foreground component such that Equation (11) can be written as

$$C_l^{\rm fg}(\nu,\, \nu') = \sqrt{C_l^{\rm fg}(\nu)\, C_l^{\rm fg}(\nu')}\,, \tag{12}$$

where $C_l^{\rm fg}(\nu) = C_l^{\rm fg}(\nu,\, \nu)$ is the single-frequency angular power spectrum.

We measured the angular power spectrum of the foreground emission at each frequency channel using a quadratic estimator (Tegmark 1997). The angular power spectrum is given by

$$C_l^{\rm fg}(\nu) = [\boldsymbol{F}^{-1}(\boldsymbol{q} - \boldsymbol{b})]_l\,, \tag{13}$$

where $\boldsymbol{F}$ is the Fisher information matrix, $\boldsymbol{q}$ is a quadratic function of the input data, and $\boldsymbol{b}$ is the bias due to thermal noise. The elements of the Fisher matrix $\boldsymbol{F}$ are given by

$$F_{ll'} = \sum_m |\boldsymbol{w}_{lm}^* \boldsymbol{C}_m^{-1} \boldsymbol{w}_{l'm}|^2, \tag{14}$$

where $F_{ll'}$ is the Fisher matrix element corresponding to the multipole numbers l and l', $\boldsymbol{w}_{lm}$ is the column of the transfer matrix corresponding to l and the azimuthal quantum number m, and $\boldsymbol{C}_m$ is the covariance matrix block corresponding to m. The convention $|\cdot|$ is used here to indicate the magnitude of a complex number. The elements of $\boldsymbol{q}$ and $\boldsymbol{b}$ are given by

$$q_l = \sum_m |\boldsymbol{w}_{lm}^* \boldsymbol{C}_m^{-1} \boldsymbol{v}_m|^2 \tag{15}$$

$$b_l = \sum_m \|\boldsymbol{w}_{lm}^* \boldsymbol{C}_m^{-1} \boldsymbol{C}_{{\rm noise},m}^{1/2}\|^2, \tag{16}$$

where $\boldsymbol{v}_m$ is the vector of m-modes corresponding to the given value of m, and $\boldsymbol{C}_{{\rm noise},m}$ is the corresponding block of the noise covariance matrix. The convention $\|\cdot\|$ is used here to indicate the magnitude of a complex vector (the usual Euclidean norm).

The result of applying this quadratic estimator to the data set at 73.260 MHz (a representative channel) can be seen in Figure 4. Broadly, the data can be described with a power law in l, but the

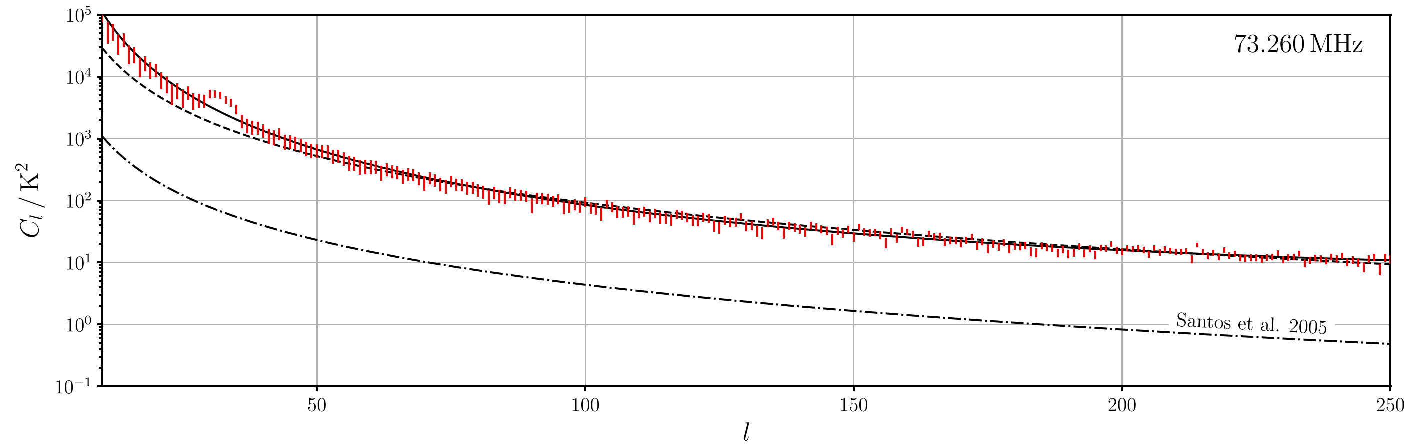


Figure 4. Angular power spectrum of the sky as measured by the OVRO-LWA at 73.260 MHz. Measurements (with 95% uncertainty) are indicated with red bars. The uncertainty is dominated by sample variance. The dashed black line is the best-fit power-law spectrum, and the solid black line is the best-fit solution when the power-law index is allowed to run. The dashed–dotted line is a model derived, in part, from the Haslam 408 MHz sky map (Haslam et al. 1981, 1982; Santos et al. 2005). The feature at $l \sim 30$ is sensitive to the choice of covariance matrix, and is therefore likely instrumental.

quality of the fit is somewhat poor. A single power-law fit gives

$$C_l \sim 92. \times \left(\frac{l}{100}\right)^{-2.5} \mathrm{K}^2. \qquad (17)$$

In fact, while this is a reasonable fit at $l > 75$, a shallower power-law index is preferred $l < 75$. If we allow for the power-law index to run, the best-fit model becomes

$$C_l \sim 85. \times \left(\frac{l}{100}\right)^{-3.2 + l/277.} \mathrm{K}^2. \qquad (18)$$

A comparison of these two models can be seen in Figure 4 in addition to a model of the Galactic synchrotron emission derived by Santos et al. (2005), which appears to underestimate the amplitude of C_l by an order of magnitude. Because the fractional bandwidth of this measurement is small, essentially all reasonable spectral indices are permitted. We adopt a fiducial spectral index of -2.5 as a compromise between the spectral indices measured by LEDA and EDGES. We will discuss the potential impact of an error in the spectral index in Section 6.4 as part of a broader discussion of calibration and bandpass errors.

4.3. Signal Covariance

Given the isotropic three-dimensional spatial power spectrum of the 21 cm brightness temperature $P_z^{21}(k)$ with the wavenumber k and at the redshift z, the multi-frequency angular power spectrum $C_l(\nu, \nu')$ is given by

$$C_l^{21}(\nu, \nu') = \frac{2}{\pi} \int P_z^{21}(k)\, j_l(kr_z)\, j_l(kr_{z'})\, k^2\, dk, \qquad (19)$$

where r_z is the comoving distance to the redshift z (specified by the frequency ν), and $j_l(x)$ is the spherical Bessel function of the first kind. In the flat-sky approximation, Equation (19) can be simplified to

$$C_l^{21}(\nu, \nu') \approx \frac{1}{\pi r_z r_{z'}} \int P_z^{21}(k_\perp, k_\parallel) \cos(k_\parallel \Delta r_z)\, dk_\parallel, \qquad (20)$$

where $k_\perp = l/r_z$ and $k_\parallel = \sqrt{k^2 - k_\perp^2}$. See the Appendix for a derivation of this approximation and the assumptions that must be satisfied for it to be a reasonable approximation.

If $P_z^{21}(k_\perp, k_\parallel)$ is additionally assumed to be a piecewise linear function, Equation (20) can be evaluated analytically. Under this assumption, $P_z^{21}(k_\perp, k_\parallel)$ can be represented using linear hat functions (triangular functions in two dimensions), such that

$$P_z^{21}(k_\perp, k_\parallel) = \sum_\alpha p_\alpha \times \mathrm{hat}_\alpha(k_\perp, k_\parallel) \qquad (21)$$

$$C_l^{21}(\nu, \nu') \approx \frac{1}{\pi r_z r_{z'}} \sum_\alpha p_\alpha H_\alpha(\Delta r_z) \qquad (22)$$

where $H_\alpha(\Delta r_z) = \int \mathrm{hat}_\alpha(k_\perp, k_\parallel) \cos(k_\parallel \Delta r_z)\, dk_\parallel$.

The flat-sky approximation is valid only when the power spectrum is smooth enough for rapid oscillations in the spherical Bessel functions to cancel out. The hat functions are non-differentiable, and so we must compute the error associated with this pixelization of the power spectrum. Figure 5 gives the relative error on the computed angular power spectrum for a fiducial hat function power spectrum. Generally the error is 10^{-4}, but can reach to 10^{-1} at values where $C_l \approx 0$. This is an acceptable error in the context of this paper, but future experiments may wish to experiment with differentiable basis functions.

When selecting a fiducial model for the 21 cm power spectrum we prefer to remain unopinionated, and therefore adopt a flat power spectrum with a single free parameter, the overall amplitude of the dimensionless power spectrum Δ_{21}:

$$P_{\mathrm{fiducial}}^{21}(k) = \frac{2\pi^2}{k^3} \Delta_{21}^2. \qquad (23)$$

Prior to the recent detection of an absorption feature centered at 78 MHz by Bowman et al. (2018), the amplitude of the power spectrum was generally predicted to be $\Delta_{21} < 20$ mK at $z \sim 20$ (e.g., Fialkov et al. 2014). However, more recent predictions in the context of the measured 78 MHz absorption feature predict a much brighter power spectrum (e.g., Barkana 2018; Kaurov et al. 2018). However, we adopt $\Delta_{21} = 20$ mK as a fiducial power spectrum amplitude in order to be somewhat conservative.

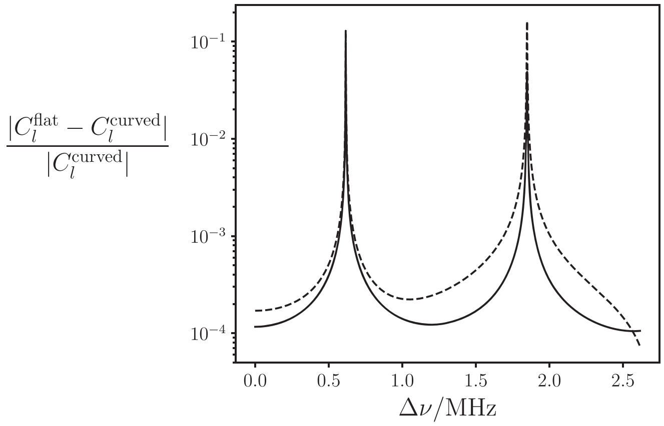


Figure 5. Relative error involved in making the flat-sky approximation for a hat function power spectrum (i.e., the relative difference between Equations (19) and (20)) with $l = 10$ (solid line) and $l = 100$ (dashed line). The hat function is centered at $k_\parallel = 0.1\ \text{Mpc}^{-1}$ with a domain that extends from $0.095\ \text{Mpc}^{-1}$ to $0.105\ \text{Mpc}^{-1}$. The spikes in relative error correspond to when $C_l^{\text{curved}}(\Delta\nu) \approx 0$.

