## CHIPS: The COSMOLOGICAL H i POWER SPECTRUM ESTIMATOR

# CHIPS: THE COSMOLOGICAL H I POWER SPECTRUM ESTIMATOR

C. M. Trott[1,2], B. Pindor[2,3], P. Procopio[2,3], R. B. Wayth[1,2], D. A. Mitchell[2,4], B. McKinley[2,3], S. J. Tingay[1,2], N. Barry[5], A. P. Beardsley[5,6], G. Bernardi[7,8,9], Judd D. Bowman[6], F. Briggs[2,10], R. J. Cappallo[11], P. Carroll[5], A. de Oliveira-Costa[12], Joshua S. Dillon[12], A. Ewall-Wice[12], L. Feng[12], L. J. Greenhill[9], B. J. Hazelton[5,13], J. N. Hewitt[12], N. Hurley-Walker[1], M. Johnston-Hollitt[14], Daniel C. Jacobs[6], D. L. Kaplan[15], H. S. Kim[2,3], E. Lenc[2,16], J. Line[2,3], A. Loeb[9], C. J. Lonsdale[11], M. F. Morales[5], E. Morgan[12], A. R. Neben[12], Nithyanandan Thyagarajan[6], D. Oberoi[17], A. R. Offringa[2,18], S. M. Ord[1,2], S. Paul[19], J. C. Pober[5,20], T. Prabu[19], J. Riding[2,3], N. Udaya Shankar[19], Shiv K. Sethi[19], K. S. Srivani[19], R. Subrahmanyan[2,19], I. S. Sullivan[5], M. Tegmark[12], R. L. Webster[2,3], A. Williams[1], C. L. Williams[12], C. Wu[21], and J. S. B. Wyithe[2,3]

[1] International Centre for Radio Astronomy Research, Curtin University, Perth, WA 6845, Australia; cathryn.trott@curtin.edu.au
[2] ARC Centre of Excellence for All-sky Astrophysics (CAASTRO), Redfern, NSW, Australia
[3] The University of Melbourne, School of Physics, Parkville, VIC 3010, Australia
[4] CSIRO Astronomy and Space Science (CASS), P.O. Box 76, Epping, NSW 1710, Australia
[5] University of Washington, Department of Physics, Seattle, WA 98195, USA
[6] Arizona State University, School of Earth and Space Exploration, Tempe, AZ 85287, USA
[7] Square Kilometre Array South Africa (SKA SA), Park Road, Pinelands 7405, South Africa
[8] Department of Physics and Electronics, Rhodes University, Grahamstown 6140, South Africa
[9] Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA
[10] Australian National University, Research School of Astronomy and Astrophysics, Canberra, ACT 2611, Australia
[11] MIT Haystack Observatory, Westford, MA 01886, USA
[12] MIT Kavli Institute for Astrophysics and Space Research, Cambridge, MA 02139, USA
[13] University of Washington, eScience Institute, Seattle, WA 98195, USA
[14] Victoria University of Wellington, School of Chemical & Physical Sciences, Wellington 6140, New Zealand
[15] University of Wisconsin–Milwaukee, Department of Physics, Milwaukee, WI 53201, USA
[16] The University of Sydney, Sydney Institute for Astronomy, School of Physics, NSW 2006, Australia
[17] National Centre for Radio Astrophysics, Tata Institute for Fundamental Research, Pune 411007, India
[18] Netherlands Institute for Radio Astronomy (ASTRON), P.O. Box 2, 7990 AA Dwingeloo, The Netherlands
[19] Raman Research Institute, Bangalore 560080, India
[20] Brown University, Department of Physics, Providence, RI 02912, USA
[21] International Centre for Radio Astronomy Research, University of Western Australia, Crawley, WA 6009, Australia

## ABSTRACT

Detection of the cosmological neutral hydrogen signal from the Epoch of Reionization (EoR) and estimation of its basic physical parameters are principal scientific aims of many current low-frequency radio telescopes. Here we describe the Cosmological H I Power Spectrum Estimator (CHIPS), an algorithm developed and implemented with data from the Murchison Widefield Array, to compute the two-dimensional and spherically-averaged power spectrum of brightness temperature fluctuations. The principal motivations for CHIPS are the application of realistic instrumental and foreground models to form the optimal estimator, thereby maximizing the likelihood of unbiased signal estimation, and allowing a full covariant understanding of the outputs. CHIPS employs an inverse-covariance weighting of the data through the maximum likelihood estimator, thereby allowing use of the full parameter space for signal estimation ("foreground suppression"). We describe the motivation for the algorithm, implementation, application to real and simulated data, and early outputs. Upon application to a set of 3 hr of data, we set a $2\sigma$ upper limit on the EoR dimensionless power at $k = 0.05$ h Mpc$^{-1}$ of $\Delta_k^2 < 7.6 \times 10^4$ mK$^2$ in the redshift range $z = [6.2\text{–}6.6]$, consistent with previous estimates.

*Key words:* Astronomical instrumentation, methods and techniques – early universe – methods: statistical – techniques: interferometric

## 1. INTRODUCTION

Detection of a neutral hydrogen signal from the Epoch of Reionization (EoR) and estimation of its basic physical parameters are primary science goals of current and future low-frequency radio telescopes, e.g., the Murchison Widefield Array (MWA;[22] Lonsdale et al. 2009; Tingay et al. 2013), the Precision Array for Probing the Epoch of Reionization (PAPER;[23] Parsons et al. 2010); the Low Frequency Array (LOFAR;[24] van Haarlem et al. 2013), the Long Wavelength Array[25] (Ellingson et al. 2009), and the Hydrogen Epoch of Reionization Array.[26] The neutral hydrogen signal during the EoR corresponds to brightness temperature fluctuations in the gas and traces conditions within the intergalactic medium through a suite of radiative and collisional processes (Furlanetto et al. 2006). The spatial structure of the signal, as a function of redshift, traces the evolution of the first ionizing sources of radiation in the universe (e.g., Mesinger et al. 2011, 2014; Fialkov et al. 2014; Pacucci et al. 2014; Pober et al. 2014; Watkinson & Pritchard 2014). These studies also predict

[22] http://www.mwatelescope.org
[23] http://eor.berkeley.edu
[24] http://www.lofar.org
[25] http://lwa.unm.edu
[26] http://reionization.org

the signal amplitude to be weak (10 s mK), compared with other sources in the sky (100 s K), and with radiometric noise associated with the internal electronics of the antennas and receiver systems. The detection experiment itself is therefore difficult, and the estimation experiment more so.

Given the current lack of a detection of the signal, and the lack of an instrument with sufficient sensitivity to directly image the brightness temperature fluctuations, research is focused on its statistical detection (Liu & Tegmark 2011; Dillon et al. 2013; Patil et al. 2014; Ali et al. 2015; Jacobs et al. 2015, and references therein). This has the twofold advantage of increasing signal strength by integrating over large areas of the sky and providing a global statistical indication of the signal (as compared with a local sample obtained by imaging a small patch of sky). The variance (power) of the temperature fluctuations is used as a statistical metric in most studies, with many computing the variance as a function of spatial scale on the sky (the power spectrum). The probability distribution function of the signal is expected to follow a zero-mean Gaussian distribution at early times and a skewed Gaussian distribution at late times (Furlanetto et al. 2006; Wyithe & Morales 2007). The variance, therefore, captures a lot of the information about the signal. Although the power spectrum has a well-defined mathematical form, factors such as the weakness and complexity of the signal, complexity of the instrumental response function (e.g., frequency- and direction-dependent [DD] antenna beam response, bandpass response), and presence of structured, contaminating foregrounds demand a careful approach to robustly demonstrate that the power spectrum of *cosmological* temperature fluctuations has been measured. Thus, there are a suite of approaches being undertaken by different research groups and instruments (Chapman et al. 2014; Choudhuri et al. 2014; Patil et al. 2014; Dillon et al. 2015; B. Hazelton et al. 2015, in preparation; Jacobs et al. 2015). In this work, we describe one of the algorithms being developed for detection of the EoR signal and computation of a two-dimensional (2D) and spherically-averaged (1D) power spectrum for the MWA. Other papers will describe the other algorithms under development (Paul et al. 2014; Dillon et al. 2015; B. Hazelton et al. 2015, in preparation). Each has a different approach to computing the same metric with the same data, allowing robust cross-referencing and benchmarking of results. Jacobs et al. (2015) provide a high-level description of the MWA EoR project and algorithms. The key motivations for the estimator we describe here are (1) inclusion of a full instrument and sky description in the statistical estimator; (2) working directly with measured visibility data, where the data description is computationally tractable; and (3) use of a maximum likelihood (ML) estimator to form the optimal estimate with full covariant error analysis.

In Section 2 we briefly describe the MWA and pertinent design properties that affect how the EoR signal is computed. In Section 3 the power spectrum is formally defined, and recent upper limits on EoR detection using the power spectrum are reviewed. We then introduce the motivation for the Cosmological H I Power Spectrum Estimator (CHIPS) algorithm and the mathematical basis of its computation. Section 5.1 describes the CHIPS approach to foregrounds and derives the expected signal from foregrounds in the power spectrum parameter space

(i.e., "the wedge"). The algorithm is then tested with realistic simulations in Sections 6. Section 7 then describes the calibration process and the observations used in this work. CHIPS is then applied to MWA data in Section 8.

Throughout we use a $\Lambda$CDM cosmology with $H_0 = 70.4 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$, $\Omega_M = 0.27$, $\Omega_k = 0$, $\Omega_\Lambda = 0.73$ (Bennett et al. 2012). The discrete Fourier transform (DFT) convention used is such that

$$\tilde{f}_k = \frac{1}{N} \sum_{j=1}^{N} f_j \exp(-2\pi i jk/N), \qquad (1)$$

for the forward transform. Vector quantities are displayed in bold italics ($\boldsymbol{r}$), Fourier-space quantities have tildes ($\tilde{f}$), Fourier-space matrices are boldfaced ($\boldsymbol{C}$), and mean values are overlined ($\overline{S}$). The notation $\tilde{S} \sim \mathcal{CN}(\mu, \boldsymbol{C})$ compactly describes a variable that is statistically distributed as a complex-normal with mean $\mu$ and covariance $\boldsymbol{C}$.

## 2. THE MURCHISON WIDEFIELD ARRAY

The MWA (Lonsdale et al. 2009; Tingay et al. 2013) is an aperture-array low-frequency radio telescope operating in Western Australia in the 80–300 MHz frequency range (instantaneous available bandwidth of 30.72 MHz). Its science themes include EoR detection, discovery and monitoring of radio transients and variables, a southern survey covering $3\pi$ sr (Wayth et al. 2015), and solar and heliospheric science (Bowman et al. 2013). The array consists of 128 tiles, each of which comprises 16 dual-polarization crossed dipole antennas in a regular $4 \times 4$ grid, with an effective area per tile $A_{\mathrm{eff}} \simeq 21 \, \mathrm{m}^2$ (150 MHz).

The antennas are connected to analog beamformers and the signals correlated within an on-site building. Correlated visibilities in full instrumental polarization flow in real time to a computing center located in Perth. The simplicity of the array design and lack of mechanical steering yield a low-cost instrument well suited to the remoteness and climatic conditions of the desert, but determine the instrumental response to signal from the sky. In particular, the regular grid of dipoles operating as a phased array yields a frequency-dependent beam response with grating lobes (beam nulls), and the analog beamforming enforces a discretized grid of pointings on the sky where the beam is well behaved. These design features demand that the position- and frequency-dependent beam response is known and that this knowledge must be incorporated into any estimator (because the beam is the "window" through which the sky signal is observed).

The 30.72 MHz of instantaneous bandwidth is natively captured at 10 kHz spectral resolution, but the standard correlation modes yield 40 kHz channels. Each of these fine channels is contained within 1.28 MHz coarse channels (24 across the band), with substantial attenuation between coarse channels. Typically, two fine channels are flagged with zero data in these nulls between the coarse channels, leading to a bandpass shape with a regular spacing of zeroes. Again, this design feature needs to be correctly accounted for in the analysis and prevents use of some standard techniques.

