The Astrophysical Journal, 833:102 (19pp), 2016 December 10 doi:10.3847/1538-4357/833/1/102
© 2016. The American Astronomical Society. All rights reserved.

# FIRST SEASON MWA EOR POWER SPECTRUM RESULTS AT REDSHIFT 7

A. P. Beardsley[1,2], B. J. Hazelton[2,3], I. S. Sullivan[2], P. Carroll[2], N. Barry[2], M. Rahimi[4,5], B. Pindor[4,5], C. M. Trott[5,6], J. Line[4,5], Daniel C. Jacobs[1], M. F. Morales[2], J. C. Pober[2,7], G. Bernardi[8,9,10], Judd D. Bowman[1], M. P. Busch[1], F. Briggs[5,11], R. J. Cappallo[12], B. E. Corey[12], A. de Oliveira-Costa[13], Joshua S. Dillon[13,14,15,16], D. Emrich[6], A. Ewall-Wice[13], L. Feng[13], B. M. Gaensler[5,17,18], R. Goeke[13], L. J. Greenhill[10], J. N. Hewitt[13], N. Hurley-Walker[6], M. Johnston-Hollitt[19], D. L. Kaplan[20], J. C. Kasper[10,21], H. S. Kim[4,5], E. Kratzenberg[12], E. Lenc[5,17], A. Loeb[10], C. J. Lonsdale[12], M. J. Lynch[6], B. McKinley[4,5], S. R. McWhirter[12], D. A. Mitchell[5,22], E. Morgan[13], A. R. Neben[13], Nithyanandan Thyagarajan[1], D. Oberoi[23], A. R. Offringa[5,24], S. M. Ord[5,6], S. Paul[25], T. Prabu[25], P. Procopio[4,5], J. Riding[4,5], A. E. E. Rogers[12], A. Roshi[26], N. Udaya Shankar[25], Shiv K. Sethi[25], K. S. Srivani[25], R. Subrahmanyan[5,25], M. Tegmark[13], S. J. Tingay[5,6], M. Waterson[6,11], R. B. Wayth[5,6], R. L. Webster[4,5], A. R. Whitney[12], A. Williams[6], C. L. Williams[13], C. Wu[27], and J. S. B. Wyithe[4,5]

[1] Arizona State University, School of Earth and Space Exploration, Tempe, AZ 85287, USA; Adam.Beardsley@asu.edu
[2] University of Washington, Department of Physics, Seattle, WA 98195, USA
[3] University of Washington, eScience Institute, Seattle, WA 98195, USA
[4] The University of Melbourne, School of Physics, Parkville, VIC 3010, Australia
[5] ARC Centre of Excellence for All-sky Astrophysics (CAASTRO), Australia
[6] International Centre for Radio Astronomy Research, Curtin University, Perth, WA 6845, Australia
[7] Brown University, Department of Physics, Providence, RI 02912, USA
[8] Square Kilometre Array South Africa (SKA SA), Park Road, Pinelands 7405, South Africa
[9] Department of Physics and Electronics, Rhodes University, Grahamstown 6140, South Africa
[10] Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA
[11] Australian National University, Research School of Astronomy and Astrophysics, Canberra, ACT 2611, Australia
[12] MIT Haystack Observatory, Westford, MA 01886, USA
[13] MIT Kavli Institute for Astrophysics and Space Research, Cambridge, MA 02139, USA
[14] Department of Astronomy, UC Berkeley, Berkeley, CA 94720, USA
[15] Radio Astronomy Laboratory, UC Berkeley, Berkeley, CA 94720, USA
[16] Berkeley Center for Cosmological Physics, Berkeley, CA 94720, USA
[17] The University of Sydney, Sydney Institute for Astronomy, School of Physics, NSW 2006, Australia
[18] Dunlap Institute for Astronomy and Astrophysics, University of Toronto, ON M5S 3H4, Canada
[19] Victoria University of Wellington, School of Chemical & Physical Sciences, Wellington 6140, New Zealand
[20] University of Wisconsin–Milwaukee, Department of Physics, Milwaukee, WI 53201, USA
[21] University of Michigan, Department of Atmospheric, Oceanic and Space Sciences, Ann Arbor, MI 48109, USA
[22] CSIRO Astronomy and Space Science (CASS), P.O. Box 76, Epping, NSW 1710, Australia
[23] National Centre for Radio Astrophysics, Tata Institute for Fundamental Research, Pune 411007, India
[24] Netherlands Institute for Radio Astronomy (ASTRON), P.O. Box 2, 7990 AA Dwingeloo, The Netherlands
[25] Raman Research Institute, Bangalore 560080, India
[26] National Radio Astronomy Observatory, Charlottesville and Greenbank, USA
[27] International Centre for Radio Astronomy Research, University of Western Australia, Crawley, WA 6009, Australia
*Received 2016 July 15; revised 2016 August 12; accepted 2016 August 20; published 2016 December 9*

## ABSTRACT

The Murchison Widefield Array (MWA) has collected hundreds of hours of Epoch of Reionization (EoR) data and now faces the challenge of overcoming foreground and systematic contamination to reduce the data to a cosmological measurement. We introduce several novel analysis techniques, such as cable reflection calibration, hyper-resolution gridding kernels, diffuse foreground model subtraction, and quality control methods. Each change to the analysis pipeline is tested against a two-dimensional power spectrum figure of merit to demonstrate improvement. We incorporate the new techniques into a deep integration of 32 hours of MWA data. This data set is used to place a systematic-limited upper limit on the cosmological power spectrum of $\Delta^2 \leqslant 2.7 \times 10^4$ mK$^2$ at $k = 0.27\,h\,\mathrm{Mpc}^{-1}$ and $z = 7.1$, consistent with other published limits, and a modest improvement (factor of 1.4) over previous MWA results. From this deep analysis, we have identified a list of improvements to be made to our EoR data analysis strategies. These improvements will be implemented in the future and detailed in upcoming publications.

*Key words:* cosmology: observations – cosmology: reionization

## 1. INTRODUCTION

Detection and characterization of the cosmic dark ages and the Epoch of Reionization (EoR) have the potential to inform our picture of the cosmos in ways analogous to the Cosmic Microwave Background (CMB) over the past several decades. The EoR, in particular, is rich with both cosmological and astrophysical dynamics, as early-universe linear evolution gives way to nonlinear structure growth, and stars and galaxies reionize the intergalactic medium (IGM).