The amplitude of the fiducial 21 cm signal is primarily used to determine which modes should be kept by the foreground filter described in the following section. Therefore, if readers are skeptical of this selection of the power spectrum amplitude, they may simply choose to interpret the results as if the foreground filter is stronger or weaker than expected. A foreground filter that is weaker than expected could lead to a biased power spectrum estimate, but a foreground filter that is stronger than expected will lead to a conservative result with error bars that may be larger than necessary.

5. Foreground Filtering

In the preceding sections, we have derived and, where appropriate, measured the contribution of thermal noise, foreground emission, and the cosmological 21 cm emission to the complete covariance matrix of the data. This was possible because the transfer matrix $\boldsymbol{B}$ is block-diagonal with respect to m, and we assumed that the sky emission is a Gaussian-random field (i.e., there are no correlations between different values of m). Without these properties the full covariance matrix is generally too large to represent and manipulate on any existing computer. Shaw et al. (2014, 2015) were therefore able to derive a new foreground filtering technique that exploits knowledge of the full covariance matrix. This filter is called the double Karhunen–Loève transform (double KL transform). In this section we will briefly summarize the action of this foreground filter and demonstrate its application to the OVRO-LWA. We will finally attempt to develop an intuitive understanding by relating its behavior to the "foreground wedge" commonly seen in the literature (e.g., Parsons et al. 2012; Vedantham et al. 2012; Thyagarajan et al. 2015).

The KL transform is closely related to the generalized eigenvalue problem. For two Hermitian, positive-definite matrices $\boldsymbol{C}_1, \boldsymbol{C}_2 \in \mathbb{C}^N$, we would like to find all pairs of eigenvalues λ_i and eigenvectors $\boldsymbol{v}_i$ for which

$$\boldsymbol{C}_1 \boldsymbol{v}_i = \lambda_i \boldsymbol{C}_2 \boldsymbol{v}_i. \tag{24}$$

Because both matrices are Hermitian, it quickly follows that the eigenvalues λ_i must be real. Because both matrices are additionally positive definite, it follows that the eigenvalues λ_i must all be positive. Furthermore we can select the normalization of the eigenvectors such that

$$\boldsymbol{v}_i^* \boldsymbol{C}_1 \boldsymbol{v}_i = \lambda_i \tag{25}$$

$$\boldsymbol{v}_i^* \boldsymbol{C}_2 \boldsymbol{v}_i = 1. \tag{26}$$

Under this convention the eigenvalues have a simple interpretation as the ratio of the mode power contained in $\boldsymbol{C}_1$ relative to $\boldsymbol{C}_2$. All N eigenvalues and eigenvectors can be conveniently found with a single call to LAPACK (Anderson et al. 1990).

In Section 4.2 we derived and measured a model for the foreground contribution to the data covariance $\boldsymbol{C}_{\text{fg}}$. In Section 4.3 we projected a fiducial model 21 cm power spectrum to a multi-frequency angular power spectrum, and therefore derived its contribution to the data covariance $\boldsymbol{C}_{21}$. We can solve the generalized eigenvalue problem for the eigenvectors (arranged as columns within the matrix $\boldsymbol{L}$) that simultaneously diagonalize both matrices (called the KL transform):

$$\boldsymbol{L} \boldsymbol{C}_{\text{fg}} \boldsymbol{L}^* = \boldsymbol{\Lambda} \tag{27}$$

$$\boldsymbol{L} \boldsymbol{C}_{21} \boldsymbol{L}^* = \boldsymbol{I}, \tag{28}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix, and $\boldsymbol{I}$ is the identity matrix. The foreground filter is simply constructed by selecting only the eigenvectors for which the corresponding eigenvalue (i.e., the foreground–signal power ratio) is less than some value ϵ_{filter} selected by the observer. The application of this filter to a fiducial set of models can be seen in the second row of Figure 6. The signal covariance matrix has been diagonalized and the power in each remaining mode is greater than the surviving power in the foreground covariance matrix. The off-diagonal elements in the foreground covariance matrix are due to numerical errors. The possible effect of these numerical errors on the efficacy of the foreground filter is noted here, but is out of the scope of the current work.

Much emphasis has been placed on maintaining the integrity of the "foreground wedge" in the next generation of 21 cm telescopes. In its simplest form, the existence of the foreground wedge is a statement that most foreground radio emission that observers have to contend with when trying to detect the cosmological 21 cm is spectrally smooth. A simple Fourier transform of an image cube therefore leads to most contamination occupying the space where $k_\parallel$ (the line-of-sight wavenumber) is small. However, due to the chromatic nature of interferometers (specifically that the fringe spacing $\propto b/\lambda$ where b is the baseline length and λ is the wavelength), this contamination is spread out into a wedge-like structure. Additional chromaticity in, for example, the bandpass or antenna primary beam leads to the contamination even leaking out of the wedge. In the event of too much leakage, the observer has lost the ability to measure the cosmological 21 cm transition.

In contrast, the KL transform automatically finds the optimal linear combination of the data set for separating foregrounds using all available information built into the models. This includes information on the frequency spectrum of the foregrounds as well as their angular structure, which can lead to scenarios where the KL transform can filter foreground emission that cannot be avoided with a delay filter. There is, of course, a caveat that the KL transform requires sufficiently detailed models for the instrument and foreground emission.

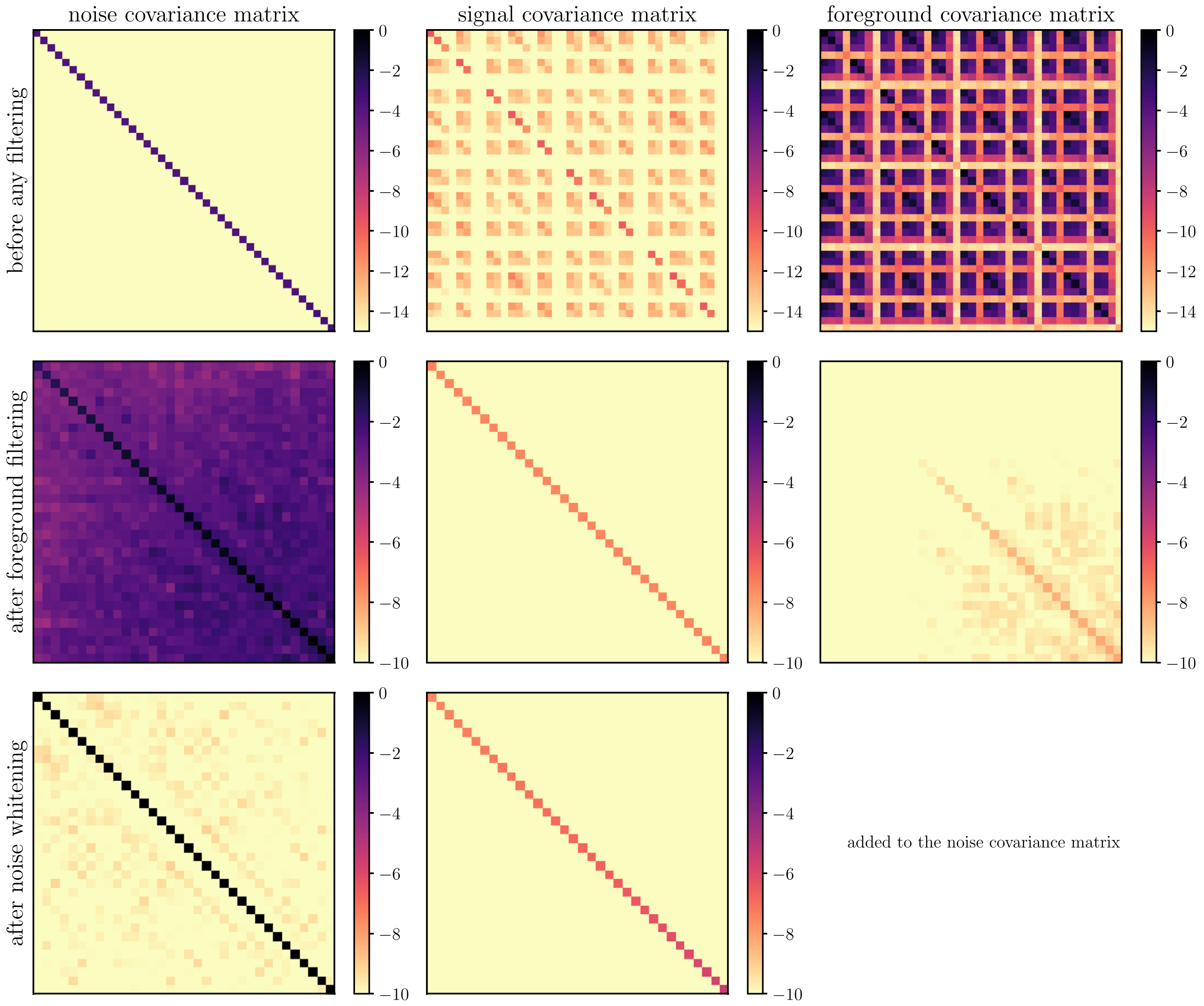


Figure 6. Illustration of the action of foreground filtering on each of the covariance matrices discussed in Section 4. The left column corresponds to the noise covariance matrix, the middle column corresponds to the high-redshift 21 cm contribution to the covariance, and the right column corresponds to the foreground covariance matrix. The top row is before any filtering has been applied, the middle row is after the first KL transform, and the bottom row is after the second KL transform. The color gives the logarithm (base 10) of the absolute value of the matrix element. Prior to any filtering, baselines are ordered by length and then by frequency. The apparent structure in the signal and foreground covariances is due to this ordering combined with the downsampling of the matrices necessary to produce this graphic. After filtering and whitening, the rows/columns are ordered by the magnitude of the corresponding eigenvalues. This diagram illustrates which matrices are diagonal, and the relative amplitude of the matrix elements after each stage in the processing.

However, it is not necessarily optimal to remain completely apathetic to the structure of foreground emission, and most collaborations are expending significant effort to characterize their instruments.