## 3. METHODOLOGY

### 3.1. The Power Spectrum Metric

The power spectral density (power spectrum) measures the spatial covariance of a signal, integrated over a spatial volume, and corresponds to the Fourier transform of the two-point correlation function (i.e., the autocorrelation function), $\xi(r)$:

$$
\begin{aligned}
P(\boldsymbol{k}) &= \int_V \xi(\boldsymbol{r}) \exp(-2\pi i \boldsymbol{k} \cdot \boldsymbol{r}) d\boldsymbol{r}, \\
&= \int_V \langle T(\boldsymbol{r}_1) T(\boldsymbol{r}_1 + \boldsymbol{r}) \rangle_{\boldsymbol{r}_1} \exp(-2\pi i \boldsymbol{k} \cdot \boldsymbol{r}) d\boldsymbol{r}, \\
&= \frac{1}{V} \int_V T(\boldsymbol{r}_1) T(\boldsymbol{r}_1 + \boldsymbol{r}) \exp(-2\pi i \boldsymbol{k} \cdot \boldsymbol{r}) d\boldsymbol{r} d\boldsymbol{r}_1, \\
&= \frac{1}{V} \tilde{T}(\boldsymbol{k}) \tilde{T}^*(\boldsymbol{k}),
\end{aligned}
\tag{2}
$$

where $T(r)$ is the temperature fluctuation (relative to the mean) at vector position $r$, and $\langle\rangle$ denotes an average over positions in the volume, $V$, as a proxy for an ensemble sampling of the distribution. The power spectrum has the units of temperature squared times volume ($K^2 Mpc^3$) and describes the integrated power on a given spatial scale, $\boldsymbol{k}$, averaged over the volume, $V$.

The power spectrum can be computed directly from an image cube, using a three-dimensional Fourier transform of image intensities to find $\tilde{S}(\boldsymbol{k})$, and then squaring and normalizing by the cube volume, where $S(\boldsymbol{x})$ is the measured brightness temperature in units of Jy beam$^{-1}$. The conversion from brightness temperature to flux density is given by

$$
S_{\mathrm{Jy}} = 10^{26} \frac{2k T_K}{\lambda^2} \Omega \, \mathrm{Jy},
\tag{3}
$$

which is the product of the specific intensity and the observation solid angle, $\Omega$, and $k$ is the Boltzmann constant.

Alternatively, the power spectrum can be computed directly from observed interferometric visibilities. The visibility is the mutual coherence of the two electrical signals detected by the antennas forming a baseline (Thompson et al. 1986). In the flat-sky approximation, where the field of view (FOV) of the instrument is small such that sky curvature can be ignored, the visibility is identically the Fourier transform of the product of the sky signal and the beam response. For wide FOV instruments, such as the MWA, this approximation breaks down, and there is a curvature convolving kernel in the measured visibility, in addition to the sky signal and beam response. Thyagarajan et al. (2015a, 2015b) studied this effect with MWA data, but current estimators, in general, ignore this effect. The visibility is the natural measurement space of the instrument (the radiometric noise is uncorrelated between visibilities) and will be used in CHIPS to compute the power spectrum directly (without tranforming to and from image space). Curvature sky terms are handled explicitly in CHIPS, and this design choice offers a natural departure from other, image-based power spectrum estimators (Liu & Tegmark 2011; Patil et al. 2014; Paul et al. 2014; Dillon et al. 2015; B. Hazelton et al. 2015, in preparation), but has been used for the angular power spectrum (Choudhuri et al. 2014). A related estimator, the "delay spectrum" (Parsons et al. 2012, 2014, and references therein), directly Fourier transforms along each visibility's frequency channels, matching a temporal delay with the position of given sky emission relative to the phase center.

This visibility-based estimator approximates the power spectrum on large angular scales, but breaks down for the longer baselines.

The spherically-averaged (1D) power spectrum computes the power on three-dimensional spatial scales, $k = |\boldsymbol{k}|$, under the assumption that the statistics of the signal are isotropic and translationally invariant (the latter being a fundamental assumption of the power spectrum, according to the Wiener–Khinchin theorem). The 1D power spectrum has the advantage of integrating over the largest parameter space, potentially increasing signal detectability. The dimensionless 1D power spectrum, which integrates the total power on a given spatial scale over the volume, is given by

$$
\Delta^2(k) = \frac{k^3}{2\pi^2} P(k).
\tag{4}
$$

PAPER has published the most competitive 1D limits to date, relative to theoretical expectations, yielding $(22.4 \, \mathrm{mK})^2$ at $z = 8.4$ and $0.15 \, h \, \mathrm{Mpc}^{-1} < k < 0.5 \, h \, \mathrm{Mpc}^{-1}$ (Ali et al. 2015). Patil et al. (2014) have demonstrated the power of the LOFAR variance statistic, which computes the overall variance in each spectral channel, with simulated data.

It is advantageous, however, to compute the 2D power spectrum as a first step, to discriminate between continuum contaminating foreground sources and the cosmological spectral-line signal. The structure of foregrounds in 2D space is described in Section 5.1; however, we comment here that it is substantially different from the cosmological signal. The 2D power spectrum has arguments $(k_\perp, k_\parallel)$, where $k_\perp = \sqrt{k_x^2 + k_y^2}$ resides in the plane of the sky (angular scales) and $k_\parallel \propto \eta$ is the line-of-sight component. Here $\eta$ is the Fourier dual of frequency, where we use the fact that the observed frequency of the neutral hydrogen spectral line is linearly proportional to distance. Following Morales & Hewitt (2004) and McQuinn et al. (2006), the transformations between Fourier dimensions and cosmological coordinates are

$$
k_\perp = \frac{2\pi |\boldsymbol{u}|}{D_M(z)},
\tag{5}
$$

$$
k_\parallel = \frac{2\pi H_0 f_{21} E(z)}{c(1+z)^2} \eta,
\tag{6}
$$

where $D_M(z), H_0, f_{21}, z$ are the transverse comoving distance, Hubble constant, rest frequency of the hyperfine transition ($\sim 1420 \, \mathrm{MHz}$), and observation redshift, respectively. $E(z)$ is (Hogg 2000)

$$
E(z) \stackrel{\text{def}}{=} \sqrt{\Omega_M(1+z)^2 + \Omega_k(1+z)^2 + \Omega_\Lambda}.
\tag{7}
$$

### 3.2. Motivation for CHIPS Algorithm

The high-level design philosophy of CHIPS is to use the data in such a way as to allow for computationally efficient and robust parameter estimation and error covariance estimation, while retaining maximum information. As such, the following design features have been employed:

1. Calibrated visibilities will be the primary data input—visibility space is the natural measurement space of interferometric data, where radiometric (stochastic) noise is uncorrelated and Gaussian distributed with color dependent on the frequency-dependent system

temperature. In addition, correlations due to point-spread function sidelobes inherent in image space are naturally accounted for in visibility space, where the measured information is clearly defined.

2. A least-squares spectral analysis (LSSA) method will be used to compute the line-of-sight transform from frequency to spectral space, not the Fourier transform (Kay 1993)—the MWA bandpass has regular missing channels, owing to the coarse bandpass edges, and intermittent missing channels owing to flagged radio frequency interference (RFI; Offringa et al. 2015). The Fourier transform cannot be used for irregular and missing data. LSSA can compute the optimal spectral representation of the data using all of the available information.

3. An ML estimate of the cosmological signal (a quadratic estimator; Kay 1993; Liu & Tegmark 2011; Dillon et al. 2013) will be computed (an inverse-covariance weighting)—the ML solution asymptotes to the optimal solution for large quantities of data and is appropriate for the data set acquired here. It retains the full information contained within the data and allows for computation of the full uncertainty covariance matrix. (Note that this is not the only estimator that can be used efficiently to approach this problem.)

4. Foregrounds will be modeled a priori and included in the estimator ("foreground suppression")—as discussed in Section 5.1, foregrounds can be approached in different ways, depending on the degree of knowledge one assumes about them (Liu & Tegmark 2011; Dillon et al. 2013; Chapman et al. 2014; Bonaldi & Brown 2015). Here we take a two-tier approach, where known foregrounds (sky model of point and extended sources) are subtracted from the data coherently, and the remaining signal is treated statistically. CHIPS therefore uses the full Fourier space for estimation (contrast with "foreground avoidance" techniques, where the contaminated regions are excised).

## 4. MATHEMATICAL FORMALISM

The derivation shown here describes the formation of the power estimate in two steps:

1. First, it introduces the computation of the coherently averaged data from measured visibilities, including the incorporation of the frequency-dependent beam shape, curvature terms (*w*-terms), and the transforming of the visibilities in frequency to line-of-sight wavenumber ($\eta$). This step is effectively the visibility gridding (Section 4.1) and the Fourier-like transform to $(u, v, \eta)$ wavenumber space (achieved using LSSA, in Section 4.2). The data set, described by $S(u, v, \nu)$, is a function of location on the *uv*-plane and frequency. Practically, it is obtained by gridding measured visibilities onto the *uv*-plane with the beam kernel (and the *w*-kernel) with a weighting that is a function of the system temperature for that observation and the visibility weight ($\propto T_{\text{sys}}/W$, where $W$ is the visibility weight and corresponds to the number of fine frequency channels averaged to form the visibility). The beam-weighted weights are also accumulated and included in the data covariance matrix, $\boldsymbol{C}$.

2. Second, the coherent information forms the power according to an ML estimate, using knowledge about the data covariance matrix (stochastic noise and foreground contaminants, Section 4.3).

This process is now described in detail.

### 4.1. Coherent Visibility Data: Angular Modes

We aim to take the large number of measured, calibrated visibilities and grid them onto a regular grid representing the *uv*-plane, in each frequency channel. The true sky signal is shaped by the antenna primary beam, and the measured visibilities sample a range of modes. Each datum represents information from a region of the *uv*-plane, according to the beam, and we grid a measurement across this region. For an instrument with linear crossed-polarization feeds, *xx* and *yy*, we model the correlator output as

$$
\begin{bmatrix} \tilde{V}_{xx} \\ \tilde{V}_{xy} \\ \tilde{V}_{yx} \\ \tilde{V}_{yy} \end{bmatrix} = \begin{bmatrix} \tilde{B}_{xx} \\ \tilde{B}_{xy} \\ \tilde{B}_{yx} \\ \tilde{B}_{yy} \end{bmatrix}
$$
$$
* \begin{pmatrix} g_{xx} & g_{xx}\cos 2\psi & g_{xx}\sin 2\psi & 0 \\ g_{xy}\Delta_{xy} & -g_{xy}\sin 2\psi & g_{xy}\cos 2\psi & i \\ g_{yx}\Delta_{yx} & -g_{yx}\sin 2\psi & g_{yx}\cos 2\psi & -i \\ g_{yy} & -g_{yy}\cos 2\psi & -g_{yy}\sin 2\psi & 0 \end{pmatrix} \begin{bmatrix} \tilde{I} \\ \tilde{Q} \\ \tilde{U} \\ \tilde{V} \end{bmatrix}, \quad (8)
$$

which maps the Stokes visibilities in sky coordinates, $(\tilde{I}, \tilde{Q}, \tilde{U}, \tilde{V})$, to the measurement set in instrumental coordinates (Hamaker et al. 1996). Here we are explicitly describing the relationship in the flat-sky approximation, although the curvature terms are introduced below. The $4 \times 4$ matrix encodes projection effects from the parallactic angle, $\psi$, polarization leakage, ($\Delta_{xy}, \Delta_{yx}$), and the direction-independent, antenna-based complex gains, ($g_{xx}, g_{xy}, g_{yx}, g_{yy}$). The convolution accounts for the primary beam shape, ($\tilde{B}_{xx}, \tilde{B}_{xy}, \tilde{B}_{yx}, \tilde{B}_{yy}$). To extract the total intensity, $\tilde{I}$, we could use all four polarization outputs from the correlator. Instead, to reduce computational load, we use only $V_{xx}$ and $V_{yy}$. This choice is made at the expense of some signal when the zenith angle is large. Failure to treat individual linear feeds will lead to polarization leakage if the visibilities are combined without consideration for beam shape differences (Moore et al. 2013).

Absorbing the complex gains into the beams, $\boldsymbol{B}$, we write the *xx* and *yy* visibilities as

$$
\tilde{V}_{xx}(u, v, w) = (\tilde{I} + \tilde{Q}\cos 2\psi + \tilde{U}\sin 2\psi)
$$
$$
* \tilde{B}_{xx} * \int \frac{\exp[-2\pi i (\boldsymbol{D} \cdot \boldsymbol{s})]}{\sqrt{1 - l^2 - m^2}} dl dm \quad (9)
$$

$$
\tilde{V}_{yy}(u, v, w) = (\tilde{I} - \tilde{Q}\cos 2\psi - \tilde{U}\sin 2\psi)
$$
$$
* \tilde{B}_{yy} * \int \frac{\exp[-2\pi i (\boldsymbol{D} \cdot \boldsymbol{s})]}{\sqrt{1 - l^2 - m^2}} dl dm, \quad (10)
$$

where

$$
\boldsymbol{D} \cdot \boldsymbol{s} = ul + vm + w(\sqrt{1 - l^2 - m^2} - 1)) \quad (11)
$$

is the projection of the baseline vector onto the sky coordinates, and we have now explicitly included the *w*-terms in the equations through the phase integral (final term).