Several probes are being used to investigate the EoR. Studies of the polarization of the CMB have placed integrated constraints on the timing of reionization (e.g., Bennett et al. 2013; Planck Collaboration et al. 2015a, 2015b). Meanwhile, observations of highly redshifted quasars have placed upper bounds on redshifts by which reionization is completed. For

example, Fan et al. (2006) showed that reionization must be complete by $z \approx 6$, a result that was confirmed independent of the model by McGreer et al. (2015). Using a $z_{\rm end} > 6$ prior, the most recent analysis of *Planck* data found a reionization redshift $z_{\rm re} = 8.8 \pm 0.9$ for a redshift-symmetric model ($z_{\rm re} = 8.5 \pm 0.9$ for redshift-asymmetric), and a duration of $\Delta z < 2.8$ (Planck Collaboration et al. 2016). Deep optical and infrared galaxy surveys are also beginning to reach the redshifts necessary to further constrain reionization (e.g., Bouwens et al. 2014), and the *James Webb Space Telescope* will improve on their sensitivity (Gardner et al. 2006).

The 21 cm hyperfine transition from neutral hydrogen residing in the IGM during reionization offers another promising method to study the EoR. Not only is the 21 cm signal a *direct* probe of the IGM, but, due to the narrow width of the transition and the relationship between the observed frequency and line-of-sight distance, it can be used to map the full three-dimensional (3D) space of the epoch. (For reviews, see Furlanetto et al. 2006; Morales & Wyithe 2010; Pritchard & Loeb 2012; Loeb & Furlanetto 2013, and Zaroubi 2013.) The first generation of instruments with primary science goals to detect the highly redshift 21 cm signal have been built, including the Giant Metrewave Radio Telescope (GMRT, Paciga et al. 2013), LOw Frequency Array (LOFAR[28] Van Haarlem et al. 2013; Yatawatta et al. 2013), PAPER (Donald C. Backer Precision Array for Probing the Epoch of Reionization[29], Parsons et al. 2010), and the Murchison Widefield Array (MWA[30], Bowman et al. 2013; Tingay et al. 2013). Due to a relatively low signal-to-noise ratio, these instruments aim for a statistical measurement of the EoR in the form of a cosmological power spectrum. Complementary to power-spectrum experiments are efforts to detect the sky-average global 21 cm signal from the EoR (e.g., Bowman & Rogers 2010; Patra et al. 2013; Voytek et al. 2014; Sokolowski et al. 2015). Meanwhile, the second generation of 21 cm EoR interferometers is on its way with the Hydrogen Epoch of Reionization Array[31] (DeBoer et al. 2016), and the low frequency Square Kilometer Array (SKA1-Low Mellema et al. 2013), which will further refine the power spectrum measurement in early build-out stages, but will ultimately be capable of imaging the ionized bubbles of reionization in its later stages (Malloy & Lidz 2013; Beardsley et al. 2015).

Using the first generation of instruments, several upper limits have been placed on the 21 cm EoR signal (Paciga et al. 2013; Dillon et al. 2014; Parsons et al. 2014; Ali et al. 2015; Jacobs et al. 2015). Furthermore, Pober et al. (2015) and Greig et al. (2016) were able to place constraints on physical reionization models. However, much work is to be done to understand the data produced by these arrays. Foreground subtraction and isolation has emerged as a priority in the field. Central to most analysis strategies is the concept of an "EoR window"—the region of Fourier space where spectrally smooth foregrounds have been isolated from the isotropic (spherically symmetric in Fourier space) cosmological signal (Morales et al. 2006; Bowman et al. 2009). More recent studies have shown the existence of a foreground "wedge," the result of instrumental mode mixing, throwing power from spectrally smooth foregrounds to higher Fourier modes (Datta et al. 2010; Morales et al. 2012; Trott et al. 2012; Vedantham et al. 2012; Hazelton et al. 2013; Pober et al. 2013; Thyagarajan et al. 2013, 2015a, 2015b; Liu et al. 2014a, 2014b), while others are investigating the spectral structure of point sources themselves (e.g., Offringa et al. 2016). The EoR window is still expected to be preserved above the wedge, and several analysis pipelines are under active development to exploit this foreground isolation (e.g., Dillon et al. 2013, 2015b; Trott 2014; B. J. Hazelton et al. 2016, in preparation; D. A. Mitchell et al. 2016, in preparation; Jacobs et al. 2016; Trott et al. 2016).

This paper serves two purposes: to demonstrate several new analysis techniques and their impact on power spectrum estimation, and to present the first deep integration power spectrum from the MWA. Using a three-hour test set of data, we introduce several novel techniques, including calibration of cable reflection contamination, high-resolution gridding kernels, subtraction of a diffuse foreground model, and development of quality control methods that will be crucial for deeper integrations. We apply these new techniques to a deep integration of data from the first semester of MWA observations (2013 August–November), with a total of 32 hours on a single EoR field and redshift range $6.2 < z < 7.5$.

Our best result is a limit on the cosmological power spectrum of $\Delta^2 \leqslant 2.7 \times 10^4$ mK$^2$ at $k = 0.27\,h$ Mpc$^{-1}$ and $z = 7.1$. This is lower than the GMRT 40-hour limit ($\Delta^2 \leqslant 6.15 \times 10^4$ mK$^2$ at $z = 8.6$; Paciga et al. 2013), but significantly higher than the latest PAPER result, which integrated over 700 hours ($\Delta^2 \leqslant 502$ mK$^2$ at $z = 8.4$; Ali et al. 2015). Several EoR experiments are reaching sensitivity levels where very low systematics become dominant. The PAPER team saw evidence that antenna cross-talk and foreground leakage was limiting their result at several scales. The LOFAR team has seen excess noise and diffuse foreground suppression due to calibration (Patil et al. 2016). It is essential for EoR experiments to understand and overcome these systematics in order to reach a cosmological detection.