A single KL transform, however, leads to large off-diagonal elements in the noise covariance matrix (see the second row of Figure 6). Therefore Shaw et al. (2014, 2015) introduced a second KL transform that diagonalizes the noise covariance matrix. This second matrix composed of eigenvectors will be denoted by $\boldsymbol{W}$. In total we therefore have

$$\boldsymbol{C}_{\text{filtered}} = \underbrace{\boldsymbol{W}^*\boldsymbol{L}^*\boldsymbol{C}_{21}\boldsymbol{L}\boldsymbol{W}}_{\boldsymbol{S}} + \underbrace{\boldsymbol{W}^*\boldsymbol{L}^*(\boldsymbol{C}_{\text{fg}} + \boldsymbol{C}_{\text{noise}})\boldsymbol{L}\boldsymbol{W}}_{\boldsymbol{I}}, \tag{29}$$

where $\boldsymbol{C}_{\text{filtered}}$ is the data covariance matrix after applying the double KL transform foreground filter, $\boldsymbol{S}$ is a real diagonal matrix, and $\boldsymbol{I}$ is the identity matrix. The diagonal elements of $\boldsymbol{S}$ give the expected signal–noise ratio in each mode. The foreground filter is applied to the measured m-modes by simply computing

$$\boldsymbol{v}_{\text{filtered}} = \boldsymbol{W}^*\boldsymbol{L}^*\boldsymbol{v}. \tag{30}$$

In this paper we will repeat the analysis using three different values for the foreground filtering foreground–signal threshold ϵ_{filter}. This will allow us to assess the performance of the foreground filter and degree to which residual foreground contamination may be affecting the measurement. We will adopt the terminology “strong,” “moderate,” and “mild” to

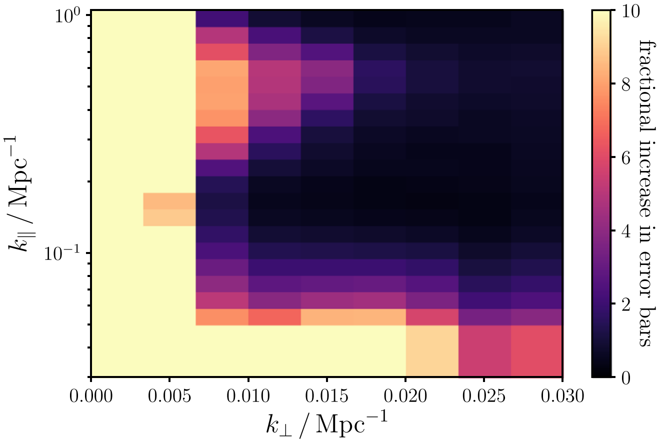


Figure 7. Fractional increase in the size of the error bars in each power spectrum bin due to the application of a double KL transform foreground filter (moderate strength).

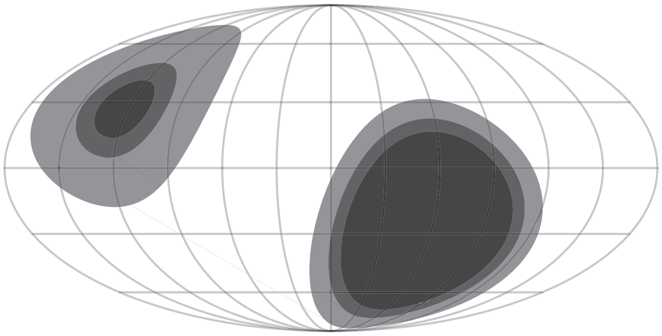


Figure 8. Mollweide projected illustration of the sky where shaded regions are down-weighted by the foreground filter. More precisely, we measured the rms along rings of constant decl. after foreground filtering. Shaded regions are where the rms is less than 1% of the maximum rms. From darkest to lightest, these regions of the sky are filtered by the mild, moderate, and strong foreground filters respectively.

mean

$$\begin{aligned} \epsilon_{\text{filter}} &= 0.1 \qquad (\text{"strong" foreground filtering}) \\ \epsilon_{\text{filter}} &= 1 \qquad (\text{"moderate" foreground filtering}) \\ \epsilon_{\text{filter}} &= 10 \qquad (\text{"mild" foreground filtering}). \end{aligned}$$

Now we will build a physical intuition for understanding the operation of the double KL transform foreground filter.

Figure 7 illustrates the fractional increase in error bars associated with applying the moderate foreground filter. In the space of a cylindrically binned power spectrum, the action of the filter is to discard linear combinations of the data set with low $k_\parallel$ and low $k_\perp$. This manifests itself as a decrease in sensitivity—equivalently an increase in the error bars—in this region of parameter space. High $k_\parallel$ modes are computed from rapid frequency differences, whereas low $k_\parallel$ modes are slowly varying in frequency. Because the foreground emission is spectrally smooth, it tends to corrupt modes with low $k_\parallel$. The pattern of this contamination is known as the foreground wedge. However, the foreground filter additionally removes emission on large angular scales (low $k_\perp$). This arises because the foreground filter is aware that the foreground emission is brighter on larger angular scales (see Figure 4 and Equation (18)).

As illustrated in Figure 8, the foreground filter also tends to remove emission in two separate parts of the sky: low declinations that are never seen at high elevations from the OVRO-LWA, and high declinations around the north celestial pole (NCP). This filtering of high and low declinations can be seen in Figure 9, which is a Tikhonov-regularized image of the sky constructed from the post-filtered data.

The OVRO-LWA is a zenith pointing drift-scanning instrument. Therefore foreground emission located far from zenith has a large path difference between antennas. This large path difference leads to additional frequency structure that allows the foreground emission to contaminate higher values of $k_\parallel$ (Morales et al. 2012). Similarly, Thyagarajan et al. (2015) derived the impact of widefield effects on the foreground contamination and found that baseline foreshortening can lead to additional Galactic synchrotron emission on large angular scales contaminating the measurement. This foreground emission from low elevations is problematic. The double KL transform suppresses the contribution of these low elevations to the measurement.

Emission from the vicinity of the NCP is characterized by its low fringe rate. As the Earth rotates, emission located here moves slowly through the fringes of the interferometer. Therefore this emission is predominantly characterized by low values of m. The foreground emission, however, is brightest relative to the cosmological 21 cm emission at low values of l (large angular scales). Because $m \leqslant l$ for a given value of l, low values of m are disproportionately contaminated by the brightest diffuse components of the foreground emission. In fact, for the fiducial foreground and signal models presented in Sections 4.2 and 4.3 respectively, the the foreground–signal ratio of the most favorable mode is $\propto m^{-3.5}$. This is a reflection of the fact that emission with a higher fringe rate tends to be smaller in angular extent. Consequently, the foreground filter aggressively discards information from small values of m and the emission located at the NCP is collateral damage because it can be difficult to separate from the diffuse foreground emission. This can be seen in Figure 8 where increasing the strength of the foreground filter increases the area around the NCP that is down-weighted.

6. Results and Error Analysis

We will use a quadratic estimator to measure the spatial power spectrum of 21 cm fluctuations (Tegmark 1997). In particular we estimate the coefficients p_α, defined in Equation (21). As described by Padmanabhan et al. (2003), the observer may tune the estimator by selecting a windowing function that produces the desired properties. For example, given the measured data $\boldsymbol{v}$, the full covariance matrix $\boldsymbol{C}$, and the Fisher information matrix $\boldsymbol{F}$, the unwindowed and minimum variance estimates of the power spectrum amplitude are

$$\hat{p}_\alpha^{\text{unwindowed}} = \sum_\beta [\boldsymbol{F}^{-1}]_{\alpha\beta}(q_\beta - b_\beta) \tag{31}$$

$$\hat{p}_\alpha^{\text{min. variance}} = \left(\sum_\beta F_{\alpha\beta}\right)^{-1}(q_\beta - b_\beta), \tag{32}$$

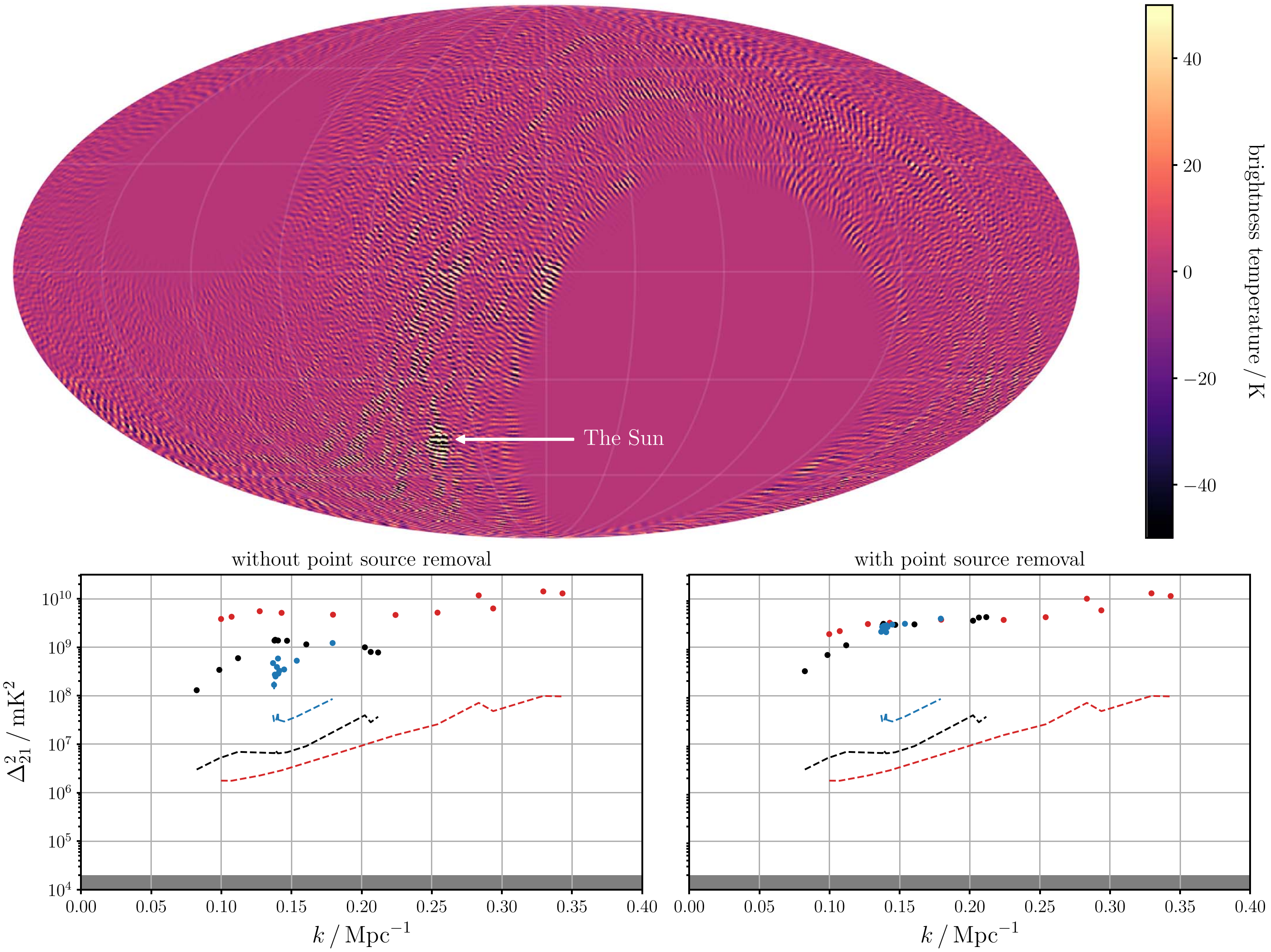


Figure 9. Top: Mollweide projected image of the sky after point source removal and moderate foreground filtering. The dominant residual feature in the residuals is associated with the Sun. Bottom: the power spectrum estimated without (left) and with (right) point source removal at a range of filter strengths. Points correspond to the estimated power spectrum amplitude and the dashed lines correspond to the computed thermal noise (95% confidence). Mild foreground filtering is red, moderate foreground filtering is black, and strong foreground filtering is blue. The shaded region represents the sensitivity required to detect or rule out optimistic models for the 21 cm power spectrum.