Using the convolution theorem, the mapping from the underlying sky Fourier representation and the measured visibilities can be written as

$$\tilde{V}_{xx} = \tilde{G}_w * \tilde{B}_{xx} * (\tilde{I} + \tilde{Q}\cos 2\psi + \tilde{U}\sin 2\psi) \tag{12}$$

$$\tilde{V}_{yy} = \tilde{G}_w * \tilde{B}_{yy} * (\tilde{I} - \tilde{Q}\cos 2\psi - \tilde{U}\sin 2\psi), \tag{13}$$

where

$$\tilde{G}_w \overset{\text{def}}{=} \int \frac{\exp[-2\pi i(w(\sqrt{1 - l^2 - m^2} - 1))]}{\sqrt{1 - l^2 - m^2}}$$
$$\times \exp[-2\pi i(ul + vm)]dldm \tag{14}$$

explicitly highlights the additional convolution due to curvature terms. These functions encode the deviation from a strict 2D Fourier transform between sky and visibilities and the spectral leakage due to the primary beam.

We now expand from considering a single visibility to a set and describe these as a vector. In doing so, we extend the convolutions to describe the transform from the underlying sky, $\tilde{I}$, to the full measurement set of data. We can write the discrete convolutions as matrix operations, to encode this transformation, and find the measurements for a set of baselines:

$$\tilde{V}_{xx} = G_w B_{xx}(\tilde{I} + \tilde{Q}\cos 2\psi + \tilde{U}\sin 2\psi) \tag{15}$$

$$\tilde{V}_{yy} = G_w B_{yy}(\tilde{I} - \tilde{Q}\cos 2\psi - \tilde{U}\sin 2\psi). \tag{16}$$

We are interested in determining the underlying Stokes *I* sky distribution at a given angular scale, $u_\alpha$, given the measured visibilities, $\tilde{V}(u)$. A vector of ($N \times 1$) measured visibilities, $\tilde{V}$, is generated from an underlying ($M \times 1$) sky distribution, $\tilde{I}$, via an ($N \times M$) matrix transform, $G_w B$. We can rewrite and combine Equations (15) and (16) such that

$$\tilde{I}(u, v, \nu) = \frac{1}{2}(B_{xx}^\dagger G_w^\dagger G_w B_{xx})^{-1} B_{xx}^\dagger G_w^\dagger \tilde{V}_{xx}$$
$$+ \frac{1}{2}(B_{yy}^\dagger G_w^\dagger G_w B_{yy})^{-1} B_{yy}^\dagger G_w^\dagger \tilde{V}_{yy}, \tag{17}$$

where the square matrix inversion involves inverting an ($M \times M$) Hermitian complex-valued matrix, which will be (almost) diagonal for independent modes, $u$. Equation (17) effectively unwraps the effects of *w*-terms and the polarized beams, before combining *xx* and *yy* information to remove the polarized component of the sky signal ($\tilde{Q}$, $\tilde{U}$). The $G^\dagger$ matrix serves to distribute the footprint of the visibilities across the *uv*-plane, as dictated by the amplitude of the *w*-term; for large *w*, the Fourier-space beam is effectively broader and the *G* matrix captures this. In practice, as discussed by Dillon et al. (2013), the *uv*-plane needs to be fully sampled (the matrix full rank) for this matrix inverse to exist (i.e., there needs to be independent information in all modes). For the MWA and using rotation synthesis, this requirement is met for parts of the *uv*-plane, but not in general.

There are multiple approaches to handling this, including computation of the pseudoinverse and simplification of the matrix by considering only diagonal components. The former inverts the matrix in regions where there is sufficient information. The latter effectively ignores the correlations introduced by the finite extent of the beam across the *uv*-plane and reduces the matrix inverse step to a simple weighting of the data according to the measurements (averaging). It does, however, still encode the distribution of power across multiple modes in the *uv*-plane. The approximation is reasonable given that the matrix is already highly diagonal. In the testing phases of CHIPS, as described here, we will implement the latter:

$$\tilde{S}(u, v, \nu) = \frac{1}{2}(B_{xx}^\dagger G_w^\dagger G_w B_{xx})_{\text{diag}}^{-1} B_{xx}^\dagger G_w^\dagger \tilde{V}_{xx}$$
$$+ \frac{1}{2}(B_{yy}^\dagger G_w^\dagger G_w B_{yy})_{\text{diag}}^{-1} B_{yy}^\dagger G_w^\dagger \tilde{V}_{yy}. \tag{18}$$

This introduces an error, for each polarization, with the *XX* polarization expression:

$$\tilde{S}_{XX}(u, \nu) = R_{XX}\tilde{I}_{XX}(u, \nu). \tag{19}$$

Here the error matrix, $R_{XX}$, is

$$R_{XX} \overset{\text{def}}{=} (\text{diag}(B^\dagger B))_{XX}^{-1}(B^\dagger B)_{XX}. \tag{20}$$

The same applies to the *YY* polarization, where the error matrix, in general, is different, depending on the shape of the beams for each telescope pointing. In general, this can invalidate the canceling of the *Q* and *U* linear polarization signals in Equation (17) because there is now the potential for mismatch between the two polarizations. In practice, the *XX* and *YY* polarizations are not combined in CHIPS at this point (the results described in this work remain in the two instrumental poarizations) because the current MWA beam model is known to be imperfect. In the future, this mismatch will need to be further studied for leakage effects.

### 4.2. Least-squares Spectral Analysis

Upon gridding the data onto the *uv*-plane for each frequency channel, the next step is to estimate the line-of-sight wavenumber information, by transforming the data at each grid point from frequency, $\nu$, to $\eta$. For a complete and regularly sampled set of $N$ complex-valued discrete data points embedded within white Gaussian noise, $\tilde{S}(\nu)$, the information contained within a spectral mode, $\eta$, can be obtained using the DFT:

$$\mathcal{S}(u, v, \eta) = \frac{1}{\sqrt{2\pi}} \sum_{j=0}^{N-1} \tilde{S}(u, v, \nu_j)\exp[-2\pi i j\eta/N], \tag{21}$$

where $\eta \in [0, N-1]$. This can be written in matrix formalism, where the Fourier transform has convenient properties as a Vandermonde matrix:

$$\tilde{S}(\eta) = \mathcal{F}^\dagger \tilde{S}(\nu), \tag{22}$$

$$\mathcal{F}^\dagger \mathcal{F} = \mathcal{I}. \tag{23}$$

Here $\tilde{S}(\eta)$ is the complex-valued estimate of the spectral information in mode $\eta$, and $\mathcal{F}$ is a square, Hermitian matrix containing the trigonometric kernel. We have also explicitly dropped the parameterization by *u*, *v* because the transform is performed *for each* point on the *uv*-plane, with the understanding that these parameters are carried along for the computations presented in this section.

When the data have generalized covariance (correlated samples, unequal weightings), the signal in some spectral

mode can be estimated using a generalized ML Fourier transform. If the data are distributed such that

$$\tilde{S}(\nu) \sim \mathcal{CN}(\overline{S}(\nu), C), \quad (24)$$

the optimal estimate is given by (i.e., the inverse-covariance weighting estimator; Liu & Tegmark 2011; Dillon et al. 2015)

$$\tilde{S}(\eta) \sim \mathcal{CN}(\mathcal{F}^{\dagger} C^{-1}\tilde{S}(\nu), \mathcal{F}^{\dagger} C \mathcal{F}), \quad (25)$$

where we have used the fact that the Fourier transform is unitary ($\mathcal{F}^{\dagger} = \mathcal{F}^{-1}$). Effectively, this prewhitens the signal (by suppressing data with a lot of noise and removing correlation) and computes the Fourier transform. The result has noise properties consistent with a prewhitening operation and with the reordering of the data according to the summing of phased data points.

Finally, when the data have incomplete sampling, the above formalism naturally generalizes and is described by the LSSA. When there are only $M$ sampled points at locations $\nu_j$, within $N$ frequency channels, the generalized least-squares estimate is

$$\tilde{S}(\eta) = \mathcal{H}\tilde{S}(\nu) \quad (26)$$

$$\stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} \sum_{j=0}^{M-1} \tilde{S}(\nu_j) \exp[-2\pi i \nu_j \eta/M], \quad (27)$$

where $\eta = [0, ..., N_{\text{freq}}]$ are evenly spaced and $N_{\text{freq}} < N$. The optimal estimator is

$$\tilde{S}(\eta) \sim \mathcal{CN}((\mathcal{H}^{\dagger}C^{-1}\mathcal{H})^{-1}\mathcal{H}^{\dagger}C^{-1}\tilde{S}(\nu), (\mathcal{H}^{\dagger}C^{-1}\mathcal{H})^{-1}), \quad (28)$$

where the LSSA operator has been labeled $\mathcal{H}$. Note that now $M < N$ and the $\mathcal{H}$ matrix is not square. In general, this method leads to correlated information between different modes, but this can be reduced by estimating $N_{\text{freq}} \lesssim M/2$ modes (Vio et al. 2013).

### 4.3. Likelihood Function and ML Estimate

Now that we have described the computation of the coherently averaged (gridded and LSSA) data from the underlying measured visibilities, we can write the likelihood function of the data to compute the power estimates.

We describe the data with a generalized multivariate Gaussian, with zero mean, and general covariance matrix, $C$. This form describes data that are drawn from a statistical distribution of possible universes, with temperature fluctuations that are Gaussian distributed. This covariance matrix contains all of the terms contributing to the signal in a visibility. The joint likelihood for the vector of complex-valued coherently averaged data, $S(u, v, \eta)$, is given by

$$L(\tilde{S}; C) = \frac{1}{\pi^N \det(C)} \exp[-\tilde{S}^{\dagger}C^{-1}\tilde{S}], \quad (29)$$

where the complex-valued covariance matrix is

$$C \stackrel{\text{def}}{=} \langle \tilde{S}\tilde{S}^{\dagger} \rangle \quad (30)$$

$$= C_{\text{FG}} + C_{\text{N}} + C_{\text{P}}, \quad (31)$$

where

$$C_{\text{P}} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{12} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{pmatrix} \quad (32)$$

describes the parameters of interest (the mode powers, $p_{\alpha\alpha}$) and the correlations between powers imprinted by the instrument and the experiment ("window functions"). Strictly, the cosmological information forms a diagonal matrix, with uncorrelated power between modes, but the imperfect data correlate contiguous modes. In this implementation of CHIPS, we assume a diagonal cosmological signal matrix, but do encode the correlations between modes in the data by capturing and propagating the covariances introduced by the line-of-sight transform (LSSA). These correlations are important when combining modes together to perform the binning to 1D (see Section 4.3.1). Note that here we are explicitly identifying the power in which we are interested with the likelihood function of the coherently averaged data. The other terms in Equation (31) describe the statistical contribution to mode $(u, v, \eta)$ from foregrounds ($C_{\text{FG}}$) and measurement (radiometric) noise ($C_{\text{N}}$). In $uv$ space, measurement noise is uncorrelated and Gaussian distributed (colored Gaussian noise [CGN]), where the "color" for a coherently averaged datum describes the weighting of the noise due to the number of visibilities contributing to that cell. The statistical foreground contribution will be treated in Section 5.1.

The ML estimate, obtained by finding the value of a given parameter that maximizes the likelihood function (or minimizes the negative log likelihood), is asymptotically efficient (achieves optimal estimation precision for large data sets). We aim to determine an expression for the parameters of interest (mode powers) in terms of the data and data covariance matrix. The log likelihood function is minimized by differentiating with respect to the parameter, setting to zero, and solving for that parameter. The derivative of the log likelihood for a zero-mean generalized Gaussian is given by

$$\frac{\partial \ln L}{\partial p_{\alpha}} = \text{tr}\left(C^{-1}\frac{\partial C}{\partial p_{\alpha}}\right) - \tilde{S}^{\dagger}(u, v, \eta)C^{-1}\frac{\partial C}{\partial p_{\alpha}}C^{-1}\tilde{S}(u, v, \eta), \quad (33)$$

where "tr" denotes the trace of the matrix. For clarity, we now drop the dependence of the data on $(u, v, \eta)$, with the understanding that we are working with the data in wavenumber space. Setting Equation (33) to zero, and then using in the first term the identity $\mathbb{I} = C^{-1}C$ and expanding the covariance matrix into its constituent parts, we find

$$\hat{p}_{\alpha} \text{tr}\left(C^{-1}C^{-1}\frac{\partial C}{\partial p_{\alpha}}\right) \stackrel{\text{def}}{=} \text{tr}\left(C_{\text{P}}C^{-1}C^{-1}\frac{\partial C}{\partial p_{\alpha}}\right) \quad (34)$$

$$= \tilde{S}^{\dagger}C^{-1}\frac{\partial C}{\partial p_{\alpha}}C^{-1}\tilde{S} - \text{tr}\left((C_{\text{FG}} + C_{\text{N}})C^{-1}C^{-1}\frac{\partial C}{\partial p_{\alpha}}\right), \quad (35)$$

where $\hat{p}_{\alpha} = \hat{p}_{\alpha\alpha}$ denotes our *estimate* of the power in mode $\alpha$, as described in the diagonal elements of the cosmological power matrix, Equation (32). Note here that the first term on the right-hand side of Equation (35) is squaring the coherently averaged data and therefore producing the desired power-like quantity.