While not expecting to detect the EoR with less than 100 hours of observations, our intermediate integration will serve to identify and diagnose systematics, allowing for improved analysis in future work. We list several directions to improve our modeling of foregrounds and the instrumental response, which, in turn, will improve our calibration and foreground subtraction. Our strategy to perform periodic deep integrations allows us to continue uncovering systematics, refine algorithms and data analysis, and understand the subtleties of the instrument to execute this challenging experiment.

The remainder of this paper is organized as follows. In Section 2, we briefly describe the MWA instrument and the observations used in this work; in Section 3, we describe our analysis pipeline; in Section 4, we describe several novel techniques in our analysis; in Section 5, we discuss our efforts to apply the analysis pipeline to a deep data integration and compare it with an alternative analysis pipeline for robust estimation; and in Section 6, we discuss future work, including a summary of planned improvements to the analysis. Throughout this paper, we use a ΛCDM cosmology with $\Omega_m = 0.73$, $\Omega_\Lambda = 0.27$, and $h = 0.7$, consistent with *WMAP* seven-year results (Komatsu et al. 2011). All distances and wave numbers are in comoving coordinates.

[28] http://www.lofar.org
[29] http://eor.berkeley.edu
[30] http://www.mwatelescope.org
[31] http://reionization.org

## 2. THE MURCHISON WIDEFIELD ARRAY AND OBSERVATIONS

The MWA is one of several first generation radio interferometers with a primary science goal of detecting the 21 cm EoR power spectrum. The remote Australian outback offers relative isolation from human-made radio-frequency interference (RFI), such as FM radio or TV stations. A recent study of the RFI environment at the Murchison Radio Observatory can be found in Offringa et al. (2015).

While the 21 cm EoR signal is priority, the MWA is a general observatory serving several science programs beyond cosmology, including Galactic and extragalactic surveys, time domain astrophysics, solar monitoring, and ionospheric science. The array was thus designed with these programs in consideration and the layout was optimized using a pseudo-random antenna placement algorithm (Beardsley et al. 2012) to obtain a well-behaved point-spread function (PSF) while retaining a dense core for EoR sensitivity (Bowman et al. 2006; Beardsley et al. 2013).[32] A full description of the science capabilities of the array is found in Bowman et al. (2013). The high imaging capability of the MWA enables us to calibrate and subtract foregrounds based on sky models, leveraging the full capability of the observatory for the EoR experiment. This strategy is in contrast to more targeted EoR experiments, like PAPER, which utilize a highly redundant layout to enhance sensitivity. Redundant arrays can exploit symmetries of the instrument for quick calibration and analysis, but have poor point-spread functions, making foregrounds more difficult to characterize.

The technical design of the MWA is reviewed in Tingay et al. (2013), and we highlight a few key characteristics that will become important in our analysis of the data. Each antenna of the MWA comprises 16 dual-polarization dipoles placed on a regular grid, lying on a ground screen. The radio-frequency signals from these dipoles feed into an analog beamformer, which uses physical delays to "point" the antenna. The MWA contains 128 of these antennas, with a tightly packed core of radius 50 m, and extending out to a radius of 1.5 km for higher resolution calibration and imaging.

The beamformed signals (one for each polarization) are then transmitted to receivers in the field, which digitize the signal and perform a first stage coarse frequency channelization of the data (1.28 MHz coarse bands; Prabu et al. 2015). This step applies a filter shape and aliases channels on the edge of the coarse bands, which will later be flagged in our analysis. The observer now selects 24 coarse bands (30.72 MHz total bandwidth) to pass onto the correlator via fiber optic link. The correlator further channelizes the data to 10 kHz resolution, cross multiplies signals between antennas to form visibilities, and averages in time and frequency to a resolution specified by the observer. For the data in this work, the correlator output resolutions were 0.5 s in time and 40 kHz in frequency. The data are then written to the disk on a cadence of 112 s, constituting a single observation, or snapshot.

The EoR observing campaign has adopted a "drift and shift" tracking strategy—we point the telescope toward a sky field of interest, allow the field to drift overhead for about 30 minutes until it begins to leave our field of view, then repoint the instrument at the field. The telescope is thus only pointed at discrete positions, or "pointings." This tracking is repeated until the field is too low in the sky to track. We have identified three EoR fields relatively devoid of Galactic emission and bright extragalactic sources for observing: "EoR0" (Right Ascension (R.A.) $= 0^{\rm h}.00$, Declination (decl.) $= -27°$), "EoR1" (R.A. $= 4^{\rm h}.00$, decl. $= -27°$), and "EoR2" (R.A. $= 11^{\rm h}.33$, decl. $= -10°$). In addition, we observe these fields in high and low bands centered at 182 MHz ($z \approx 6.8$) and 154 MHz ($z \approx 8.2$), respectively.

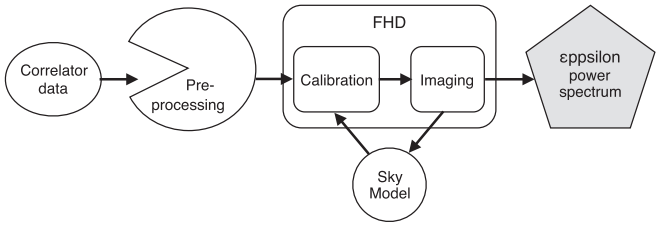


**Figure 1.** Schematic of the analysis pipeline used in this work. The raw correlator data is preprocessed by performing RFI and other flagging. The data are then calibrated and imaged using an iterative foreground modeling approach with the FHD package. Finally, a power spectrum is formed using the ϵppsilon package.

The data for this work includes two sets. First we use a set of 94 two-minute high-band observations of the EoR0 field from 2013 August 23, which were taken early in telescope operations and found to be particularly well behaved. The MWA EoR collaboration has designated this set as a "golden data set," which is used to build analysis tools and compare early results (e.g., Dillon et al. 2015a; Jacobs et al. 2016; Pober et al. 2016; Trott et al. 2016). All techniques demonstrated here will use this golden data set. In Section 5, we will use a larger data set to move toward a deep integration. This data set consists of all EoR0 high-band observations beginning 2013 August 23 (including the golden data set), and concluding when the field was no longer accessible for the season: when the Sun was above the horizon for most or all of the field's transit. The final observation of this set was on 2013 November 29. The individual snapshots are 112 s of data at 0.5 s, 40 kHz resolution. In total, we analyze 2780 snapshots, or about 86.5 hours of data.