where

$$q_\alpha = \boldsymbol{v}^* \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{C}}{\partial p_\alpha} \boldsymbol{C}^{-1} \boldsymbol{v} \tag{33}$$

$$b_\alpha = \mathrm{tr}\left(\boldsymbol{C}^{-1} \frac{\partial \boldsymbol{C}}{\partial p_\alpha} \boldsymbol{C}^{-1} \boldsymbol{C}_{\mathrm{noise}} \right) \tag{34}$$

$$F_{\alpha\beta} = \mathrm{tr}\left(\boldsymbol{C}^{-1} \frac{\partial \boldsymbol{C}}{\partial p_\alpha} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{C}}{\partial p_\beta} \right). \tag{35}$$

Directly computing $F_{\alpha\beta}$ from its definition is computationally expensive, and so we compute an approximation of the Fisher information matrix using the iterative Monte Carlo scheme described by Padmanabhan et al. (2003) and Dillon et al. (2013).

We will make exclusive use of the minimum variance estimator in this paper because it is relatively insensitive to errors in the Fisher information matrix, which are inevitable due to the Monte Carlo computation. Additionally, the unwindowed estimator can compound numerical errors when the condition number of $\boldsymbol{F}$ is large.[13]

In Figure 9 we present the results of the quadratic estimator with and without point source removal, and across the range of foreground filter strengths. These estimates are, across the board, severely limited by systematic errors. This is readily apparent due to the extreme amplitude of the estimated power. We therefore interpret these measurements as upper limits $\Delta^2_{21} \lesssim (10^4\ \mathrm{mK})^2$ at $k \approx 0.10\ \mathrm{Mpc}^{-1}$.

As the strength of the foreground filter is increased more information is lost by the filter. This is seen in the window functions of the quadratic estimator. With mild foreground filtering, the window functions are roughly evenly spaced between $k = 0.10\ \mathrm{Mpc}^{-1}$ and $0.35\ \mathrm{Mpc}^{-1}$. With strong foreground filtering, all of the measurements are instead concentrated around $k = 0.15\ \mathrm{Mpc}^{-1}$, which reflects the loss of information at other values of the wavenumber k.

---

[13] The condition number of a matrix $\boldsymbol{A}$ is $\kappa(\boldsymbol{A}) = \|\boldsymbol{A}\|\,\|\boldsymbol{A}^{-1}\|$ and describes the error introduced when solving the linear equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ for the vector $\boldsymbol{x}$. As a general rule of thumb, if $\log_{10}\kappa(\boldsymbol{A}) = N$, one can expect to lose N digits of precision after computing $\boldsymbol{A}^{-1}\boldsymbol{b}$.

After initial calibration and stationary component removal, we attempted to subtract the eight brightest point sources in the northern hemisphere in addition to the Sun. The brightest of these sources were removed with direction-dependent calibrations. The fainter sources were simply subtracted after fitting for their flux and position (attempting to account for ionospheric scintillation and refraction). The Sun was removed using a resolved source model. With mild foreground filtering, this point source removal leads to a $\sim 2\times$ reduction in the power spectrum amplitude.

However, the efficacy of the foreground filter materially differs between the data sets where bright point sources have and have not been removed. Without point source removal, increasing the strength of the foreground filter leads to a reduction of the estimated power. This reflects the fact that the foreground filter is removing increasing amounts of foreground contamination. In contrast, if point sources have been subtracted, the power spectrum amplitude is insensitive to the strength of the foreground filter. While the point source removal routine leads to less foreground contamination in the absence of foreground filtering, it also restricts the effectiveness of the foreground filter. This suggests that the point source removal routine introduces additional errors into the data set that inhibit the action of the foreground filter.

We will now attempt to diagnose the source of these residual systematic errors that limit this measurement. While doing this, we will adopt the moderate foreground filter as the fiducial foreground filter due to its action as a compromise between the amount of foreground emission removed and resolution in the wavenumber k.

6.1. *Even–Odd Jackknife*

Errors arising from variations on rapid timescales—the timescale of a single correlator dump—can be revealed through a comparison of results obtained data using only even-numbered integrations and the interleaving odd-numbered integrations. These two halves of the data set have independent thermal noise with additional errors due to ionospheric scintillation, radio frequency interference (RFI), and source subtraction errors.

In prior work we observed that ionospheric scintillation generates $\sim 10\%$ fluctuations in the flux of a point source on 13 s timescales at 73 MHz (Eastwood et al. 2018). The position of a source varies more slowly by up to $4'$ on 10 minute timescales. Therefore comparing even- and odd-numbered integrations will reveal errors arising from ionospheric scintillation, but not necessarily from variable ionospheric refraction.

Figure 10 contains a map of the sky constructed from differencing the even and odd data sets (after point source removal). This map is almost featureless. If ionospheric scintillation was contributing a substantial amount of additional noise to the measurement, we would expect to see enhanced residuals in the vicinity of bright point sources. Instead, the dominant features are a ~ 50 K residual at the location of the Sun, and some artifacts at low declinations that do not rise above 10° elevation (likely generated by RFI). We therefore conclude that over a long 28 hr integration, the ionospheric scintillation has averaged down and is not the dominant source of error.

The bottom panel of Figure 10 compares the amplitude of the estimated power spectrum after differencing the even- and odd-numbered integrations. Differencing the two halves of the data set cancels out the majority of the residual contamination of foreground emission into the measurement. Therefore the power spectrum decreases in amplitude. The improvement is roughly one order of magnitude before source subtraction and only a factor of 2–3 after source subtraction. Point source removal is conducted independently on each integration; sporadic errors and source subtraction residuals will therefore also tend to manifest on the timescale of a single integration. This measurement therefore suggests that source subtraction residuals could be a limiting factor for this estimate of the 21 cm power spectrum.

6.2. *Day–Night Jackknife*

The dominant subtraction residual in the preceding section is associated with the Sun, which is a difficult source to cleanly subtract due to its complex structure. We can therefore split the data into two halves: data collected while the Sun is above the horizon, and below the horizon. The data collected during the night have a number of advantages. Specifically, subtraction residuals associated with the Sun cannot impact data collected during the night. Additionally the ionospheric total electron content (TEC) is lower during the night because the Sun acts as a source of ionization for the ionosphere. Specifically, the median vertical TEC measured within 200 km of OVRO rose to 20 TECU (total electron content unit) during the day, but dropped to 6 TECU during the night. There were no geomagnetic storms during the observing period and the fact that these observations were collected during the winter months generally contributes to a reduction in the ionospheric TEC. Finally, due to the time of year, the sky temperature is lower at night. For these reasons, we generally expect an improvement in the nighttime data with respect to the daytime data.

In principle, m-mode analysis requires that data be collected for a full sidereal day because the m-modes are computed from the Fourier transform of the visibilities with respect to sidereal time. We relax that requirement here. When selecting half the data, we additionally apply a Blackman–Harris window function to prevent ringing. Tikhonov-regularized images made from just the daytime and nighttime data can be seen in Figure 11. These dirty images serve as a proof of concept that m-mode analysis can reasonably be applied to data sets without a full sidereal day's worth of data.

We estimated the power spectrum from each half of data and the results are presented in Figure 11. Restricting the observations to nighttime only leads to a substantial improvement in the power spectrum limits both with and without point source removal. In fact the measurements with and without point source removal are now comparable. This suggests the point source subtraction residuals are less of an issue in the nighttime data due to the fact that (due to the time of year) there are fewer bright point sources that were removed.

6.3. xx–yy *Jackknife*

The polarization angle of linearly polarized emission rotates as it propagates through a magnetized plasma (e.g., Jelić et al. 2014). The rotation angle is $\propto \lambda^2$, where λ is the wavelength of the radiation. Therefore instrumental polarization leakage from a linear polarization (Stokes-Q or Stokes-U) into total intensity (Stokes-I) can introduce additional spectral structure into the foreground emission that is not accounted for in our currently unpolarized analysis.