Therefore, the estimate of the power in mode $\alpha$ is given by computation of

$$\hat{p}_\alpha = \frac{1}{\mathrm{tr}\left(C^{-1}C^{-1}\frac{\partial C}{\partial p_\alpha}\right)}\left(\tilde{S}^\dagger C^{-1}\frac{\partial C}{\partial p_\alpha}C^{-1}\tilde{S}\right.$$
$$\left. - \mathrm{tr}\left((C_{\mathrm{FG}} + C_{\mathrm{N}})C^{-1}C^{-1}\frac{\partial C}{\partial p_\alpha}\right)\right), \quad (36)$$

where we decouple dependence of the estimator on the mode powers by making the approximation

$$C \approx C_{\mathrm{FG}} + C_{\mathrm{N}}, \quad (37)$$

under the assumption that $C_{\mathrm{FG}} + C_{\mathrm{N}} \gg C_{\mathrm{P}}$.

In practice, the data can be split into two and cross-correlated. This removes thermal noise power from the final power estimate, at the expense of a factor of $\sqrt{2}$ in the final signal-to-noise ratio (S/N) (in temperature; a factor of 2 in power). Although noise power has been removed from the estimate, thermal noise *uncertainty* remains in the final covariance of the power estimates. Typically, alternate correlator output data sets are processed into alternate coherent data vectors, to ensure uniform *uv* coverage between the two (this is important for applying the same data covariance matrix to both data sets). Using this scheme, the cross-correlation power formed from coherent data vectors, $_1S$ and $_2S$, is

$$\hat{p}_\alpha = \frac{1}{\mathrm{tr}\left(C^{-1}C^{-1}\frac{\partial C}{\partial p_\alpha}\right)}\left(_1S^\dagger C^{-1}\frac{\partial C}{\partial p_\alpha}C^{-1}{}_2S\right.$$
$$\left. - \mathrm{tr}\left(C_{\mathrm{FG}}C^{-1}C^{-1}\frac{\partial C}{\partial p_\alpha}\right)\right). \quad (38)$$

The expected values and covariances are derived in the [Appendix](#) and demonstrate that the expected value of the ML estimate yields the cosmological power.

The second term in Equation ([38](#)) is the foreground power bias. As discussed in a number of recent papers (Liu & Tegmark [2011](#); Dillon et al. [2013](#), [2015](#)), subtraction of this term requires an accurate model for the foregrounds. Inaccurate subtraction leads to bias in the power. Given the relative power of the foreground and cosmological signal, mis-subtraction will occur when the foreground power is incorrect at the fraction of a percent level. To avoid this, we retain the foreground power bias and allow the noise covariance to indicate which modes are contaminated (i.e., the second term is set to zero).

### 4.3.1. Explicitly Including LSSA

To connect the LSSA transform with the ML estimate, we can take the expression for the optimal power, Equation ([38](#)), which is a function of $(u, v, \eta)$, and insert the LSSA estimate of $\tilde{S}(u, v, \eta)$ from the underlying gridded data, $\tilde{S}(u, v, \nu)$. To do so, we combine Equations ([28](#)) and ([38](#)) to compute the optimal (ML) power estimate:

$$\hat{p}_\alpha = \frac{(\mathcal{H}^\dagger C^{-1}{}_1\bar{S}(\nu))^\dagger \frac{\partial C}{\partial p_\alpha}(\mathcal{H}^\dagger C^{-1}{}_2\bar{S}(\nu))}{\mathrm{tr}\left((\mathcal{H}^\dagger C^{-1}\mathcal{H})\frac{\partial C}{\partial p_\alpha}(\mathcal{H}^\dagger C^{-1}\mathcal{H})\right)}. \quad (39)$$

The denominator (the normalization) reduces to

$$\sum_i |A_{i\alpha}|^2, \quad (40)$$

where $A_{ij}$ are the elements of the inverse-covariance matrix,

$$A \stackrel{\mathrm{def}}{=} \mathcal{H}^\dagger C^{-1}\mathcal{H}. \quad (41)$$

This then yields the sum over the square of the matrix elements for the normalization. Finally, the vector of power estimates is $\chi^2$-distributed:

$$\hat{p}(\eta) \sim \chi^2(p, 2((\mathcal{H}^\dagger C^{-1}\mathcal{H})(\mathcal{H}^\dagger C^{-1}\mathcal{H}))^{-1}), \quad (42)$$

where the (identity) diagonal cosmological signal matrix, $dC/dp$, has been dropped.

To compute the spherically-averaged (1D) power spectrum from the 2D output, we average in elongated cylindrical shells, allowing for differences in the line of sight and angular modes for the cosmological signal in redshift space in the conversion from $u$, $v$, $\eta$ to $k$ (Mpc$^{-1}$). The ML estimate of the 1D power is given by weighting the individual 2D mode powers contributing to a cell, $k$, with the inverse of their power covariance matrix. This is the inverse of the covariance term from Equation ([42](#)), with

$$\mathcal{A} \stackrel{\mathrm{def}}{=} (\mathcal{H}^\dagger C^{-1}\mathcal{H})(\mathcal{H}^\dagger C^{-1}\mathcal{H}). \quad (43)$$

Thus,

$$P_k = \frac{\sum_{i \in k} \mathcal{A}_{,i}p}{\mathrm{tr}(D^\dagger \mathcal{A}_{,i}D)}, \quad (44)$$

where the covariance matrix of data is used quadratically to reflect that we are now working to optimally average *power*, and $D$ is a binning matrix that reduces the full 2D space to the subspace spanned by $i \in k$. The numerator terms here come from performing the matrix multiplication from Equation ([42](#)) and including all of the covariances between 2D modes, $i$, contributing to binned mode $k$. The denominator provides the variances and covariances between the 1D $k$ modes, encoding the degree of uncertainty on each parameter and the relationship between estimates. The syntax, $_{,i}$, denotes the matrix, $\mathcal{A}$, as formed from projecting onto the bin, including parameter covariances.

## 5. BUILDING THE DATA COVARIANCE MATRIX

### 5.1. Foregrounds

Dealing with foreground contamination can be approached in different ways, all of which rely on the basic distinction that foreground sources are continuum emitters and therefore have a smooth spectrum over a small bandwidth, and the cosmological signal is a spectral line with structure reflecting the brightness temperature at different spatial depths. This distinction fundamentally dissociates angular modes from line-of-sight modes and is the reason for operating with the 2D power spectrum. In 2D, the power from flat-spectrum foregrounds that could be observed through a perfect instrument (no frequency-dependent instrumental response, no spectral incompleteness, and complete sampling of the Fourier plane) would be contained within the DC ($\eta = 0$) mode of $k_\parallel$ because the flat-spectrum foregrounds add a simple amplitude across frequency. For a real instrument and nonflat

sources, foregrounds occupy a broader region of parameter space, colloquially termed the "wedge" because of its broad wedge-like structure in $(k_\perp, k_\parallel)$ space. The wedge has been well studied (Datta et al. 2010; Morales et al. 2012; Parsons et al. 2012; Trott et al. 2012; Vedantham et al. 2012; Hazelton et al. 2013; Thyagarajan et al. 2013). It results from the incomplete sampling of the Fourier modes and/or spectral modes by an interferometer and the migration of the angular mode sampled by a given baseline as a function of observation frequency ("mode mixing"). A mathematical derivation of the expected structure of the wedge for the MWA is included in this section.

Broadly, there are two primary foreground treatment design options employed in the literature: (1) "avoidance," where it is attempted to contain foregrounds to a region of parameter space and ignore this region (e.g., PAPER's approach; Parsons et al. 2012; Jacobs et al. 2015); and (2) "suppression," where the foreground contribution to the signal is modeled (using an a priori source model, nonparametric methods, or a data-driven model) and the estimator uses this knowledge to suppress contaminated information. The latter option includes the nonparametric fitting of low-order polynomials (Bowman et al. 2009) and the more sophisticated Component Analysis techniques discussed by Chapman et al. (2012, 2014), Bonaldi & Brown (2015), and references therein. It also includes model- and data-driven (parametric) statistical descriptions of the sky itself and incorporation of that information (Liu & Tegmark 2011; Dillon et al. 2013, 2015). These approaches all have their own advantages and disadvantages. Broadly, nonparametric approaches are simple to implement but, requiring no input physical knowledge, can destroy cosmological information and retain foreground signal. Typically, the metric for an adequate solution is not well defined. Conversely, parametric models are designed to include as much physical information about the foregrounds and cosmological signal as is available. However, they can be computationally difficult to implement, are limited by our knowledge of the sky at low radio frequencies, and can also destroy information if the models are incorrect. Surveys such as the MWA's GLEAM (Wayth et al. 2015) and LOFAR's MSSS (Heald et al. 2015) will be crucial for improving our understanding.

Ultimately, the key to a robust estimator is to have an understanding of the limitations and biases of one's method and incorporate that knowledge into the methodology and results. CHIPS aims to do this and handles foregrounds using a two-tiered approach: (1) known foregrounds (sky model of point and extended sources, with updated calibration solutions and ionospheric corrections) are subtracted from the data coherently (visibility data) using the Real-Time System (RTS) calibration (see Section 7); (2) remaining signal is handled statistically, using an a priori model for the expected distribution and spectral structure of sources (CHIPS). We then use a perturbative calculation to ascertain our degree of confidence in the model (Liu et al. 2015).

We initially employ a two-component statistical foreground model, consisting of extragalactic point sources and Galactic synchrotron, described here. The component describing the Galactic plane is omitted. The Galactic plane is more difficult to describe statistically, with definite sky position and skewed statistics. We leave this as an open problem for future work.

### Point-source Covariance Matrix

We consider the additional noise-like signal contained within a visibility due to unmodeled point sources present within the primary beam (where the Poisson-distributed number of sources within any differential patch of sky yields the variant, noise-like signal). To compute the contribution of these sources, we perform the following calculation. The number of sources within a small area of the sky is assumed to be Poisson distributed. For a Poisson distribution, the variance is equal to the mean.

1. Calculate the Poisson noise due to the random number of sources within a small patch of sky and a small range of source flux density ($N(S, S + dS; l, l + dl; m, m + dm)$):

$$N(S, S + dS; l, l + dl; m, m + dm) = \frac{dN}{dS} dS\, dl\, dm, \quad (45)$$

where $dN/dS$ is the source density per unit flux density and is given parametrically by

$$\frac{dN}{dS}(\nu) = \alpha \left(\frac{S_{\rm Jy}}{S_0}\right)^{-\beta} \text{Jy}^{-1}\,\text{sr}^{-1}. \quad (46)$$

For this work we use the number counts of Intema et al. (2011), with $\alpha = 4100$ Jy$^{-1}$ sr$^{-1}$ and $\beta = 1.59$ at 150 MHz.

2. Compute the variance on a visibility measurement due to the Poisson noise from a differential patch of sky at $(l', m')$; for $N$ sources with flux density $S(\nu)$ located at sky position $(l', m')$ within the beam $B(l, m, \nu)$, the mean visibility is given by

$$\langle V(u, v) \rangle = NSB(l', m', \nu)$$
$$\times \exp[-2\pi i(ul' + vm')]. \quad (47)$$