## 3. ANALYSIS PIPELINE

In order to ensure consistency in analysis, the international MWA EoR collaboration has defined two independent reference analysis pipelines; the details of this strategy are outlined in Jacobs et al. (2016). This study is based on the FHD-to-ϵppsilon pipe, which is in turn based on the Fast Holographic Deconvolution (FHD[33], Sullivan et al. 2012) and Error Propagated Power Spectrum with Interleaved Observed Noise (ϵppsilon[34], B. J. Hazelton et al. 2016, in preparation) packages. The general flow of the pipeline is shown in Figure 1, and we give an overview below. As a point of comparison, and to demonstrate robustness of the results, we compute the output power spectra with an independent calibration and power spectrum estimation pipeline, and demonstrate consistency.

[32] A drone flyover view of the array layout is available on YouTube at https://youtu.be/yDWdTUzTUMo.

[33] https://github.com/EoRImaging/FHD

[34] https://github.com/EoRImaging/eppsilon

### 3.1. Preprocessing

The preprocessing step converts the data from the non-standard memory dump format from the correlator's GPUs to a standard UVFITS format (Greisen 2012), phases the data to the center of the pointing, performs flagging, averages in time and frequency, and writes to the disk. For this step we use the COTTER package (Offringa et al. 2015), which in turn calls the AOFLAGGER[35] to perform RFI flagging (Offringa et al. 2010).

Besides RFI, we manually flag data to remove three other effects. As mentioned in Section 2, the edges of the coarse bands contain aliasing from the filter applied in the digital receivers. We flag two 40 kHz channels on either side of the coarse-band edges (total of four channels per coarse-band edge). In principle, the aliasing could be calibrated out and these channels could be recovered, but this is left for future work. Second, early data from the MWA also contained some instances where antennas were not pointed properly at the beginning of observations due to a beamformer error. This problem has since been resolved, but exists in our data. We therefore flag the first two seconds of each observation to avoid any potentially mis-pointed data. Finally, we flag the central 40 kHz channel of each coarse band. This channel corresponds to the coarse-band DC mode, which has been observed to contain anomalous power, likely due to very low-level rounding errors in the polyphase filter bank of the digital receivers.

In the final step of preprocessing, COTTER averages the data to 2 s, 80 kHz resolution and writes to the disk. At this stage, each snapshot is contained in a single UVFITS file, which serves as the base unit for calibration and imaging.

### 3.2. Calibration and Imaging

The next step in our pipeline is to calibrate the visibilities. As shown in Figure 1, a sky model is derived from our imaged data, which feeds back into the calibration. This is naturally an iterative process as the sky model is refined with improved calibration solutions. We will describe the sky model in detail in Section 4, but here we present the production mode of analysis after the model has been determined.

Calibration is accomplished within the FHD package in two steps. First we match the raw data to model visibilities formed using a realistic simulation of the array response to our foreground model. This step results in independent complex gains for each antenna, 80 kHz frequency channel, and polarization. In the second step we impose restrictions on these gains motivated by our understanding of the spectral response of the instrument. By reducing the number of free parameters in this step we increase our signal to noise on calibration solutions and avoid absorbing unmodeled confusion source flux density into our gain solutions which has been shown to contaminate the EoR window (Barry et al. 2016). Recently, Patil et al. (2016) demonstrated excess noise and a suppression of diffuse foregrounds in LOFAR images after calibration and subtraction of bright discrete foreground sources. They suggested a multi-frequency calibration solution as a potential solution to these systematics, similar to the scheme implemented here. Indeed we have not seen evidence of these systematics through simulations and direct propagation of our noise (B. J. Hazelton et al. 2016, in preparation; Barry et al. 2016).

In the first calibration step, we allow our complex gains to account for any direction-independent response on a per antenna, per 80 kHz frequency channel, per polarization basis. To define our gains in this way, we have made two simplifications. First, we push all direction-dependence of the antenna response into the model of the primary beam. FHD is capable of using separate models for each antenna, though this has not been implemented in this work. The second simplification is to ignore any cross terms between our different axes; for example, a mutual coupling between two antennas. In principle, these terms would introduce baseline dependent gains, rather than antenna dependent. We have not yet seen evidence of these terms in MWA data, and so we currently neglect them in our solutions.

Under these assumptions we can express the measured, uncalibrated visibility for an antenna pair $(i, j)$ as

$$V'_{ij}(\nu) \approx g_i(\nu) g_j^*(\nu) V_{ij}(\nu), \tag{1}$$

where $g_i(\nu)$ and $g_j(\nu)$ are the complex gains for antennas $i$ and $j$, respectively, and $V_{ij}(\nu)$ is the true visibility, which we aim to recover through calibration. Because we are treating the polarizations independently, we allow the antenna subscript to run over both east–west (E–W) and north–south (N–S) polarizations.

FHD utilizes the StEFCal algorithm described in Salvini & Wijnholds (2014) to find the minimum $\chi^2$ estimate with respect to the complex gains. The result of this operation is estimated gains for every antenna, frequency channel, and polarization. However, with certain known properties of the antenna response we can reduce the number of free parameters in our solution. This takes us to the second step of calibration.

Our initial gain model included an amplitude bandpass common to all antennas, $B(\nu)$, and an antenna-dependent low-order polynomial in frequency. Mathematically, we can express our restricted gain as

$$\hat{g}_i(\nu) = B(\nu) P_i(\nu), \tag{2}$$

where the polynomial term can be further decomposed as

$$P_i(\nu) = (A_{i,0} + \nu A_{i,1} + \nu^2 A_{i,2}) e^{i(\phi_{i,0} + \nu \phi_{i,1})}. \tag{3}$$

The coefficients $B(\nu)$, $A_{i,n}$, and $\phi_{i,n}$ are real quantities. This model allows us to capture arbitrary spectral responses due to common antenna factors (such as the polyphase filter shapes or phased array response), as well as slowly varying antenna-dependent deviations. We also found it necessary to include terms dependent on the length of the cable connecting the beamformers to the receivers, which we will discuss in Section 4.1.

Once all calibration terms are found, we complete the calibration by dividing the raw data by the gain estimates to produce calibrated visibilities. The entire calibration process is done for every snapshot of EoR data, providing both a two-minute resolution time dependence of calibration solutions, as well as model visibilities that are used for foreground subtraction and diagnostic purposes.