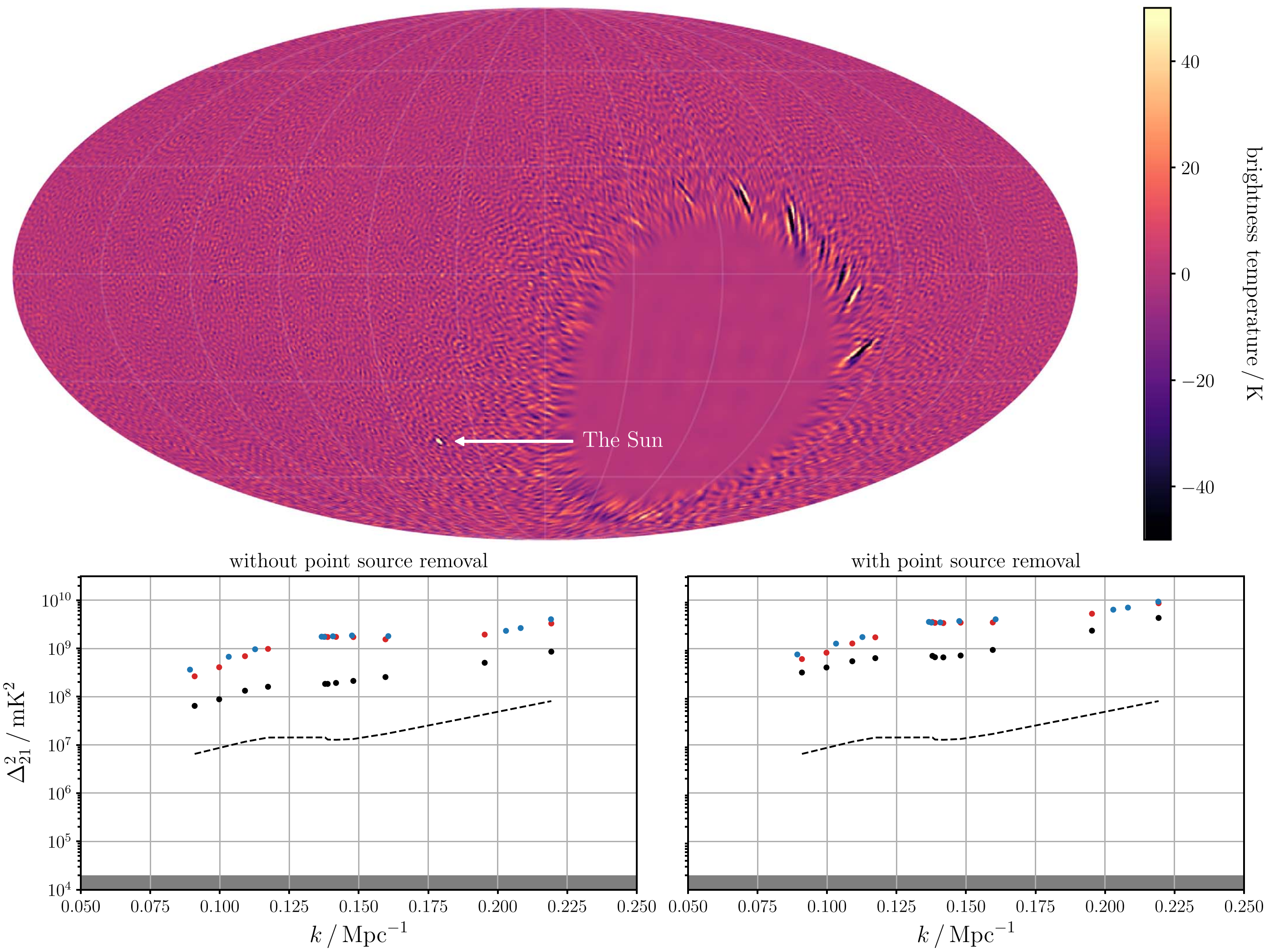


Figure 10. Top: Mollweide projection of the sky in Galactic coordinates after differencing even- and odd-numbered integrations. The Sun is the dominant artifact in this image due to the sporadic failure of source subtraction. Large residuals are also present at low declinations that do not rise above 10° elevation. These low-elevation artifacts are generated by RFI. Bottom: the power spectrum estimated without (left) and with (right) point source removal. Point correspond to the estimated power spectrum amplitude and the dashed line corresponds to the computed thermal noise (95% confidence). Red and blue points are estimates from the even- and odd-numbered integrations respectively. Black points are estimates after computing the difference between the two halves of data. The shaded region represents the sensitivity required to detect or rule out optimistic models for the 21 cm power spectrum.

If Faraday-rotated, linearly polarized emission is a problem, it will be exacerbated by computing the power spectrum from the xx correlations and yy correlations separately. For this comparison, the transfer matrix $\boldsymbol{B}$ must be recomputed using the correct response pattern for the individual dipoles. An image of the sky computed from the difference of the xx and yy correlations is shown in Figure 12. This map is related to the linear polarization of the sky, but does not account for the full polarization of the beam, and therefore includes some amount of instrumental polarization.

We estimated the 21 cm power spectrum from the xx and yy correlations. This estimate is shown in Figure 12. The estimates are comparable to the total intensity estimate and we therefore conclude that polarization leakage is not currently a major source of systematic error.

6.4. Calibration Errors

Many authors have investigated the impact of calibration errors on an experiment's ability to separate foreground emission from the cosmological 21 cm signal (Barry et al. 2016; Ewall-Wice et al. 2017). In this section we will compute the impact of calibration errors on the double KL transform foreground filter.

In this calculation we will simulate a realistic set of visibilities for the foreground emission, and introduce errors into the calibration before applying the double KL transform filter. Finally we will estimate the power spectrum amplitude as a way to characterize the amount of contamination associated with the calibration errors.

The angular structure of the foreground model used here is measured from the actual data (shown in the bottom panel of Figure 2), but the frequency dependence of this emission is chosen to be a power law with a fiducial spectral index of −2.3. This spectral index was chosen to be consistent with the results reported by LEDA (Price et al. 2018), but due to the small fractional bandwidth of this measurement, we expect these results to be insensitive to the specific choice of spectral index. The set of m-modes we expect to measure with the

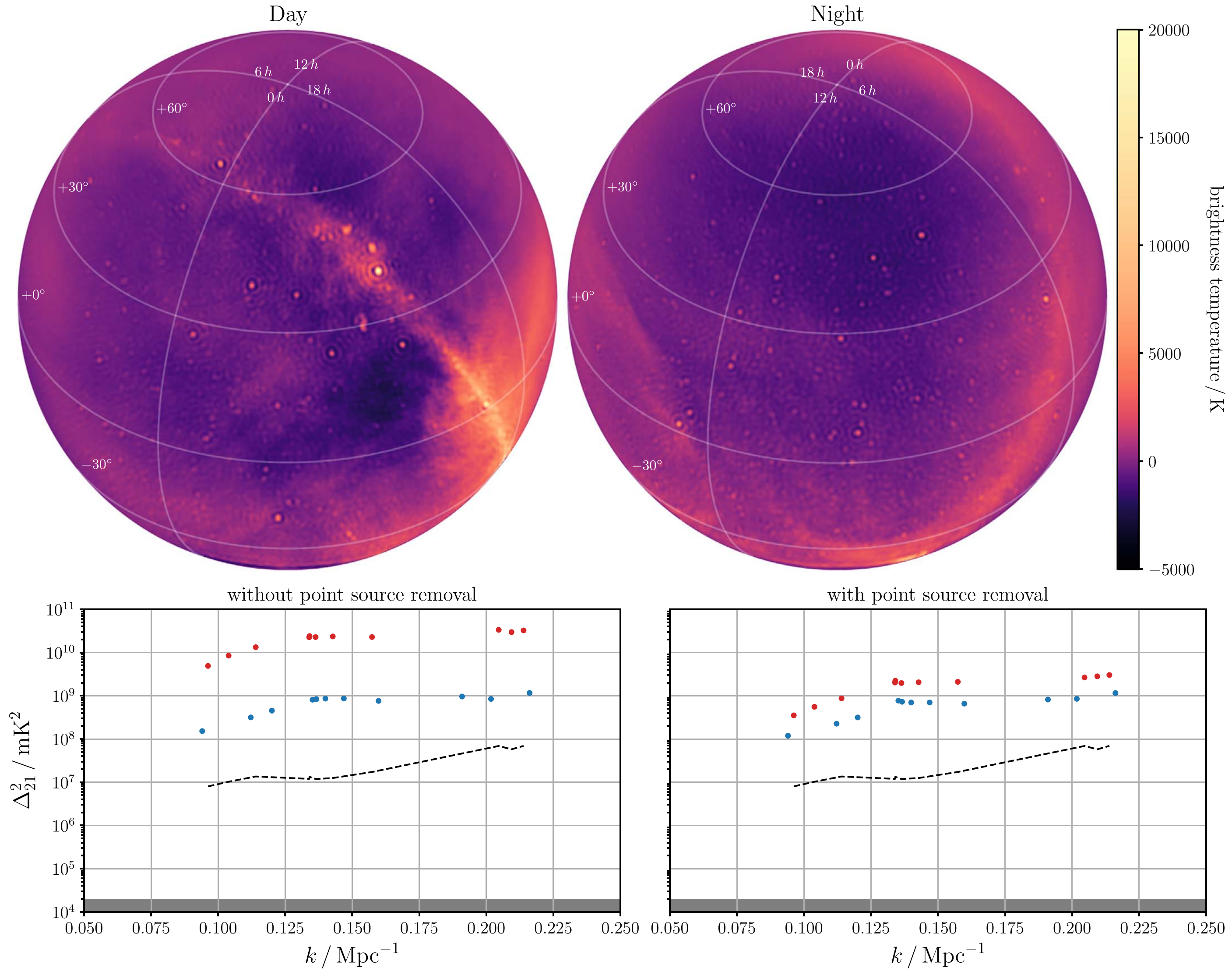


Figure 11. Top: orthographic projection of the sky constructed from data collected only during the day (left), and only during the night (right). Bottom: the power spectrum estimated without (left) and with (right) point source removal. Points correspond to the estimated power spectrum amplitude and the dashed line corresponds to the computed thermal noise (95% confidence). Measurements from the day are red, and those from the night are blue. The shaded region represents the sensitivity required to detect or rule out optimistic models for the 21 cm power spectrum.

interferometer $\boldsymbol{v}_{\text{simulated}}$ is computed by

$$\boldsymbol{v}_{\text{simulated}} = \boldsymbol{B}\boldsymbol{a}_{\text{simulated}}, \tag{36}$$

where $\boldsymbol{B}$ is the transfer matrix, and $\boldsymbol{a}_{\text{simulated}}$ is a vector of the spherical harmonic coefficients of the foreground model.

At this point the simulated m-modes are corrupted with calibration errors. We explore two possibilities:

1. each antenna and frequency channel receives an incorrect gain calibration;
2. each frequency channel receives an incorrect gain calibration, but this error is coherent across antennas.

In each case, the gain errors are drawn from a complex normal distribution, and the amplitude of the error is varied between 0.1%, 1%, and 10%. The former case is coined "random gain errors" to indicate that each antenna is given an error in its complex gain calibration. The latter case is coined "random bandpass errors" to indicate that the overall bandpass of the interferometer is perturbed. The impact of these calibration errors can be seen in Figure 13.

In order to avoid biasing the 21 cm power spectrum, these results indicate that the gain calibration must be derived to an accuracy better than 0.1%. A general rule of thumb for the OVRO-LWA is that, for equal amplitude errors, the foreground contamination generated by bandpass errors that are coherent across all antennas are an order of magnitude worse than for the random gain errors (in units of Δ^2_{21}). Therefore to achieve a comparable level of foreground contamination, the overall bandpass of the interferometer must be known to better than 0.01%.