3. Compute the total variance within a visibility from all sources (in the simple model where there is no source clustering, covariance matrices sum because sources are independent); the total covariance between two visibilities on the same baseline (different frequencies) can be computed by considering the response of the instrument at different frequencies, for a fixed location in the $uv$-plane. The source number density, primary beam, and $uv$-sampling all change, with the latter (the physical source of the wedge) being encapsulated in the term $f_\nu = (\nu'' - \nu')/\nu_{\rm low}$, which produces the frequency phase wrapping that yields the wedge feature in power. Here $\nu_{\rm low}$ is the lowest measurement frequency channel, and $\nu'$ and $\nu''$ denote different frequencies within the band. The covariance is:
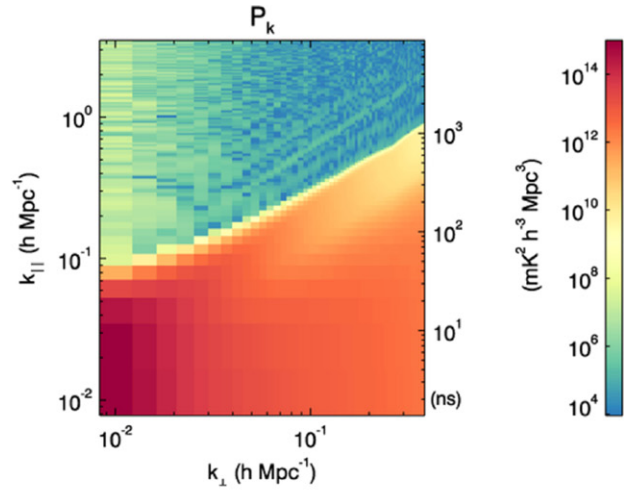
$$\boldsymbol{C}_{\rm PS} = \iiint S^2 \left(\frac{\sqrt{\nu''\nu'}}{\nu_{\rm low}}\right)^{-\gamma} B(l, m; \nu'') B(l, m; \nu')$$

$$\times \frac{dN}{dS} dS\, dl\, dm \quad (48)$$

$$= \frac{\alpha}{3 - \beta} \left(\frac{\sqrt{\nu''\nu'}}{\nu_{\rm low}}\right)^{-\gamma} \frac{S_{\rm max}^{3-\beta}}{S_0^{-\beta}} \iint B(\boldsymbol{l}; \nu'') B(\boldsymbol{l}; \nu')$$

$$\times \exp[-2\pi i(\boldsymbol{u} \cdot \boldsymbol{l}) f_\nu] dl\, dm \text{ Jy}^2, \quad (49)$$

where $S_{\rm max}$ is the brightest unmodeled source in the field (the peeling limit), here taken as 1 Jy. This expression
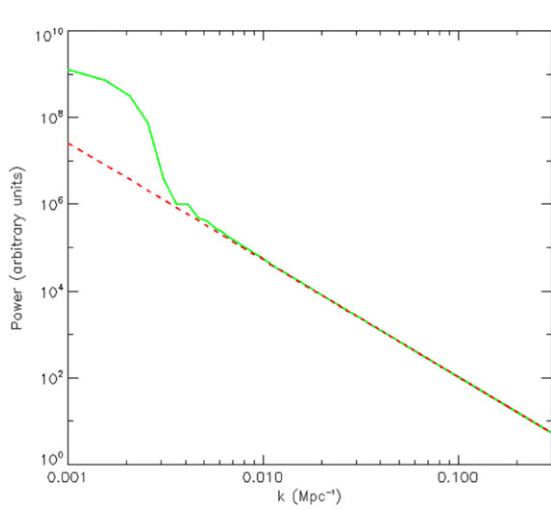
(a) Beam-convolved Galactic synchrotron component.



(b) Foreground model.

**Figure 1.** (a) Intrinsic (red dashed line) and beam-convolved (green solid line) power spectra for the Galactic synchrotron model. (b) Power spectrum for the input foreground covariance model, including contributions from extragalactic point sources and Galactic synchrotron.

simplifies for a circularly symmetric beam, where the 2D Fourier transform can be written as a 1D Hankel transform:

$$C_{\rm PS} = \frac{\alpha}{3-\beta}\left(\frac{\sqrt{\nu''\nu'}}{\nu_{\rm low}}\right)^{-\gamma}\frac{S_{\rm max}^{3-\beta}}{S_0^{-\beta}}$$
$$\times \int_0^\infty B(l;\nu'')B(l;\nu')J_0(2\pi(ul)(\nu''-\nu'))l\,dl \ {\rm Jy}^2.$$
(50)

For a top-hat beam truncated at the horizon ($l = 1$), the foreground covariance has a simple form, which is similar to a sinc function:

$$C_{\rm PS} = \frac{\alpha}{3-\beta}\left(\frac{\sqrt{\nu''\nu'}}{\nu_{\rm low}}\right)^{-\gamma}\frac{S_{\rm max}^{3-\beta}}{S_0^{-\beta}}\frac{J_1(2\pi uf_\nu)}{f_\nu u},$$
(51)

where $f_\nu = (\nu''-\nu')/\nu_{\rm low}$ and $u = |x|\nu_{\rm low}/c$. Similarly, a frequency-dependent Gaussian-shaped beam is often a reasonable approximation to the beam shape:

$$B(l;\nu) \propto \exp[-l^2/\sigma^2],$$
(52)

where

$$\sigma(\nu) \simeq \epsilon c/\nu D,$$
(53)

and $\epsilon \simeq 0.42$ and $D \simeq 4$ m are the scalings from an Airy disk to a Gaussian width and the tile diameter, respectively. Using this and the Hankel transform of a Gaussian yields

$$C_{\rm PS} = \frac{\alpha}{3-\beta}\Gamma(\nu'')\Gamma(\nu')\left(\frac{\sqrt{\nu''\nu'}}{\nu_{\rm low}}\right)^{-\gamma}$$
$$\times \frac{S_{\rm max}^{3-\beta}}{S_0^{-\beta}}\frac{\pi c^2\epsilon^2}{D^2}\frac{1}{\nu''^2+\nu'^2}\exp\left(\frac{-u^2c^2f_\nu^2\epsilon^2}{4(\nu''^2+\nu'^2)D^2}\right),$$
(54)

where $\Gamma$ indicates the use of a frequency taper function to reduce spectral leakage (e.g., a Hanning window, as used

here). For a real, frequency-dependent MWA beam, the foreground covariance can be computed numerically.

### *Galactic Synchrotron Covariance Matrix*

Galactic synchrotron emission, from electrons spiraling along our Galaxy's magnetic field lines, produces emission on large scales. We use the parameters of Jelić et al. (2008, 2010) to motivate our model, which also includes the effects of the MWA primary beam and instrumental chromaticity.

The intrinsic temperature power spectrum is modeled as

$$\langle\Delta T_{\rm GS}^2\rangle(u,\nu) = (\eta T_B)^2\left(\frac{u}{u_0}\right)^{-2.7}\left(\frac{\nu}{\nu_0}\right)^{-2.55} {\rm K}^2,$$
(55)

where $\eta = 0.01$ is the fluctuation level relative to the uniform brightness temperature, $T_B = 253$ K, $u_0 = 10$ wavelengths, and $\nu_0 = 100$ MHz is the reference frequency. The covariance matrix is a function of the parameter $u$, to reflect that we expect the statistics to be rotationally invariant. The apparent steep spectral index in temperature units is flattened once converting to flux density (integrated) units for a given instrument, such that the intrinsic power is

$$P_{\rm GS}(u,\nu) = \left(\frac{2k}{\lambda^2}\right)^2\Omega(\eta T_B)^2\left(\frac{u}{u_0}\right)^{-2.7}\left(\frac{\nu}{\nu_0}\right)^{-2.55} {\rm Jy}^2$$
(56)

$$= \left(\frac{2k}{\lambda}\right)^2\frac{1}{A_{\rm eff}}(\eta T_B)^2\left(\frac{u}{u_0}\right)^{-2.7}\left(\frac{\nu}{\nu_0}\right)^{-2.55} {\rm Jy}^2$$
(57)

$$= \left(\frac{(2k)^2}{A_{\rm eff}}\right)(\eta T_B)^2\left(\frac{u}{u_0}\right)^{-2.7}\left(\frac{\nu}{\nu_0}\right)^{-0.55} {\rm Jy}^2.$$
(58)

The truncation of the spectrum due to the primary beam can be obtained via convolution (via a Fourier transform), and Figure 1(a) shows the intrinsic and convolved power spectra. Here the Fourier transform is used (rather than LSSA) because the plot uses complete and regularly sampled spacing for

9

display purposes. The actual covariance contains the same incompleteness as the data, when implemented in CHIPS. Note that the primary beam convolution is frequency dependent, generating complex structure at small wavenumbers.

Finally, the instrumental chromaticity is included in a similar manner as for the point-source model, yielding the following model for the spectral covariance matrix:

$$C_{GS}(\nu, \nu'; u) = 2\pi \sqrt{P'_{GS}(u, \nu) P'_{GS}(u, \nu')}$$
$$\times \int J_0(2\pi(ul)f(\nu)) l \, dl \ \mathrm{Jy}^2, \qquad (59)$$

where $P'$ denotes the beam-convolved power.

### 5.1.1. Total Covariance Matrix

The two foreground components are summed to form the complete foreground contribution to the data covariance matrix, as observed by the instrument. After performing a spectral transform to $\eta$-space and converting to cosmological units, the final total foreground contribution to the power spectrum is shown in Figure 1(b) (coarse bandpass missing channels are omitted for this plot, for clarity). Note the presence of coherent streaks of emission beyond the foreground wedge region but parallel to it. This is caused by spectral structure in the window taper function (in this example, a Blackman–Nuttall taper is used) interacting with the non-smooth shape of the MWA primary beam. Note also the apparent brightening toward the edge of the wedge region. Naively, the wedge shape in the $k_\parallel$ direction is dictated by the primary beam (for a uniform foreground brightness distribution), suggesting that the edge should be attenuated. However, as predicted in Thyagarajan et al. (2013) and demonstrated in Thyagarajan et al. (2015b), wide-field curvature effectively compresses the sky, increasing the density of emission at the edge of the wedge.

### 5.2. Thermal Noise

The final component of the data covariance corresponds to the thermal (measurement, radiometric) noise. This is the measurement uncertainty on each visibility due to the finite number of data samples (information) that contribute to it. We are effectively estimating the sky signal with a fixed amount of information, which depends on the signal strength and the number of samples. The former is set by the system temperature, which is dominated by sky temperature at low frequencies, and the latter is set by the bandwidth and sampling time for each visibility. For a single polarization,

$$\sigma = 10^{26} \frac{2k T_{sys}}{A_{eff}} \frac{1}{\sqrt{\Delta\nu \Delta t}} \ \mathrm{Jy}. \qquad (60)$$

When gridding the visibilities onto the $uv$-plane, as described in Section 4.1, the noise decreases coherently (with square-root improvement). This noise reduction therefore follows the evolution of sampled points in the $uv$-plane for a nightly track of the EoR field. Identical observations on subsequent nights yield the same coherent reduction in power. The thermal noise contribution to the power spectrum therefore maps the distribution of points in the $uv$-plane for a nightly track, thereby reflecting the array configuration.
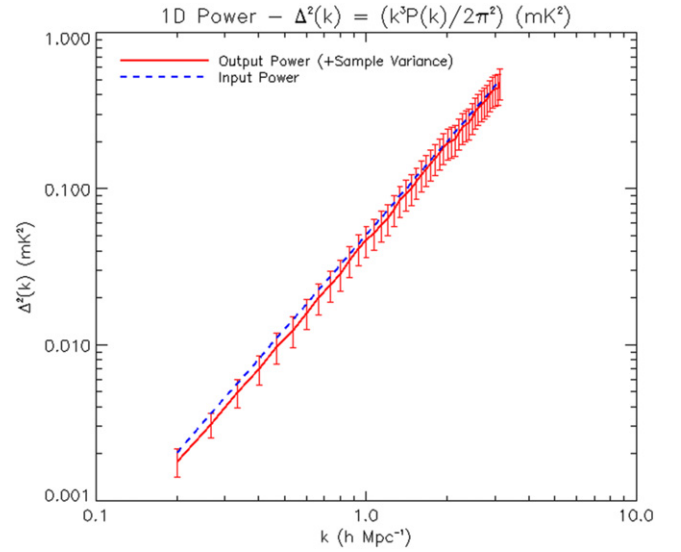


**Figure 2.** Spherically-averaged power spectrum of simulated power spectrum (red solid line) and input spectrum (blue dashed line). The error bars quantify $2\sigma$ uncertainties due to sample variance.

### 6. SIMULATIONS

To test the estimator, we generated a set of end-to-end noise-free simulations of a single 2-minute snapshot of visibilities and passed them through the pipeline, using a power-law input power spectrum. The aims of these simulations were to verify that the slope and normalization were unbiased. We chose the amplitude of the spectrum arbitrarily ($A = 1 \ \mathrm{K}^2$), and the index was set to $n = -1$, where $P(|k|) = A k^n \ \mathrm{K}^2$. To produce realistic visibilities, we included the following in the simulations: (1) the full frequency-dependent primary beam shape of the instrument, (2) the actual $uv$ distribution of a zenith-pointed observation, and (3) curvature of the sky ($w$-terms).

The simulations were produced by starting with an image-cube Gaussian random field of brightness temperature fluctuations ($l, m, \nu$), performing a 3D Fourier transform to $k$-space, multiplying by the square root of the input power spectrum, and inverse Fourier transforming back to the real space-frequency cube. The data were then multiplied by the frequency-dependent beam shape and regridded to a curved coordinate system. Finally, the $uv$-plane for each frequency channel was generated by performing a 2D Fourier transform to ($u, v, \nu$)-space. This cube formed the underlying data from which the visibilities were sampled according to the baseline distribution of the MWA (Beardsley et al. 2013).

Figure 2 displays the computed spherically-averaged power spectrum and input power spectrum, showing agreement between the two within a few percent.

### 7. THE RTS AND OBSERVATIONS

Data are calibrated using the MWA RTS (Mitchell et al. 2008; D. A. Mitchell et al. 2015, in preparation). The RTS has been specifically designed to calibrate MWA data at high cadence, using the full frequency and DD beam response, ionospheric modeling and correction, and multisource in-field calibration using point-source and extended-source models.