After calibration, we form snapshot image cubes (frequency mapping to line-of-sight direction). The visibilities are gridded using the primary beam as the gridding kernel, placing the data in the "holographic frame," which has been shown to be the

[35] http://aoflagger.sourceforge.net/

optimal weighting scheme for combining images in the sense that it preserves all information for parameter estimation (Bhatnagar et al. 2008; Morales & Matejek 2009). We use a simulation-based primary beam model for MWA antennas developed by Sutinjo et al. (2015), which incorporates an average embedded element pattern for the dipoles, as well as mutual coupling between dipoles within an antenna. At this stage, we also average in frequency, by a factor of two, by gridding pairs of frequency channels to the same $(u, v)$ plane, resulting in a frequency resolution of 160 kHz. This is done to reduce the data volume and has little impact on our results, because our sensitivity to the EoR is extremely low at the corresponding line-of-sight $k$ modes (Beardsley et al. 2013).

The data are gridded into separate cubes for even/odd interleaved time samples. The sum of the even and odd cubes contains power from the sky as well as noise, while the difference between the two contains only noise power. Subtracting the power spectrum of the difference from that of the sum yields an unbiased estimate of the sky power (Jacobs et al. 2016).

The gridded $(u, v)$ data are then Fourier transformed to create sky-frame images. To avoid aliasing, we image out 90° from phase center (gridding resolution of a half wavelength), and crop the image. We found that cropping the image based on beam value resulted in hard edges when integrating images from different snapshots having beams pointed in slightly different directions. To avoid the hard edges, we predetermine a set of HEALPix pixels to interpolate to, and use the same set for all snapshots on a given field. The final cropped field of view for this work is a 21° square centered at R.A. $= 0^{\rm h}$, decl. $= -27°$.

In principle the foregrounds could be subtracted from the data immediately after calibration, before gridding and imaging. However, for diagnostic reasons, we have found it beneficial to carry the model through the entire pipeline, and only subtract just before squaring to power spectrum units, allowing us to form power spectra of the dirty, model, and residual data.

FHD provides all inputs needed for the ϵppsilon package, which include weight cubes (for proper accumulation of data), variance cubes (for error propagation), and even/odd interleaved data cubes (for an unbiased estimator and a direct measurement of the noise). We also retain both E–W and N–S polarizations. At this time, we can remap our coordinates to cosmological units according to the relationships given in Morales & Hewitt (2004), where angular and frequency units $(\theta_x, \theta_y, f)$ map to cosmological distance units $(r_x, r_y, r_z)$. Often, $r_z$ is referred to as the line-of-sight dimension, and denoted as $r_\parallel$. Similarly, $(r_x, r_y)$ is the plane perpendicular to the line of sight and is denoted as $\boldsymbol{r}_\perp$.

### 3.3. Power Spectra

In the final steps of our pipeline, we integrate the snapshot image cubes, and calculate a power spectrum estimate. Though the integration step is formally an imaging component, we conceptually group it with the power spectrum step. This is because all steps of the pipeline up to this point have been performed on a per-snapshot basis (no communication between snapshots beyond foreground modeling), and the integration step can be used to select subsets of data for diagnosing power-spectrum artifacts. The images produced by FHD are in the holographic frame, which is already properly weighted for combining images, so the integration is simply adding the cubes together and propagating the weights.

The ϵppsilon package performs a discrete Fourier transform (DFT) on each $r_\parallel$ slice of the integrated HEALPix cube, forming a cube in $(\boldsymbol{k}_\perp, r_\parallel)$, where $\boldsymbol{k}_\perp$ represents the cosmological wave number in the plane perpendicular to our line of sight, and $r_\parallel$ is the distance to the observed redshift slice along the line of sight. The Fourier transform along the $r_\parallel$ dimension is treated separately due to incomplete $(u, v)$ sampling and flagged frequency channel; this leads to structure in the frequency sampling along any given $k_\perp$ pixel. We thus adopt the Lomb–Scargle least-squares method to determine the orthogonal eigenfunctions given our sampling function and estimate the total power in each $k_\parallel$ mode (Scargle 1982). In addition, because the spectrally smooth foregrounds contain vastly more power than the expected EoR signal, we apply a Blackman–Harris window function prior to the $r_\parallel$ to $k_\parallel$ transform. This has the effect of trading lower effective bandwidth for higher dynamic range (e.g., Thyagarajan et al. 2013, 2016).

The pipeline as described up to this point is applied to both the calibrated data and the foreground model. At this point, we subtract the model from the data to form residual data as well. All three sets of data (dirty, model, and residual) are carried through the remaining steps to form corresponding power spectra.

After squaring and dividing by our observation window function (Bowman et al. 2006), we arrive in 3D power spectrum space. We use the even/odd interleaved cubes to form a signal (odd plus even) and noise (odd minus even) power spectrum, which we subtract from one another to form an unbiased estimate of our signal power. This is mathematically equivalent to cross multiplying the even and odd cubes. The 3D power spectrum cube can next be averaged in annuli of constant $k_\perp$ and $k_\parallel$ (i.e., orthogonal to the $k_\parallel$ axis) to form 2D power spectra, or spherical shells to form one-dimensional (1D) power spectra where we will ultimately constrain the EoR.