The data set presented in this paper is systematically limited at roughly $\Delta^2_{21} \sim (10^4\,\text{mK})^2$. These limits are therefore consistent with ~1% errors in the overall bandpass of the interferometer. The top panel of Figure 13 therefore presents the fractional difference in the sky images between two adjacent frequency channels (after averaging down to 240 kHz channel resolution). The residuals in this sky map are typically 2%–3%, but generally do not correlate with the sky brightness. We therefore conclude that between adjacent 240 kHz channels, the bandpass error is less than 1%. In fact, the

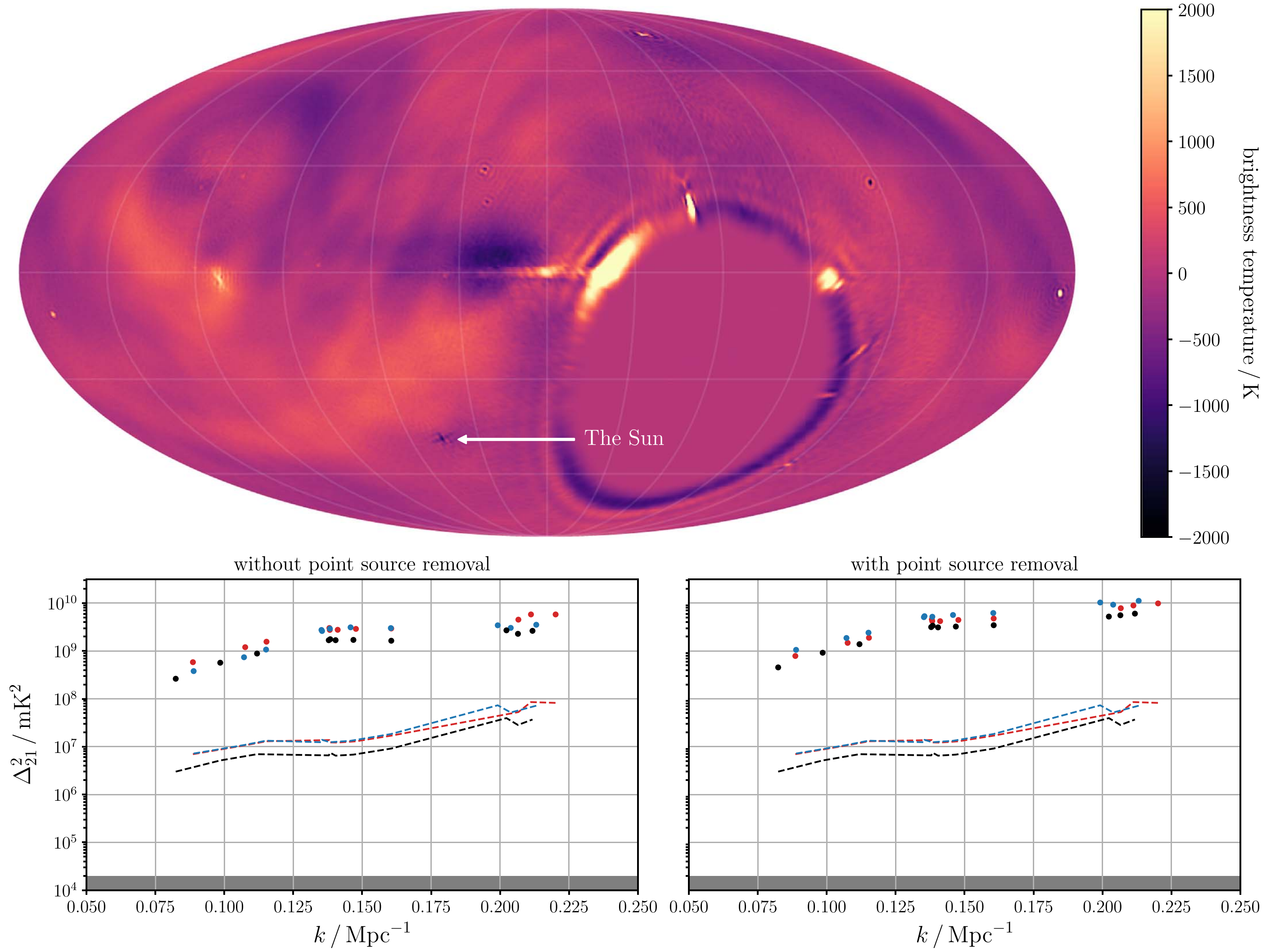


Figure 12. Top: Mollweide projection of the sky in Galactic coordinates after differencing the *xx* and *yy* correlations. Note that this is not true linear polarization because it does not account for the full polarization of the antenna response pattern. Bottom: the power spectrum estimated without (left) and with (right) point source removal. Points correspond to the estimated power spectrum amplitude and the dashed line corresponds to the computed thermal noise (95% confidence). Red and blue points are estimates from the *xx* and *yy* correlations respectively. Black points are estimates from the mean of the *xx* and *yy* correlations. The shaded region represents the sensitivity required to detect or rule out optimistic models for the 21 cm power spectrum.

structure of the residuals in Figure 13 suggests a different terrestrial source. Terrestrial sources of radio emission do not move through the sky at a sidereal rate. Therefore when constructing images of the sky, this contaminating emission tends to be smeared along rings of constant decl. These ring-like structures are clearly visible in Figure 13 alongside some larger-scale diffuse structures.

However, if we attribute the residual emission entirely to gain errors, then these simulations suggest that the antenna gains are known only to within a couple percent, which when considered alongside the RFI that contaminates the measurement, is likely sufficient to explain the current systematic limitations of our data set.

The results of this section additionally put some constraints on the impact of errors in the foreground covariance computed in Section 4.2. For instance, a ∼0.1 error in the spectral index of the foreground emission may be interpreted as a ∼0.5% spectrally smooth bandpass error that will degrade the performance of the foreground filter. However, because these errors are spectrally smooth, the performance degradation is less than what has been simulated in this section (random bandpass errors with no correlation between frequency channels). We therefore interpret the calculations performed here as an upper bound. A ∼0.1 error in the spectral index of the foreground emission will lead to systematic errors less than a comparable ∼0.5% spectrally unsmooth error in the bandpass, which may be inferred from Figure 13.

6.5. Non-stationarity

One implicit assumption of *m*-mode analysis is that all measured components are stationary: the foreground emission, the 21 cm signal, and the thermal noise.

For foreground components, we assumed that the emission is a Gaussian random field. Shaw et al. (2014, 2015) demonstrated that under ideal circumstances this assumption on its own does not substantially bias CHIME estimates of the 21 cm power spectrum. However, when the antenna beam is not perfectly measured, foreground emission leaks through the foreground filters to corrupt the measurement. We also see this in Section 6.4, where bandpass errors lead to residual foreground emission contaminating the measurement, but in

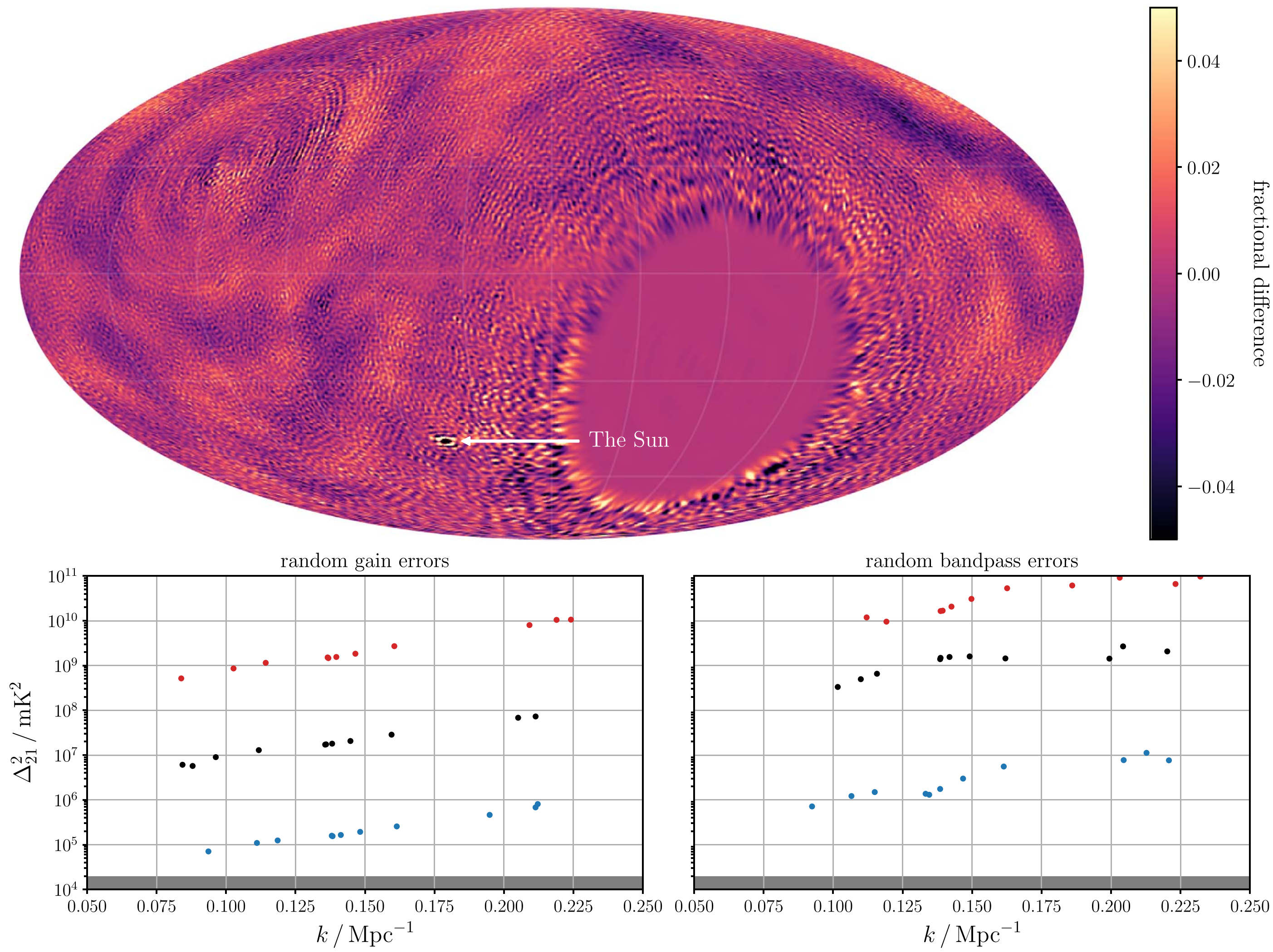


Figure 13. Top: Mollweide projection of the sky in Galactic coordinates after differencing two adjacent 240 kHz frequency channels. Bottom: simulated power spectrum estimates as a result of a foreground model and gain errors that are incoherent (left) and coherent (right) between antennas. Blue corresponds to 0.1% errors, black corresponds to 1% errors, and red corresponds to 10% errors in the complex gains. The shaded region represents the sensitivity required to detect or rule out optimistic models for the 21 cm power spectrum.

the absence of bandpass errors the real non-Gaussian foreground emission does not directly limit the measurements.