MWA EoR observations used in this work were collected in 112 s integrations during which the MWA correlator outputs

raw visibilities every 0.5 s. Data are used from the MWA EOR0 field, centered at R.A. = 0$^h$, decl. = −27°. This sky position is chosen to avoid the Galactic plane and yield a relatively cold sky. RFI was detected and excised using AOFlagger (Offringa et al. 2012). Each such observation is independently calibrated by the RTS in a two-stage process. First, the entire observation is used to determine the direction-independent Jones matrices (complex gains) of each MWA tile by fitting the observed visibilities to a model that consists of the 1000 apparently[27] brightest sources that lie within 20° of the pointing center. This model is constructed from catalogs of known radio sources using the PUMA algorithm (J. Line et al. 2015, in preparation).[28] This method of combining faint sources into a single high-S/N calibrator is similar to source clustering as proposed by Kazemi et al. (2013), although in this case we only construct a single compound source. Following this averaged calibration stage, the sky model sources are subsequently individually passed through the RTS Calibrator Measurement Loop (CML). For each source within the CML, (i) the model visibilities of all other calibrator sources are subtracted, (ii) an ionospheric offset and gain[29] term is measured by fitting a phase ramp to the visibilities when rotated toward the catalog position of the calibrator, and (iii) for the brightest sources DD corrections to the antenna gains are fitted to the residual visibilities following ionospheric correction. In practice, a combination of limited processing time, S/N, and available degrees of freedom limits the number of sources that can receive this full DD treatment. In this work, five sources are treated as full DD calibrators, and the rest are only updated for ionospheric effects as above. For the EOR0 field, which has relatively few bright sources, the DD calibrators have flux densities of ∼10−20 Jy. Once the CML model has converged, the best-fit model of each calibrator source is subtracted from the visibilities. The RTS is parallelized over frequency, so that each of the 24 coarse channels (1.28 MHz) is effectively independently calibrated. One exception to this are the ionospheric offset fits that are fitted to a $\lambda^2$ dependence over the entire bandwidth. The second calibration and subtraction step is performed on an 8 s cadence in order to resolve in time ionospheric fluctuations. Note that for sources subtracted during this step, the requirement that they must lie within 20° of the pointing center is removed so that sufficiently bright sources can be subtracted from anywhere in the sky. As a result, the lists of sources used in the two calibration stages are not identical. Subtracting 1000 sources corresponds to a subtraction threshold of ∼350 mJy at the center of the beam. All calibration operations are performed at 40 kHz frequency resolution, but the residual visibilities are averaged to 80 kHz for power spectrum estimation. Hence, the output from the RTS to the power spectrum estimation module is calibrated, with residual visibilities averaged to 80 kHz and 8 s.

---

[27] As attenuated by the primary beam.

[28] In addition to the positional matching, PUMA uses entries from known catalogs and fits a power law to these data points to assess the reliability of the matching. In practice, each source in the catalog almost always has more than one flux entry. Contiguous entries are then used in the RTS to compute the spectral index for that frequency band. The relatively large frequency intervals between the catalog entries enforce a generally smooth behavior of the sources in the band considered while preserving the overall spectral energy distribution features.

[29] Ostensibly, this term is intended to account for ionospheric attenuation. However, at present it subsumes and it is likely dominated by errors in interpolated catalog source fluxes.

Different calibration and peeling strategies were tested to find the optimal calibration settings for these data. Beyond the usual image-quality and calibration-stability metrics employed to assess calibration performance, we also used the power spectrum to assess the impact of different strategies. Three such modes were (1) self-calibrate with 300 sources, peel 300 sources; (2) self-calibrate with 1000, peel 300 sources; and (3) self-calibrate with 300 sources, peel 1000 sources. Figure 3 shows the ratio of cross power for strategies 2 and 3, relative to strategy 1 (which we expect to leave the most power in the power spectrum), for the 15 zenith-only observations of the data used in this work. Calibrating on more sources leads to a more precise calibration solution and improves the regions of interest for EoR science. Peeling additional sources reduces power in the wedge, as expected for removing signal power. An interesting feature of removing more wedge power is the reduction in power at higher $k_\parallel$, where copies of the wedge caused by harmonics introduced by the regular coarse bandpass missing channels leak power beyond low-$k$ modes. Hence, peeling more sources allows a reduction in foreground power throughout the 2D power spectrum parameter space. Ultimately, a strategy self-calibrating with 1000 sources and peeling 1000 sources was employed for the work presented in this paper.

## 8. RESULTS

### Application to Data

The set of high-band EoR data were taken from a single night of observations during the first semester of MWA EoR observing (2013 August 23). These data were chosen to test, refine, and compare different calibration and analysis methodologies, as described in Jacobs et al. (2015). The original data set consists of 160 112 s observations. These data were refined to a final set of 94 observations, removing pointings that were >25° from zenith and a pointing heavily affected by Galactic emission in the sidelobes. As described in Section 7, the data were calibrated and provided as 8 s visibilities, yielding 14 time steps per observation and seven per interleaved data set per observation (temporally interleaved data are used to compute the cross-power spectrum). After calibration, we performed a uv cut at a maximum distance of 300 wavelengths at the lowest frequency, within which the EoR signal is expected to fall, and used the full bandwidth data set for the initial analysis (30.72 MHz).

The cross-power spectra are produced with CHIPS both with and without the foreground covariance matrix included in the estimator. Without foreground covariance, the data are weighted purely by the system temperature for that observation, the relative weight of the visibility (determined from the number of 10 kHz channels contributing), and the uv-sampling of the instrument and correspond to the most straightforward application of a power spectrum analysis. With foreground covariance, the data are downweighted according to the modeled and measured noise and signal contamination within the data.

Figure 4 shows the instrumental north–south and east–west cross-power polarizations for the full bandwidth. The solid and dashed diagonal lines correspond roughly to the expected locations of foregrounds within the main lobe of the MWA primary beam and the horizon, respectively. There is clear foreground contribution within the expected region. Note that
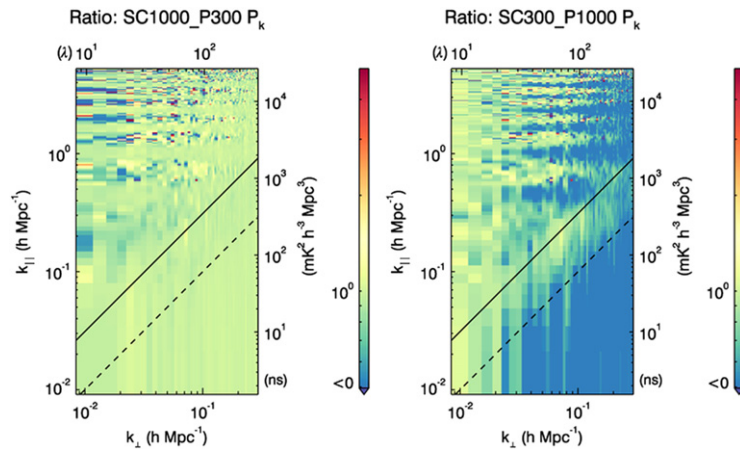
**Figure 3.** Ratio of the power in two calibration and peeling strategies, relative to a standard strategy of self-calibrating with 300 sources and peeling 300 sources, as described in the text. Self-calibrating on more sources (left) produces a more well-calibrated data set and yields improvement in the non-foreground-dominated regions of the power spectrum. Leaving the calibration the same but peeling more sources removes power from the foregrounds, thereby yielding lower power in the wedge and coarse bandpass harmonics. The diagonal dashed and solid lines refer to where flat-spectrum foregrounds would lie if contained within the main lobe of the primary beam and the horizon, respectively.
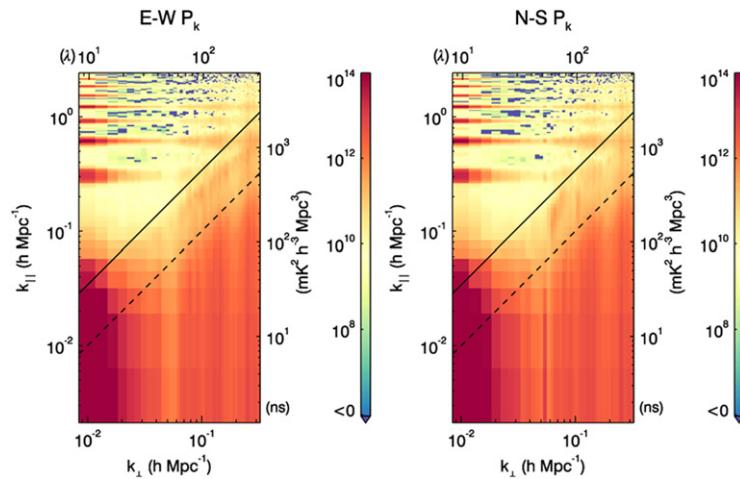


**Figure 4.** E-W (left) and N-S (right) cross-power spectra when foregrounds are *excluded* from the estimator. Both polarizations demonstrate foreground power consistent with smooth point sources and large-scale Galactic emission. The N-S polarization has additional power beyond the theoretical edge of the main primary beam lobe (into the region approaching the horizon) and corresponds to Galactic emission rising or setting during the observations.

this wedge emission is consistent with the input foreground model, as shown in Figure 1. There is also the clear imprint of the MWA's $uv$-sampling function, which has few very short and few very long ($k > 0.08 \, \mathrm{h \, Mpc^{-1}}$) baselines. The copies of the wedge are visible at regular intervals in $k_\parallel$, corresponding to the comb-like bandpass sampling function of the MWA. The errors and expected noise and the measured signal-to-error ratio can also be computed (Figures 5 and 6). The error is computed using Equation (42) for the thermal-noise-only data covariance matrix, as considered for the estimator (foreground data covariance contribution is omitted here). The expected thermal noise is computed by gridding visibilities with a thermal noise contribution given an average system temperature of $T_{\mathrm{sys}} = 240 \, \mathrm{K}$ and propagating through the same estimator. This system temperature was estimated by matching the observed noise (obtained through the difference in the even and odd data sets) to unity-valued gridded visibilities. Note that this is only the thermal noise contribution, and the signal-to-error plot shows all of the high ratio detections expected in the foreground-dominated regions. The signal-to-error plot shows

behavior consistent with thermal noise away from the contaminated wedge and bandpass harmonics regions.

When the foreground contribution is *included* in the data covariance matrix, power is effectively removed from the contaminated regions, and the errors reflect those parts of the parameter space (Figure 7(b)). The errors at the locations of the coarse bandpass features are also elevated, but this is not obvious from the plots as shown. The small signal-to-error ratio in the wedge indicates that these modes are highly contaminated and should be downweighted in the final binning from 2D to 1D. The ratio is close to unity in the wedge, suggesting that the foreground covariance is capturing the contamination appropriately. These data can then be averaged in cylindrical bins and normalized to obtain the spherically-averaged dimensionless power spectrum. Instead of treating the full bandwidth, which corresponds to a redshift range of 6.2–7.5 and therefore is highly likely to contain signal evolution, we follow Dillon et al. (2015) and split the coherent data into three contiguous 10.24 MHz bins, corresponding to redshifts $z = [6.2–6.6]$, $[6.6–7.0]$, and $[7.0–7.5]$, and compute the 1D power spectra for those (Figure 8) for the E-W polarization
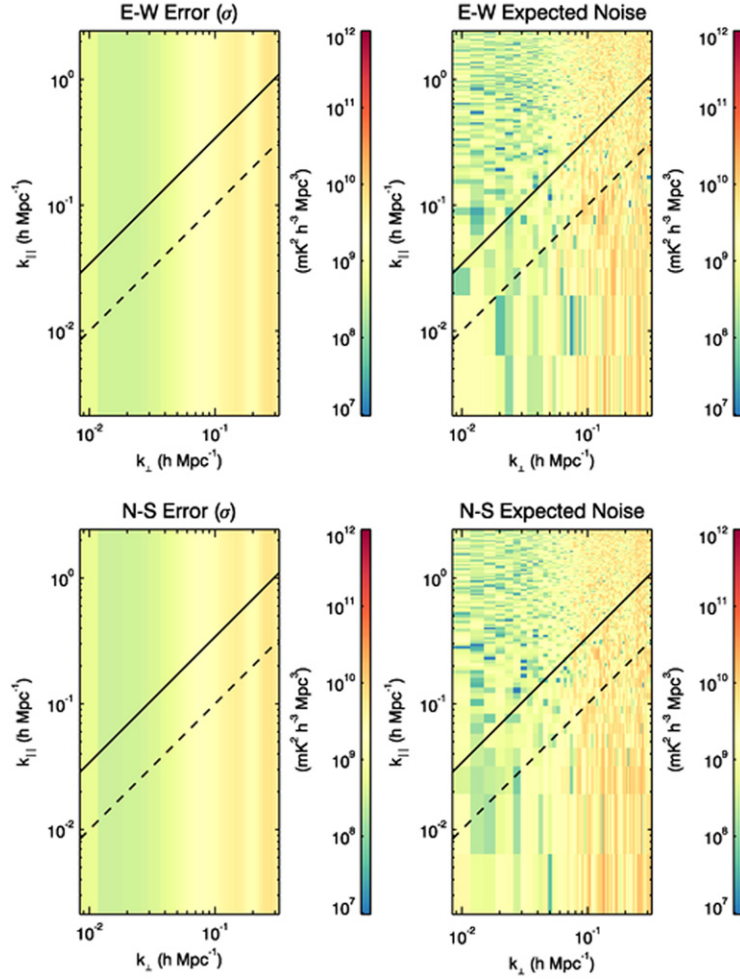
**Figure 5.** Error and expected noise computed from the input visibility weights and a system temperature of 240 K at 170 MHz, for both instrumental polarizations.
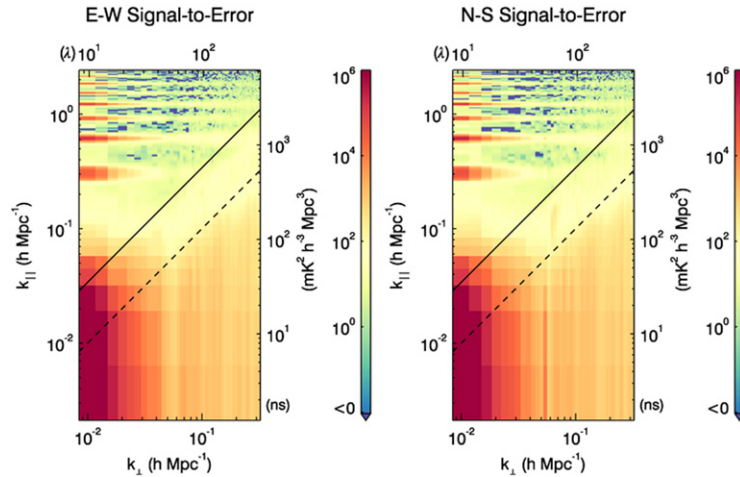