An example 2D residual power spectrum formed from the three-hour golden data set is shown in Figure 2. The bulk of the residual power is in the so-called foreground "wedge," indicated with the diagonal black lines where the solid line corresponds to sky emission at the horizon, and the dashed line corresponds to emission at the edge of the MWA field of view. Above the wedge, we see horizontal lines of contamination due to the periodic frequency sampling function. These lines are often referred to as the coarse-band harmonics. We see vertical streaks at high $k_\parallel$ and high $k_\perp$. These are due to sparse $(u, v)$ sampling beyond the dense core of the MWA (starting $\sim 70\lambda$), which results in non-uniform spectral sampling after gridding. This, in turn, causes the foreground power to mix to high spectral modes (Bowman et al. 2009). The frequency-dependent sampling creates a PSF in the $k_\parallel$ for each $(u, v)$ cell. In principle, covariance weighting along the frequency dimension may be able to mitigate this leakage (e.g., Liu & Tegmark 2011), and an ϵppsilon implementation is under development. We provide axes more closely related to the measurements to assist in connecting these instrumental effects. The right axis shows delay (Fourier dual to frequency), and the top axis shows the baseline length measured in wavelengths. Between the coarse-band lines and to the left of the vertical streaks we see regions that contain both positive and negative

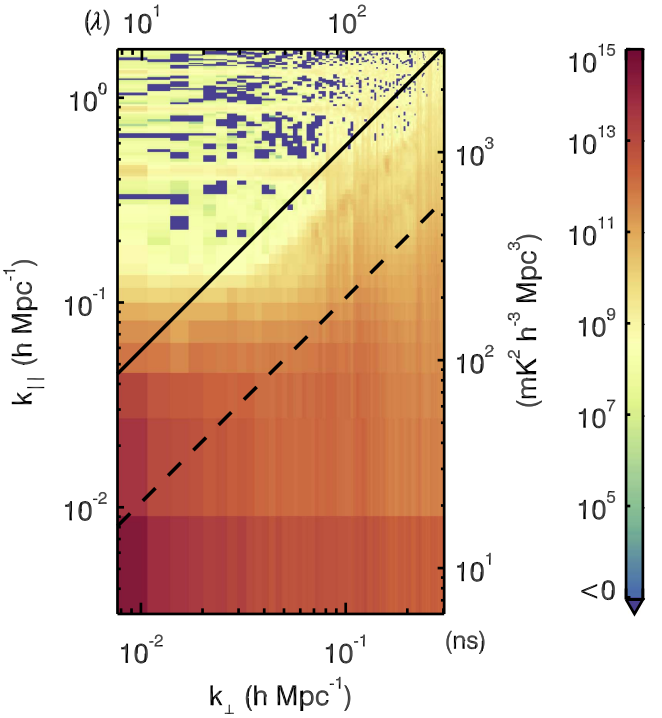


**Figure 2.** Example two-dimensional, E–W polarization residual power spectrum formed using the three-hour golden data set from 2013 August 23. The foreground wedge dominates the region below the solid black horizon line, while regions above are mostly noise-like, with the exception of horizontal coarse-band harmonic lines. The vertical streaks at high $k_{||}$ and high $k_{\perp}$ are due to sparse $(u, v)$ sampling beyond the dense core of the MWA. The right axis shows the delay (Fourier dual to frequency) corresponding to the $k_{||}$ axis, and the top axis shows the baseline length (in wavelengths) corresponding to the $k_{\perp}$ axis.

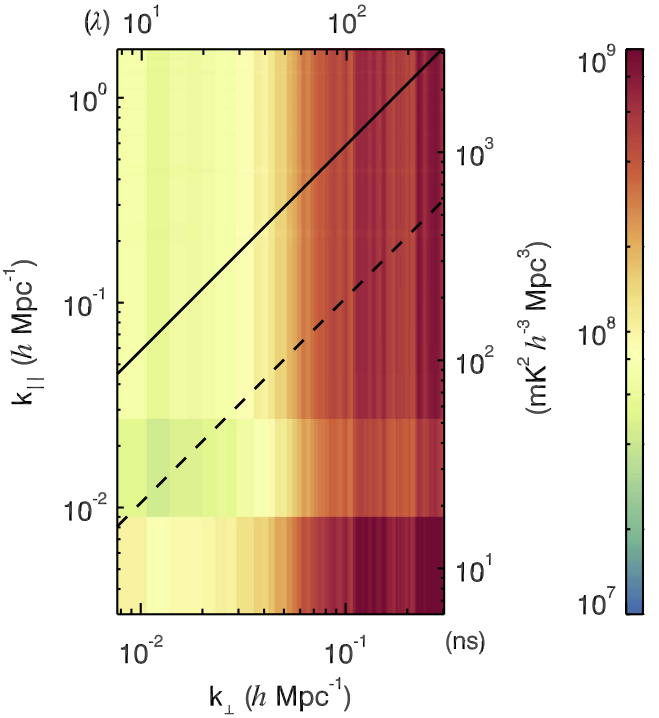


**Figure 3.** Uncertainty level corresponding to Figure 2. The noise is relatively low at short baseline length (low $k_{\perp}$) due to the dense MWA core. The array is more sparse at longer baselines, and the noise increases. The second $k_{||}$ bin has slightly lower noise than the rest of the $k_{||}$ modes due to the interaction between the Lomb–Scargle periodogram with the Blackman–Harris window function applied to the data.

(indicated by blue on the color bar) pixels. These regions are where we have successfully isolated the foregrounds and retained a noise-like EoR window in the three-hour integration.

Figure 3 shows the uncertainty level corresponding to the 2D power spectrum in Figure 2. These levels are found by propagating noise measured from calibrated visibilities through the entire pipeline. The dense core of the MWA results in relatively low noise at low $k_{\perp}$. As the $(u, v)$ density of the array decreases at longer baselines, the noise increases. The noise is mostly constant across $k_{||}$ because the thermal noise amplitude is relatively constant over frequency and the Fourier transform distributes the noise power across all modes (Morales & Hewitt 2004). The exception is the second $k_{||}$ bin, which has slightly lower noise than the rest of the $k_{||}$ modes. As explained by B.J. Hazelton et al. (2016, in preparation), this is a result of the Lomb–Scargle periodogram interacting with the Blackman–Harris window function we apply to the data. Most of the bins have the noise equally distributed between cosine and sine terms, with the exception of the $k_{||} = 0$ bin, where the cosine term carries all the information. Due to the correlations introduced by the Blackman–Harris window function, the first non-zero $k_{||}$ mode is highly covariant with the bottom bin—therefore, the noise on this mode's cosine term is lower than other higher order cosines.

## 4. ANALYSIS METHODS FOR TESTING DATA QUALITY

The two-dimensional (2D) power spectrum is a useful figure of merit (FoM) as we improve and refine our analysis pipeline. Foregrounds and systematics often manifest with characteristic shapes in this space, enabling us to diagnose problems and quantify improvements (Morales et al. 2012). We use the MWA three-hour "golden" data set from 2013 August 23 to repeatedly form power spectra to test and refine our analysis. While an exhaustive catalog of these improvements is outside the scope of this paper, we demonstrate the utility of the 2D power spectrum as an FoM with a few key techniques we have employed in our analysis.