An additional source of non-stationarity can be seen in Figure 3, which shows that the system temperature varies by approximately 50% over the course of a sidereal day. We simulated the impact of this variation in system temperature by computing visibilities from a realistic sky model (see Figure 2) and injecting noise into the measured visibilities. No 21 cm signal is injected into the measurement because we are simply trying to measure the impact of the injected noise on foreground filtering and power spectrum estimation. We ran the simulation twice. In the first case, the system temperature varied sidereally and was directly sampled from the data presented in Figure 3. In the second case, the system temperature was held fixed at the mean system temperature. After foreground filtering and power spectrum estimation, the two estimates were indistinguishable within the computed uncertainties. We therefore conclude that sidereal variation of the system temperature is not a limiting factor in our experiment.

7. Conclusion

In this paper we estimated the amplitude of the 21 cm power spectrum of the Cosmic Dawn with 28 hr of data from the OVRO-LWA. This measurement was severely limited by systematic errors and therefore we interpret our measurements as upper limits, which are currently the most sensitive at this epoch and the first measurement at $z > 18$. We measured $\Delta^2_{21} \lesssim (10^4\ \mathrm{mK})^2$ at $k \approx 0.10\ \mathrm{Mpc}^{-1}$.

In making this measurement, we demonstrated the first application of the double KL transform foreground filter to a measured data set. We demonstrated that the application of this foreground filter can lead to improved power spectrum limits, and in combination with Tikhonov-regularized imaging, we developed a physical intuition for the action of the foreground filter. The double KL transform derives its action from models for the foreground and 21 cm signal covariance. We measured the angular power spectrum of the foreground emission and found that the power-law index appears to steepen on large angular scales ($l < 50$). The 21 cm signal covariance is derived from the flat-sky approximation, which we derive in the Appendix.

Although application of the foreground filter leads to some improvement in our measurement of the 21 cm power spectrum, the improvement was relatively modest. This is essentially a reflection of the fact that the true covariance of the data does not match the expectations of the models. We performed a series of jackknife tests and simulations that appear to implicate a combination of source subtraction errors, terrestrial interference, and calibration errors as limiting factors in this measurement.

If this measurement was thermal noise limited, we may have expected to place limits at the level of $\Delta_{21}^2 \sim (10^3\ \mathrm{mK})^2$ to $(10^{3.5}\ \mathrm{mK})$ depending strongly on the amount of foreground filtering that is needed to adequately suppress the foreground contamination. Because the action of the foreground filter is to exchange thermal noise sensitivity for more comprehensive removal of foreground emission, extrapolating from here to find the requisite sensitivity of an interferometer that will be able to make a detection is difficult. An improved calibration may greatly reduce the required integration time, while an increased need for foreground filtering may extend the required integration time. However, from simple scaling arguments, and in an optimistic scenario where the experiment does attain its noise limited sensitivity, the OVRO-LWA could begin to constrain the brightest models of the 21 cm power spectrum ($\Delta_{21}^2 \sim (10^2\ \mathrm{mK})^2$) with between 10^3 and 10^4 hr of observing time. However, a detection of the Cosmic Dawn 21 cm spatial power spectrum will require that gain errors be restricted to less than 0.1% and bandpass errors to less than 0.01%. Improved results will certainly require improving the instrumental calibration and source removal, which will help to prevent foreground emission leaking through the measurement and into the power spectrum estimate.

This material is based in part upon work supported by the National Science Foundation under grants AST-1654815 and AST-1212226. The OVRO-LWA project was initiated through the kind donation of Deborah Castleman and Harold Rosen.

Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration, including partial funding through the President's and Director's Fund Program.

This work has benefited from open-source technology shared by the Collaboration for Astronomy Signal Processing and Electronics Research (CASPER). We thank the Xilinx University Program for donations; NVIDIA for proprietary tools, discounts, and donations; and Digicom for collaboration on the manufacture and testing of DSP processors.

We thank the Smithsonian Astrophysical Observatory Submillimeter Receiver Lab for the collaboration of its members, and the observatory for research and development funds.

Development, adaptation, and operation of the LEDA real-time digital signal-processing systems at OVRO-LWA have been supported in part by NSF grants AST/1106059, PHY/0835713, OIA/1125087, and AST/1616709.

Appendix
Converting a Spatial Power Spectrum to an Angular Power Spectrum

The multi-frequency angular power spectrum $C_l(\nu, \nu')$ is measured from the spherical harmonic coefficients of the sky $a_{lm}(\nu)$ at the frequencies ν and ν':

$$C_l^{21}(\nu, \nu') = \frac{1}{2l+1} \sum_{m=-l}^{l} \langle a_{lm}^{21}(\nu)\, a_{lm}^{21*}(\nu') \rangle, \tag{37}$$

where the angled brackets should be understood as an ensemble average over sky realizations. Here the average over m is indicated with an explicit sum to distinguish it from the ensemble average.

The spherical harmonic coefficients themselves are computed from an integral over the 21 cm brightness temperature $T_\nu^{21}(\boldsymbol{r})$ over a spherical shell of the universe:

$$a_{lm}^{21}(\nu) = \int T_\nu^{21}(\boldsymbol{r})\, Y_{lm}^*(\hat{r})\, \delta(r - r_z)\, d^3r, \tag{38}$$

where $Y_{lm}(\hat{r})$ is a spherical harmonic function, and the Dirac delta function $\delta(r{-}r_z)$ is used to pick out the spherical shell of the universe at the comoving distance r_z to the redshift z.

The 21 cm brightness temperature is related to the power spectrum $P_z^{21}(\boldsymbol{k})$ through its Fourier transform:

$$T_\nu^{21}(\boldsymbol{r}) = \int T_\nu^{21}(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\, \frac{d^3k}{(2\pi)^3} \tag{39}$$

$$\langle T_\nu(\boldsymbol{k}) T_{\nu'}(\boldsymbol{k}') \rangle = (2\pi)^3\, \delta^3(\boldsymbol{k} - \boldsymbol{k}')\, P_z^{21}(\boldsymbol{k}). \tag{40}$$

Finally, we will need the "plane wave expansion" that describes a plane wave in terms of spherical harmonics:

$$e^{i\boldsymbol{k}\cdot\boldsymbol{r}} = 4\pi \sum_{lm} i^l\, j_l(kr)\, Y_{lm}(\hat{r})\, Y_{lm}^*(\hat{k}), \tag{41}$$

where the function $j_l(kr)$ is the spherical Bessel function of the first kind.

Putting this all together we can find

$$\begin{aligned}
C_l^{21}(\nu, \nu') &= \frac{1}{2l+1} \sum_{m=-l}^{l} \Bigg\langle \iint\!\!\iint T_\nu^{21}(\boldsymbol{k}) T_{\nu'}^{21*}(\boldsymbol{k}')\, e^{i(\boldsymbol{k}\cdot\boldsymbol{r} - \boldsymbol{k}'\cdot\boldsymbol{r}')} \\
&\quad \times Y_{lm}^*(\hat{r}) Y_{lm}(\hat{r}')\, \delta(r - r_z)\, \delta(r' - r_{z'})\, d^3r\, d^3r'\, \frac{d^3k}{(2\pi)^3}\, \frac{d^3k'}{(2\pi)^3} \Bigg\rangle \\
&= \frac{4\pi(-i)^l}{2l+1} \sum_{m=-l}^{l} \Bigg\langle \iiint T_\nu^{21}(\boldsymbol{k}) T_{\nu'}^{21*}(\boldsymbol{k}')\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\, j_l(k' r_{z'})\, Y_{lm}^*(\hat{r}) \\
&\quad \times Y_{lm}(\hat{k}')\, \delta(r - r_z)\, d^3r\, \frac{d^3k}{(2\pi)^3}\, \frac{d^3k'}{(2\pi)^3} \Bigg\rangle \\
&= \frac{(4\pi)^2}{2l+1} \sum_{m=-l}^{l} \Bigg\langle \iint T_\nu^{21}(\boldsymbol{k}) T_{\nu'}^{21*}(\boldsymbol{k}')\, j_l(kr_z)\, j_l(k' r_{z'}) \\
&\quad \times Y_{lm}^*(\hat{k})\, Y_{lm}(\hat{k}')\, \frac{d^3k}{(2\pi)^3}\, \frac{d^3k'}{(2\pi)^3} \Bigg\rangle \\
&= \frac{(4\pi)^2}{2l+1} \sum_{m=-l}^{l} \int P_z^{21}(\boldsymbol{k})\, j_l(kr_z)\, j_l(kr_{z'})\, Y_{lm}^*(\hat{k})\, Y_{lm}(\hat{k})\, \frac{d^3k}{(2\pi)^3}.
\end{aligned}$$

In the first two steps we used the plane-wave expansion, and in the final step we used the definition of the spatial power spectrum.

At this point if we assume that the power spectrum is isotropic and has no dependence on the orientation of the wave vector $\hat{k}$, then the angular component of the remaining integral

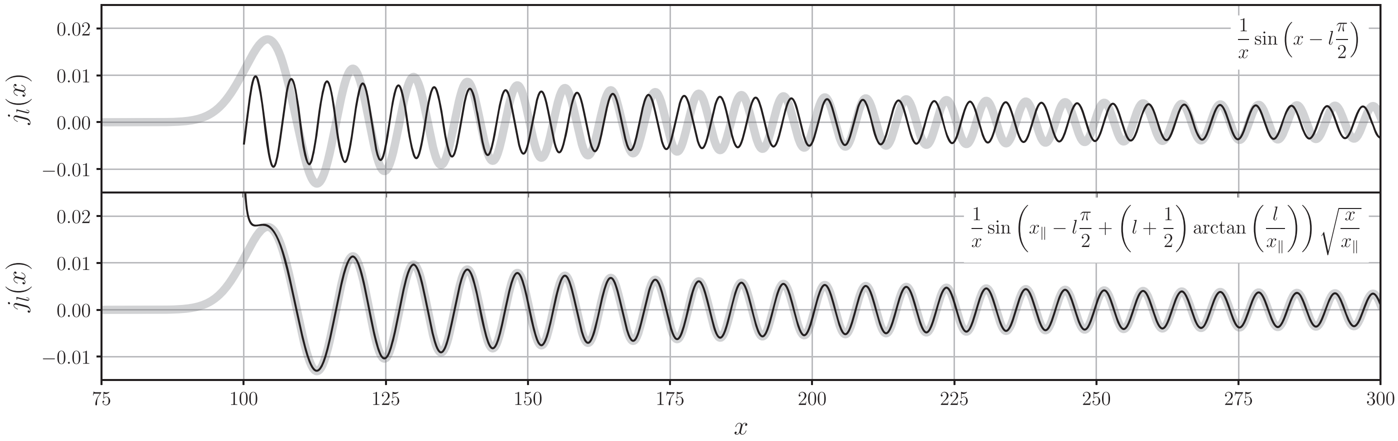


Figure 14. Comparison of two approximations (thin black lines) to the spherical Bessel function $j_l(x)$ with $l = 100$ (thick gray lines). The top panel shows the approximation derived from the limiting behavior of $j_l(x)$ as $x \to \infty$. This is a poor approximation near $x \sim l$. The bottom panel shows the approximation derived from the method of steepest descent. This approximation maintains the same limiting behavior as $x \to \infty$ and greatly improves the accuracy of the approximation near $x \sim l$.