**Figure 6.** Signal-to-error ratio (ratio of cross power to errors) for both polarizations where the error term only includes the thermal noise contribution to the estimator (no foreground contribution). As expected when the foreground contribution is excluded from the estimator, the foreground-dominated wedge and coarse bandpass harmonics show high signal-to-error ratio detections of signal.

(which shows reduced foreground leakage from the Galaxy). This is effectively using the top and bottom plots from Figure 7 as the power and errors, but with the reduced bandwidth (it is not exactly this process, because the inverse-covariance estimator also uses all of the off-diagonal error terms, which are not represented in the 2D error plots). Unlike Dillon et al.

(2015), however, we do not exclude any wedge contribution here and *instead allow the estimator to "work within the wedge"* and downweight contaminated modes. The only data cut performed is to remove the $k_\perp = 0$ and $k_\parallel = 0$ bins, which show large contamination and a poor foreground model response. Both the thermal-noise-only and the full thermal
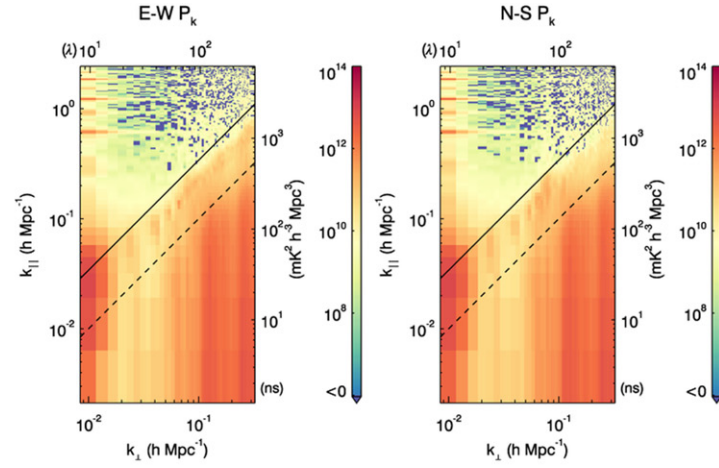
(a) Cross power spectra with the foreground model included in the estimator.
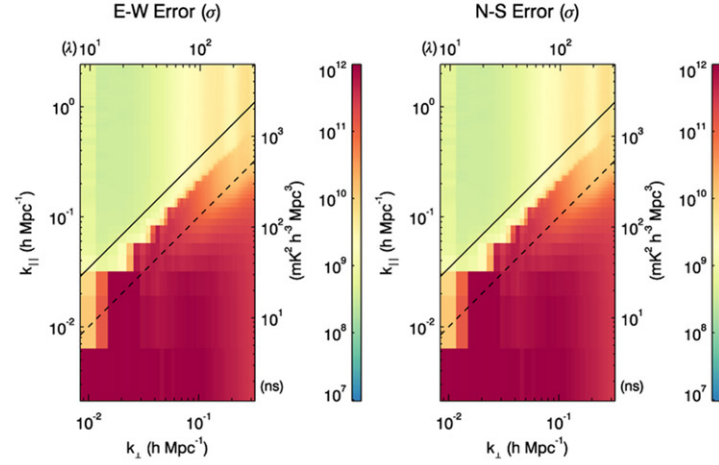


(b) Errors, when the foregrounds are included in the data covariance matrix.
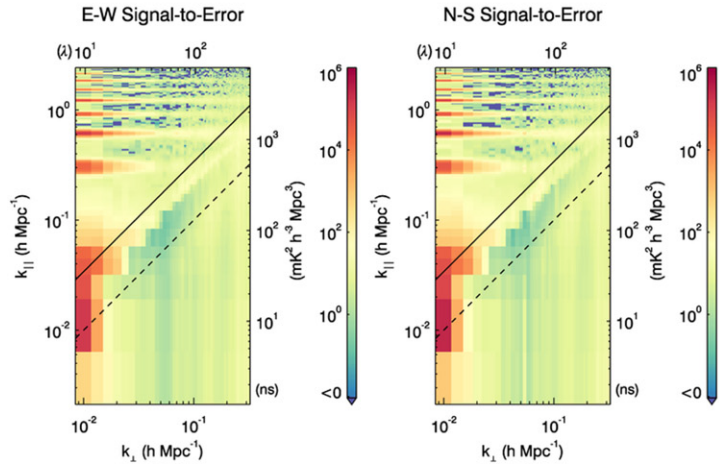


(c) Signal-to-error ratio plot demonstrating the large suppression of modes within the wedge. Compare with Figure 6 where no foregrounds are included in the estimator.

**Figure 7.** Cross-power spectra, errors, and signal-to-error ratios for the two instrumental polarizations, when the foreground model is included in the estimator. The small signal-to-error ratio in the wedge indicates that these modes are highly contaminated and should be downweighted in the final binning from 2D to 1D.

(a) $\Delta z$=6.2–6.6.



(b) $\Delta z$=6.6–7.0.



(c) $\Delta z$=7.0–7.5.
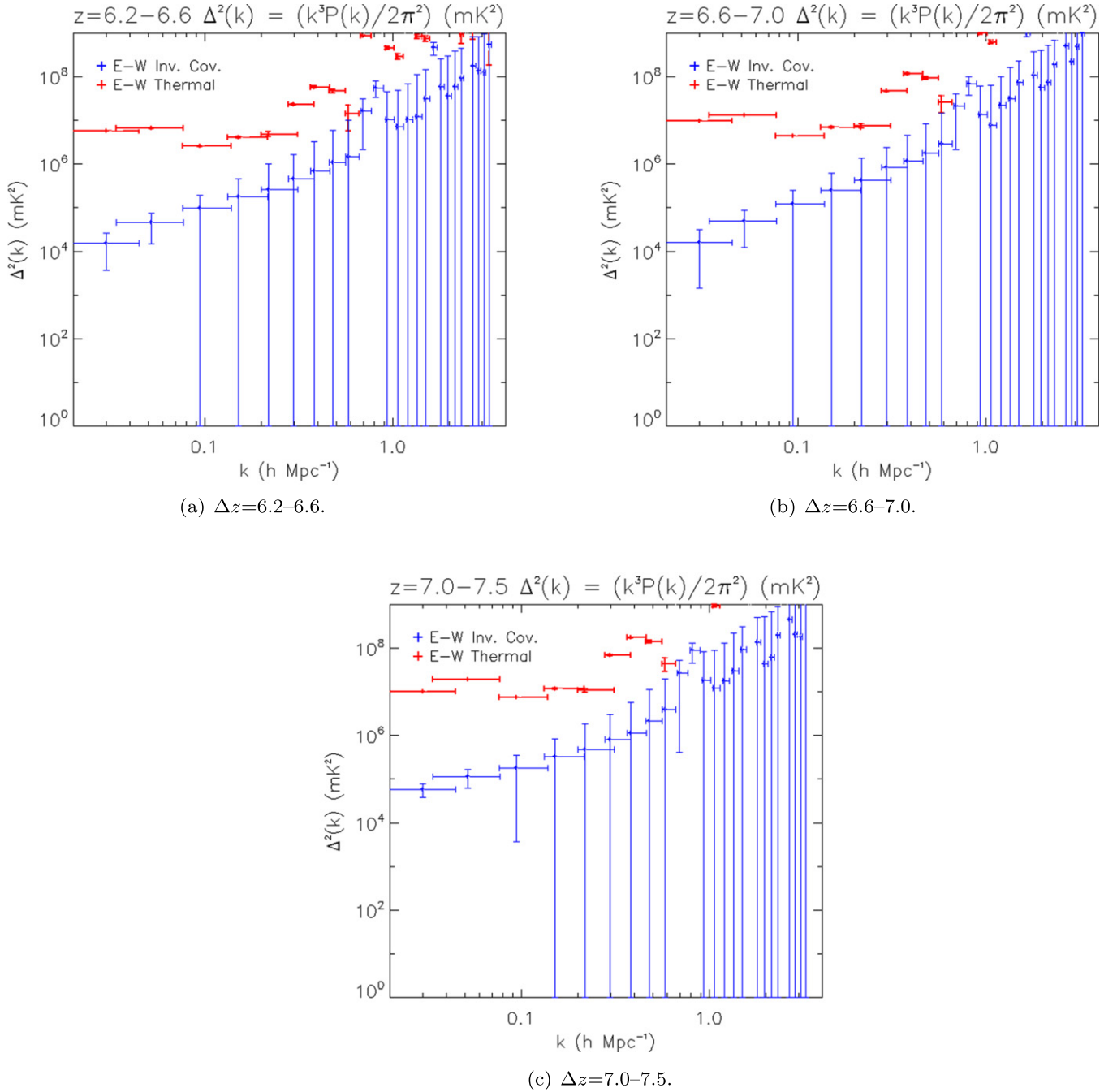
**Figure 8.** 1D spherically-averaged E-W polarization power spectra for three redshift ranges, corresponding to the upper, midddle, and lower 10.24 MHz bands of the data. Both a thermal-noise-only (red) and a full foreground (blue) data covariance model has been used, as well as the full parameter space (no wedge excision).

noise+foreground data covariance matrices are shown. The error bars reflect 95% confidence regions in both dimensions. Inclusion of the foreground model increases the uncertainties on those modes, corresponding to the substantial contamination in those regions. The full data covariance model increases the uncertainties such that the results are consistent with thermal noise+foregrounds across most of parameter space. The "detections" in the $k_{\perp}$ = [0.03–0.10] $h\,\mathrm{Mpc}^{-1}$ region are due to power bias from unmodeled foregrounds, recalling that the power bias term from Equation (36) is omitted from the estimator.

From these results, we can set $2\sigma$ upper limits on the EoR power spectrum at the point where the 1D power is at a

**Table 1**
95% Confidence ($2\sigma$) Upper Limits on the EoR Power in Three Redshift Bins at $k = 0.05$ h $\mathrm{Mpc}^{-1}$, Using the Full CHIPS Estimator and Including All Data

| Redshift | $\Delta_k^2$ (mK$^2$) |
| --- | --- |
| $z = [6.2\text{–}6.6]$ | $7.6 \times 10^4$ |
| $z = [6.6\text{–}7.0]$ | $8.8 \times 10^4$ |
| $z = [7.0\text{–}7.5]$ | $16.5 \times 10^4$ |

minimum and we achieve a "detection," corresponding to $k = 0.05\,h\,\mathrm{Mpc}^{-1}$. (We omit the inner detection because the power associated with this bin includes power on the scale of

the primary beam size). Table 1 lists the upper limits at this wavenumber for the 3 hr of data in the E-W polarization. While not competitive with the results from deeper studies with other telescopes, it is consistent with previously published values, with a best detection value of $\Delta_k^2 = (275 \text{ mK})^2$ in the redshift range $z = [6.2–6.6]$.