### 4.1. Cable Dependent Calibration

Early in our analysis it became apparent that the bandpass of each MWA antenna requires more free parameters than those described in Equation (2). In particular, differing lengths of cable connecting the beamformers to the receivers lead to different bandpass shapes due to signal attenuation. We therefore allow the amplitude bandpass factor to depend on cable length, $B_{\alpha}(\nu)$, where $\alpha$ denotes the cable type.[36]

The top panel of Figure 4 shows a 2D power spectrum using the gain model described so far. Three of the horizontal lines in the EoR window can be attributed to the coarse-band gaps, as they reside at the harmonics expected for 1.28 MHz periodic sampling. But the sampling cannot account for the fourth line, highlighted by the arrow. The corresponding delay time of the line, $\tau \approx 1.23$ $\mu$s, corresponds almost exactly to twice the signal travel time through the 150 m cables (with velocity 0.81c). We therefore introduced a reflection term into our gain

[36] For logistical purposes, the beamformer to receiver cables were installed in six set lengths. Cables of length 90, 150, and 230 m are RG-6, while 320, 400, and 524 m cables are LMR400-75.

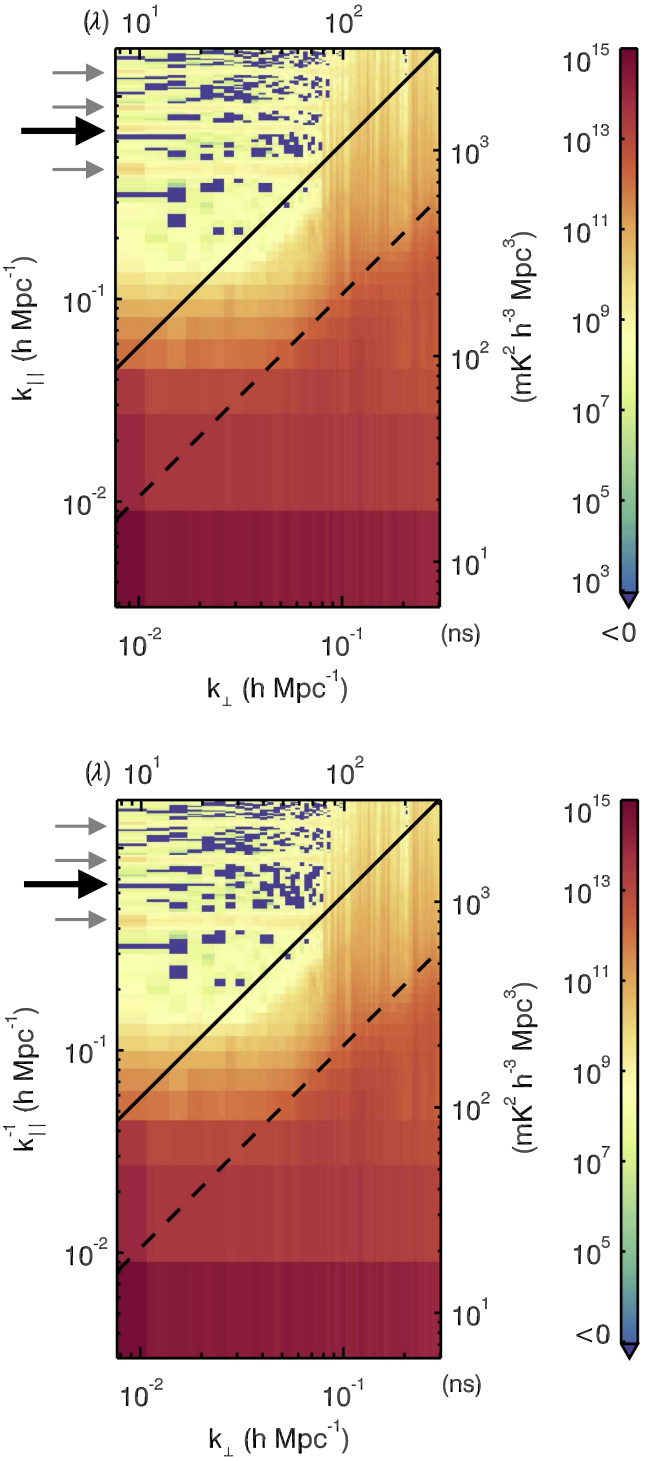


**Figure 4.** Top: single polarization dirty power spectrum formed from the golden data set, before implementing cable reflection fitting into the calibration loop. The small gray arrows point to the bands resulting from the periodic coarse-band sampling. The black arrow points to a horizontal band at $k_{||} \approx 0.7\,h\,\mathrm{Mpc}^{-1}$, which cannot be accounted for by the coarse bands. Bottom: after we implement the cable reflection fitting in our calibration solutions, we see the reflection line disappears.

model, which allows our restricted gains to account for the interference of the incident signal with a round-trip reflected signal in the cable. A similar approach was taken at lower frequencies in Ewall-Wice et al. (2016). The full gain model is expressed as

$$\hat{g}_i(\nu) = B_\alpha(\nu)(P_i(\nu) + R_i(\nu)), \tag{4}$$

where the reflection term can be further decomposed as

$$R_i(\nu) = R_{i,0}e^{-2\pi i\tau_i\nu}. \tag{5}$$

The reflection coefficient, $R_{i,0}$, is allowed to be complex, while the reflection delay, $\tau_i$, is real. After fitting for all other parameters, we fit for the reflection mode. We found that if we did not first fit and remove the polynomial terms, our reflection fits would be dominated by the large power in the smooth structure, resulting in solutions that did not match physical reality. The reflection fitting is done for antennas with suspected reflections, the most offensive of which is seen in the antennas with cables of 150 m. Because the cables were not cut at exact lengths (variations on order tens of centimeters), we solve for both the reflection coefficient and delay by performing a direct Fourier transform to a highly over-resolved delay grid and selecting the mode where the reflection amplitude is largest. This fitting produces a single reflection coefficient for each antenna with nominal cable length of 150 m.