can be performed to find

$$C_l^{21}(\nu, \nu') = \frac{2}{\pi}\int P_z^{21}(k)\, j_l(kr_z)\, j_l(kr_{z'})\, k^2\, dk. \tag{42}$$

Typically $r_z \sim 10{,}000\,\mathrm{Mpc}$ and $k \sim 0.1\,\mathrm{Mpc}^{-1}$ so the spherical Bessel functions $j_l(x)$ are typically evaluated in the limit of $l < x < l^2$. Equation (42) is exact for an isotropic power spectrum, but in practice integrals over the product of two spherical Bessel functions are numerically challenging due to their oscillatory behavior, so we will look for a scheme to approximate this integral. This approximation is simple in the regimes where $x \ll l$ and $x \gg l$, because

$$\lim_{x\to 0} j_l(x) \propto x^l \approx 0 \tag{43}$$

$$\lim_{x\to\infty} j_l(x) = \frac{1}{x}\sin\left(x - l\frac{\pi}{2}\right) + \mathcal{O}\left(\frac{l^2}{x^2}\right). \tag{44}$$

However, we are primarily interested in the intermediate regime where a better approximation can be obtained using the method of steepest descent (this method is also used, for example, to derive Stirling's approximation to log n! for large n). Starting with the integral representation of the spherical Bessel functions, the integration contour is deformed slightly to pass through saddle points of the integrand, approaching along paths of steepest descent. This allows the integral to be approximated as a Gaussian integral, which can be analytically evaluated to

$$j_l(x) \approx \frac{1}{x}\sin\left(x_\parallel - l\frac{\pi}{2} + \left(l + \frac{1}{2}\right)\arctan\left(\frac{l}{x_\parallel}\right)\right)\sqrt{\frac{x}{x_\parallel}}, \tag{45}$$

where we have suggestively defined $x_\parallel = \sqrt{x^2 - l^2}$. This approximation holds for $l \gg 1$ and $x > l$ (see Figure 14). The product of the two spherical Bessel functions in Equation (42) therefore results in rapid oscillations on top of a slower beat frequency. After computing the integral—provided the power spectrum is sufficiently smooth—the rapid oscillations will average down. Therefore this integral can be approximated using only the latter term:

$$C_l^{21}(\nu, \nu') \approx \frac{1}{\pi r_z r_{z'}}\int P_z^{21}(k_\perp, k_\parallel) \times \cos\left(k_\parallel \Delta r_z + \left(l + \frac{1}{2}\right)\arctan\left(\frac{\Delta r_z / r_z}{k_\perp/k_\parallel + k_\parallel/k_\perp}\right)\right) dk_\parallel, \tag{46}$$

where $k_\parallel = \sqrt{k^2 - k_\perp^2}$, $k_\perp = l/r_z$, and $\Delta r_z = r_{z'} - r_z$. Typically for nearby frequency channels, the argument to the arctangent will be small, and consequently we arrive at the "flat-sky approximation" used by Bharadwaj & Ali (2005) and Datta et al. (2007):

$$C_l^{21}(\nu, \nu') \approx \frac{1}{\pi r_z r_{z'}}\int P_z^{21}(k_\perp, k_\parallel)\cos(k_\parallel \Delta r_z)\, dk_\parallel. \tag{47}$$

In order to derive this approximation we require $l \gg 1$, $\Delta r_z / r_z \ll k_\parallel / k_\perp$, and that the power spectrum be smooth on scales that allow the rapid oscillations of the spherical Bessel functions to average down. The flat-sky approximation is advantageous because piecewise linear representations of the power spectrum can be rapidly evaluated analytically. However, we separately need to verify that the piecewise linear representation is smooth enough to permit this approximation. This assumption is evaluated in Section 4.3 and Figure 5 for a fiducial piecewise linear power spectrum.

ORCID iDs

Michael W. Eastwood https://orcid.org/0000-0002-4731-6083
Jayce Dowell https://orcid.org/0000-0003-1407-0141
Danny C. Price https://orcid.org/0000-0003-2783-1608
Frank K. Schinzel https://orcid.org/0000-0001-6672-128X
Harish Vedantham https://orcid.org/0000-0002-0872-181X

References

Ali, Z. S., Parsons, A. R., Zheng, H., et al. 2015, ApJ, 809, 61
Anderson, E., Bai, Z., Dongarra, J., et al. 1990, in Proc. 1990 ACM/IEEE Conf. on Supercomputing, Supercomputing '90 (Los Alamitos, CA, USA: IEEE Computer Society), 2

Anderson, M. M., Hallinan, G., Eastwood, M. W., et al. 2018, ApJ, 864, 22

Baars, J. W. M., Genzel, R., Pauliny-Toth, I. I. K., & Witzel, A. 1977, A&A, 61, 99

Barkana, R. 2018, Natur, 555, 71

Barry, N., Hazelton, B., Sullivan, I., Morales, M. F., & Pober, J. C. 2016, MNRAS, 461, 3135

Beardsley, A. P., Hazelton, B. J., Sullivan, I. S., et al. 2016, ApJ, 833, 102

Bharadwaj, S., & Ali, S. S. 2005, MNRAS, 356, 1519

Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., & Mahesh, N. 2018, Natur, 555, 67

Datta, K. K., Choudhury, T. R., & Bharadwaj, S. 2007, MNRAS, 378, 119

DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017, PASP, 129, 045001

Dillon, J. S., Liu, A., & Tegmark, M. 2013, PhRvD, 87, 043005

Dowell, J., & Taylor, G. B. 2018, ApJL, 858, L9

Eastwood, M. W., Anderson, M. M., Monroe, R. M., et al. 2018, AJ, 156, 32

Ewall-Wice, A., Chang, T.-C., Lazio, J., et al. 2018, ApJ, 868, 63

Ewall-Wice, A., Dillon, J. S., Hewitt, J. N., et al. 2016, MNRAS, 460, 4320

Ewall-Wice, A., Dillon, J. S., Liu, A., & Hewitt, J. 2017, MNRAS, 470, 1849

Fialkov, A., Barkana, R., & Cohen, A. 2018, PhRvL, 121, 011101

Fialkov, A., Barkana, R., Pinhas, A., & Visbal, E. 2014, MNRAS, 437, L36

Fialkov, A., Barkana, R., Visbal, E., Tseliakhovich, D., & Hirata, C. M. 2013, MNRAS, 432, 2909

Field, G. B. 1958, PIRE, 46, 240

Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, PhR, 433, 181

Gehlot, B. K., Mertens, F. G., Koopmans, L. V. E., et al. 2018, arXiv:1809.06661

Greig, B., & Mesinger, A. 2017, MNRAS, 472, 2651

Greig, B., & Mesinger, A. 2018, MNRAS, 477, 3217

Haslam, C. G. T., Klein, U., Salter, C. J., et al. 1981, A&A, 100, 209

Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1982, A&AS, 47, 1

Helmboldt, J. F., & Kassim, N. E. 2009, AJ, 138, 838

Hicks, B. C., Paravastu-Dalal, N., Stewart, K. P., et al. 2012, PASP, 124, 1090

Hills, R., Kulkarni, G., Meerburg, P. D., & Puchwein, E. 2018, Natur, 564, E32

Jelić, V., de Bruyn, A. G., Mevius, M., et al. 2014, A&A, 568, A101

Kaurov, A. A., Venumadhav, T., Dai, L., & Zaldarriaga, M. 2018, ApJL, 864, L15

Kocz, J., Greenhill, L. J., Barsdell, B. R., et al. 2015, JAI, 4, 1550003

Koopmans, L., Pritchard, J., Mellema, G., et al. 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14), ed. T. L. Bourke et al. (Trieste: PoS), 1

Mellema, G., Koopmans, L. V. E., Abdalla, F. A., et al. 2013, ExA, 36, 235

Mitchell, D. A., Greenhill, L. J., Wayth, R. B., et al. 2008, ISTSP, 2, 707

Morales, M. F., Hazelton, B., Sullivan, I., & Beardsley, A. 2012, ApJ, 752, 137

Mozdzen, T. J., Bowman, J. D., Monsalve, R. A., & Rogers, A. E. E. 2017, MNRAS, 464, 4995

Offringa, A. R., McKinley, B., Hurley-Walker, N., et al. 2014, MNRAS, 444, 606

Paciga, G., Albert, J. G., Bandura, K., et al. 2013, MNRAS, 433, 639

Padmanabhan, N., Seljak, U., & Pen, U. L. 2003, NewA, 8, 581

Parsons, A. R., Pober, J. C., Aguirre, J. E., et al. 2012, ApJ, 756, 165

Patil, A. H., Yatawatta, S., Koopmans, L. V. E., et al. 2017, ApJ, 838, 65

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13

Price, D. C., Greenhill, L. J., Fialkov, A., et al. 2018, MNRAS, 478, 4193

Pritchard, J. R., & Loeb, A. 2012, RPPh, 75, 086901

Readhead, A. C. S., & Wilkinson, P. N. 1978, ApJ, 223, 25

Salvini, S., & Wijnholds, S. J. 2014, A&A, 571, A97

Santos, M. G., Cooray, A., & Knox, L. 2005, ApJ, 625, 575

Shaw, J. R., Sigurdson, K., Pen, U.-L., Stebbins, A., & Sitwell, M. 2014, ApJ, 781, 57

Shaw, J. R., Sigurdson, K., Sitwell, M., Stebbins, A., & Pen, U.-L. 2015, PhRvD, 91, 083514

Singh, S., Subrahmanyan, R., Udaya Shankar, N., et al. 2018, ApJ, 858, 54

Taylor, G. B., Carilli, C. L., & Perley, R. A. (ed.) 1999, in ASP Conf. Ser. 180, Synthesis Imaging in Radio Astronomy II (San Francisco, CA: ASP)

Tegmark, M. 1997, PhRvD, 55, 5895

Thyagarajan, N., Jacobs, D. C., Bowman, J. D., et al. 2015, ApJ, 804, 14

Vedantham, H., Udaya Shankar, N., & Subrahmanyan, R. 2012, ApJ, 745, 176

Venumadhav, T., Chang, T.-C., Doré, O., & Hirata, C. M. 2016, ApJ, 826, 116

Venumadhav, T., Oklopčić, A., Gluscevic, V., Mishra, A., & Hirata, C. M. 2017, PhRvD, 95, 083010

Wouthuysen, S. A. 1952, AJ, 57, 31