## 9. DISCUSSION AND CONCLUSIONS

The CHIPS estimator is one of several EoR power spectrum estimators being developed and applied within the MWA collaboration and broadly among the community with other low-frequency telescopes. It takes one possible approach to the substantial systematic problems of structured and bright foregrounds and complex instrumentation. The primary design principles for CHIPS include (1) a full instrumental model (bandpass, frequency- and pointing-dependent primary beam shape, uv-sampling, observation-dependent system temperature, instrumental chromaticity); (2) a model-driven foreground component to the data covariance, drawing on previous observational studies to extract a realistic statistical model for an extragalactic point-source population and Galactic synchrotron emission; and (3) an ML estimator to tie together the full sky and instrumental information in a consistent and robust framework and yield a fully covariant set of output uncertainties.

This approach is not unique and likely not the full solution to addressing this complex task. In particular, the approach to foregrounds has many options, and others have demonstrated the benefits of empirical (Dillon et al. 2015) and blind parametric and nonparametric (e.g., Chapman et al. 2014) approaches. Ultimately, the approach to foregrounds may require a combination of these techniques. Conversely, an understanding of the full impact of the instrumental and analysis signal chain on the final data product is crucial for complicated low-frequency telescopes. In this regard, a CHIPS-like approach is likely to be required for most current and future instruments.

The results presented here are from a very small amount of data and are meant to be interpreted as a proof of principle for the general approach. The outputs are visually as we would expect given our understanding of the instrument and expectations of system noise, and quantitatively in one dimension they are consistent with previous estimates. In the regions of parameter space that we expect to be thermal noise limited, this small data set shows consistency with those expectations.

The foreground model employed here, while multicomponent, is still very simple and includes none of the intricacies of different underlying extragalactic populations (for example, star-forming galaxies and active galactic nuclei). Being a statistical model, it also lacks the ability to treat any outlier emission that remains after peeling (for example, when peeling sources, the Galactic center and main Galactic plane are not treated in this current calibration strategy, but are very bright), leaving additional power in the output power spectrum that would be contained in the unsubtracted bias term. As we probe deeper into the data, we will use the new information gleaned to form better sky models, but this is ultimately an iterative process at low frequencies, where deep observations are sparse.

We thank the referee for a very thorough reading of the manuscript and many excellent suggestions to improve its

## APPENDIX
## EXPECTED POWERS AND COVARIANCES

To assess the performance of the estimator, we can take the expected value of the estimate, $\langle \hat{p}_\alpha \rangle$, noting that $\langle \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \rangle = \text{tr}(\boldsymbol{A} \boldsymbol{C})$,

$$\langle \hat{p}_\alpha \rangle = \frac{\text{tr}\left( \boldsymbol{C}_P \boldsymbol{C}^{-1} \boldsymbol{C}^{-1} \frac{\partial \boldsymbol{C}}{\partial p_\alpha} \right)}{\text{tr}(\boldsymbol{C}^{-1} \boldsymbol{C}^{-1})} \qquad (61)$$

$$= \frac{\sum_\beta p_\beta |(\boldsymbol{C}^{-1})_{\alpha\beta}|^2}{\sum_\beta |(\boldsymbol{C}^{-1})_{\alpha\beta}|^2}. \qquad (62)$$

The estimate of the power in mode $\alpha$ is therefore a mixture of the cross power between mode $\alpha$ and others (denoted $\beta$ in the sum), with whitening (decorrelation and weighting) by the data covariance matrix. This expression reduces to this simple form

because the derivative

$$\frac{\partial C}{\partial p_\alpha} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \tag{63}$$

has only a single nonzero entry, with a corresponding single nonzero *row* in the quantity $C^{-1}\frac{\partial C}{\partial p_\alpha}$. This is for the case where the cosmological signal is not contaminated, and is confined to a single mode.

The final weighting matrix for the vector of estimates contains all of the weighting and correlation due to the primary beam and weighting and correlation due to the foreground structure.

*Example*

In the simplest case, there is no foreground or noise contribution, and the covariance matrix $C = C_P$. Then, Equation (62) reduces to

$$\langle \hat{p}_\alpha \rangle = \frac{\text{tr}\left( C_P^{-1} \frac{\partial C}{\partial p_\alpha} \right)}{\text{tr}\left( C_P^{-1} C_P^{-1} \frac{\partial C}{\partial p_\alpha} \right)} \tag{64}$$

$$= \frac{1/p_\alpha}{\sum_\beta |(C_P^{-1})_{\alpha\beta}|^2}, \tag{65}$$

yielding the true power estimate, weighted by the correlations between mode $\alpha$ and all other modes. In the limit where there are no covariances between modes,

$$\langle \hat{p}_\alpha \rangle = \frac{1/p_\alpha}{1/p_\alpha^2} = p_\alpha, \tag{66}$$

yielding an unbiased estimator.

The covariance matrix of the estimator provides a measure of its performance. In the case of a general data covariance, the covariance between powers $p_\alpha$ and $p_\beta$ is given by

$$\text{cov}(p_\alpha, p_\beta) \stackrel{\text{def}}{=} \langle \hat{p}_\alpha \hat{p}_\beta \rangle - \langle \hat{p}_\alpha \rangle \langle \hat{p}_\beta \rangle \tag{67}$$

$$= \frac{K}{\text{tr}(E_\alpha)\text{tr}(E_\beta)}, \tag{68}$$

where

$$K = 2\text{tr}(C^{-1}C_\alpha C^{-1}C_\beta) + \text{tr}(C^{-1}C_\alpha C^{-1}C_\beta C^{-1}C_{N,\text{FG}})$$
$$+ \text{tr}(C^{-1}C_\alpha C^{-1}C_{N,\text{FG}} C^{-1}C_\beta)$$
$$+ \text{tr}(C^{-1}C_\alpha C^{-1}C_{N,\text{FG}} C^{-1}C_\beta C^{-1}C_{N,\text{FG}}), \tag{69}$$

with $C_{N,\text{FG}} \stackrel{\text{def}}{=} C_{\text{FG}} + C_N$ and

$$E_\alpha = C^{-1}C^{-1}\frac{\partial C}{\partial p_\alpha} \tag{70}$$

$$C_\alpha = \frac{\partial C}{\partial p_\alpha}. \tag{71}$$

We have used the expression for the covariance of a zero-mean bilinear quadratic form,

$$\text{cov}(V_1^\dagger A_1 V_2, V_1^\dagger A_2 V_2) = 2\text{tr}(A_1 C A_2 C), \tag{72}$$

where $A_1$ and $A_2$ are general matrices and the data sets have means $V_1$ and $V_2$ and covariance $C$.

Given that we are using derived data as input into the estimator (coherently averaged visibilities, rather than individual measured visibilities), the squaring operates on a small number of visibilities, yielding a $\chi^2$-distribution for the data. (Although the quadratic form formally sums power over the whole *uv* range, in practice the localization of the beam makes most added powers have zero weight. This is by design to obtain a mostly diagonal covariance matrix and coupling only between neighboring *uv* points.) The covariance expression described above is therefore an input to an underlying skewed distribution.

## REFERENCES

Ali, Z. S., Parsons, A. R., Zheng, H., et al. 2015, ApJ, 809, 61
Beardsley, A. P., Hazelton, B. J., Morales, M. F., et al. 2013, MNRAS, 429, L5
Bennett, C. L., Larson, D., Weiland, J. L., et al. 2012, arXiv:1212.5225
Bonaldi, A., & Brown, M. L. 2015, MNRAS, 447, 1973
Bowman, J. D., Cairns, I., Kaplan, D. L., et al. 2013, PASA, 30, 31
Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2009, ApJ, 695, 183
Chapman, E., Abdalla, F. B., Harker, G., et al. 2012, MNRAS, 423, 2518
Chapman, E., Zaroubi, S., & Abdalla, F. 2014, arXiv:1408.4695
Choudhuri, S., Bharadwaj, S., Ghosh, A., & Ali, S. S. 2014, MNRAS, 445, 4351
Datta, A., Bowman, J. D., & Carilli, C. L. 2010, ApJ, 724, 526
Dillon, J. S., Liu, A., & Tegmark, M. 2013, PhRvD, 87, 043005
Dillon, J. S., Neben, A. R., Hewitt, J. N., et al. 2015, PhRvD, 91, 123011
Ellingson, S. W., Clarke, T. E., Cohen, A., et al. 2009, IEEEP, 97, 1421
Fialkov, A., Barkana, R., & Visbal, E. 2014, Natur, 506, 197
Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, PhR, 433, 181
Hamaker, J. P., Bregman, J. D., & Sault, R. J. 1996, A&AS, 117, 137
Hazelton, B. J., Morales, M. F., & Sullivan, I. S. 2013, ApJ, 770, 156
Heald, G. H., Pizzo, R. F., Orrú, E., et al. 2015, A&A, 582, A123
Hogg, D. W. 2000, arXiv:9905116v4
Intema, H. T., van Weeren, R. J., Röttgering, H. J. A., & Lal, D. V. 2011, A&A, 535, A38
Jacobs, D., Procopio, P., Mitchell, D., et al. 2015, ApJ, submitted
Jacobs, D. C., Pober, J. C., Parsons, A. R., et al. 2015, ApJ, 801, 51
Jelić, V., Zaroubi, S., Labropoulos, P., et al. 2008, MNRAS, 389, 1319
Jelić, V., Zaroubi, S., Labropoulos, P., et al. 2010, MNRAS, 409, 1647
Kay, S. M. 1993, Fundamentals of Statistical Signal Processing: Estimation Theory (Englewood Cliffs, NJ: Prentice-Hall)
Kazemi, S., Yatawatta, S., & Zaroubi, S. 2013, MNRAS, 430, 1457
Liu, A., Parsons, A. R., & Trott, C. M. 2015, PhRvD, 90, 023018
Liu, A., & Tegmark, M. 2011, PhRvD, 83, 103006
Lonsdale, C. J., Cappallo, R. J., Morales, M. F., et al. 2009, IEEEP, 97, 1497
McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L., & Furlanetto, S. R. 2006, ApJ, 653, 815
Mesinger, A., Ewall-Wice, A., & Hewitt, J. 2014, MNRAS, 439, 3262
Mesinger, A., Furlanetto, S., & Cen, R. 2011, MNRAS, 411, 955
Mitchell, D. A., Greenhill, L. J., Wayth, R. B., et al. 2008, ISTSP, 2, 707
Moore, D. F., Aguirre, J. E., Parsons, A. R., Jacobs, D. C., & Pober, J. C. 2013, ApJ, 769, 154
Morales, M. F., Hazelton, B., Sullivan, I., & Beardsley, A. 2012, ApJ, 752, 137
Morales, M. F., & Hewitt, J. 2004, ApJ, 615, 7
Offringa, A. R., van de Gronde, J. J., & Roerdink, J. B. T. M. 2012, A&A, 539, A95
Offringa, A. R., Wayth, R. B., Hurley-Walker, N., et al. 2015, PASA, 32, e008
Pacucci, F., Mesinger, A., Mineo, S., & Ferrara, A. 2014, MNRAS, 443, 678
Parsons, A., Pober, J., McQuinn, M., Jacobs, D., & Aguirre, J. 2012, ApJ, 753, 81
Parsons, A. R., Backer, D. C., Foster, G. S., et al. 2010, AJ, 139, 1468
Parsons, A. R., Liu, A., Aguirre, J. E., et al. 2014, ApJ, 788, 106
Patil, A. H., Zaroubi, S., Chapman, E., et al. 2014, MNRAS, 443, 1113
Paul, S., Sethi, S. K., Subrahmanyan, R., et al. 2014, ApJ, 793, 28
Pober, J. C., Liu, A., Dillon, J. S., et al. 2014, ApJ, 782, 66
Thompson, A. R., Moran, J. M., & Swenson, G. W. 1986, Interferometry and Synthesis in Radio Astronomy (New York: Wiley)
Thyagarajan, N., Jacobs, D. C., Bowman, J. D., et al. 2015a, ApJ, 804, 14
Thyagarajan, N., Jacobs, D. C., Bowman, J. D., et al. 2015b, ApJL, 807, L28

Thyagarajan, N., Udaya Shankar, N., Subrahmanyan, R., et al. 2013, ApJ, 776, 6

Tingay, S. J., Goeke, R., Bowman, J. D., et al. 2013, PASA, 30, 7

Trott, C. M., Wayth, R. B., & Tingay, S. J. 2012, ApJ, 757, 101

van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2

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

Vio, R., Diaz-Trigo, M., & Andreani, P. 2013, A&C, 1, 5

Watkinson, C. A., & Pritchard, J. R. 2014, MNRAS, 443, 3090

Wayth, R. B., Lenc, E., Bell, M. E., et al. 2015, PASA, 32, 25

Wyithe, J. S. B., & Morales, M. F. 2007, MNRAS, 379, 1647