The resulting 2D power spectrum after fitting for cable reflections is shown in the bottom panel of Figure 4, where the reflection line is suppressed below the noise level. This example demonstrates the power of the 2D power spectrum as an FoM. The power in the reflection line was about five orders of magnitude below the foregrounds, making it very difficult to detect in image-based metrics. However, the power spectrum is specifically designed to be sensitive to low-level spectral structure, and using the two-dimensional spectrum allows us to identify the "shape" of contamination, enabling a precise diagnosis.

### 4.2. Gridding Kernel Resolution

A similar, low-level effect that had the potential to contaminate the EoR window is shown in Figure 5. The spectra shown are the *model* spectra—calculated by propagating the sky model visibilities through the entire pipeline. Despite our foreground model not containing spectral structure, the EoR window in the top panel seems to have a floor at a level $\sim 10^7$ mK$^2$ $h^{-3}$ Mpc$^3$, comparable to predicted EoR signals. In addition, there is a faint line in the upper left of the plot that has the same slope as the wedge, but seems to originate far beyond the horizon.

The source of this floor and super-horizon line was traced to the resolution at which we formed the gridding kernel while gridding visibilities. In the interest of computational speed and efficiency, the kernel is pre-calculated at a high resolution, then a nearest-neighbor lookup table is used to approximate the beam values at discrete pixels. This is a common practice in most analysis software packages. The kernel resolution used in the top panel of Figure 5 was 0.04 wavelengths—much smaller than the half-wavelength grid corresponding to horizon-to-horizon imaging. However, as a baseline migrates in $(u, v)$ coordinates across frequencies, it undergoes discrete steps between kernel values used. This effectively results in small baseline position errors, which shift periodically in frequency, resulting in power being mixed into the window and a relatively strong harmonic at the frequency of the shifting (the faint line in the upper left of the window).

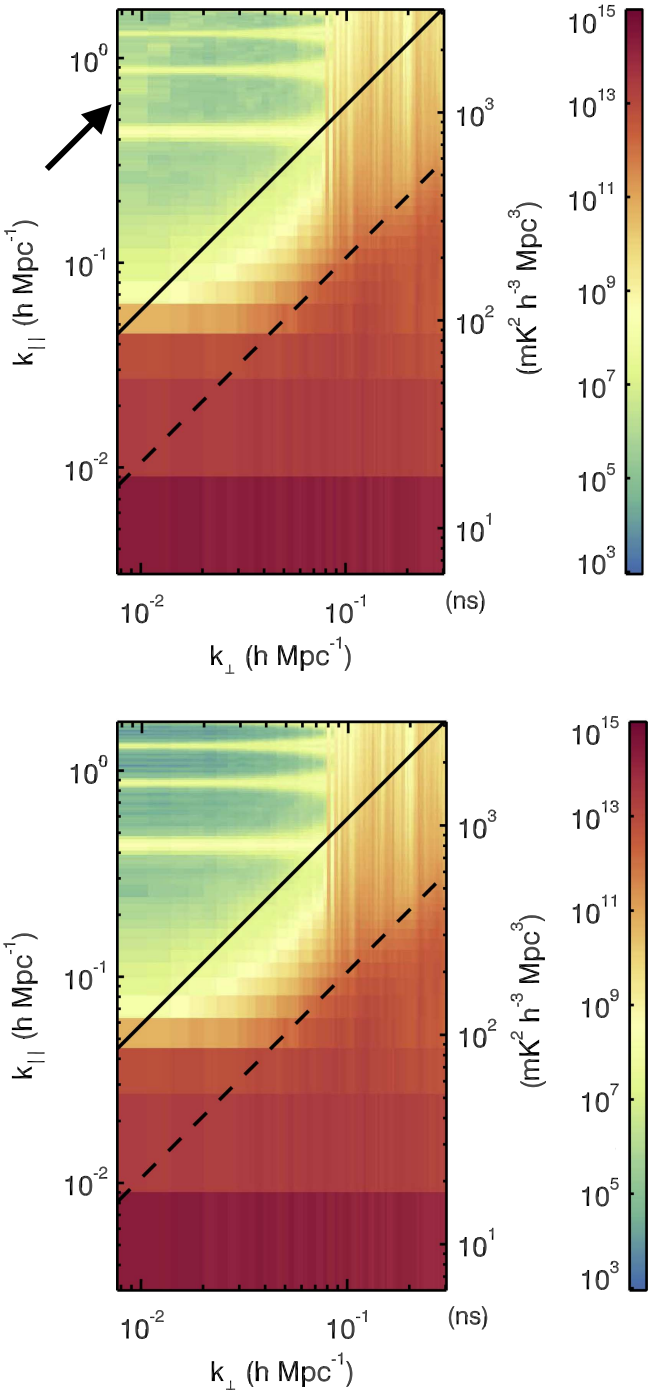


**Figure 5.** Top: single polarization model power spectrum demonstrating the contamination in the EoR window when using insufficient gridding kernel resolution. The window has a floor comparable to an expected EoR signal, and a faint super-horizon line appears at high $k_{||}$. The black arrow indicates the location and direction of the faint line. Bottom: model power spectrum after increasing the gridding kernel resolution from 0.04 wavelengths to 0.007 wavelengths. The floor is now far below the expected EoR signal level.

While we illustrate this effect with a model power spectrum, the same problem exists in the data, but is below the noise levels at three hours of integration.

This effect is distinct from more fundamental wedge effects (e.g., Morales et al. 2012; Hazelton et al. 2013) in that it is a result of the computational limitation of the analysis. The primary contribution to the foreground mode mixing is due to the loss of information as each baseline samples a range of $(u, v)$ modes. On the other hand, the effect discussed here is due to small positional offsets in the gridding kernel and only exists after gridding.

We resolved this issue by increasing the resolution at which we form the kernel. While the most accurate answer is to model at infinite resolution, it is not computationally feasible to do so. Instead, we chose a resolution at which the effect no longer materially impacts the power spectrum. With experimentation, we found that a kernel resolution of 0.007 wavelengths was sufficient. The resulting model power spectrum is shown in the bottom panel of Figure 5. The contamination within the EoR window now drops significantly lower. Of course, without knowing the exact level of the cosmological signal, we cannot know if this level is sufficiently low. This effect may need to be revisited if the EoR power spectrum is lower than predicted. The improvement in beam model resolution came at the cost of 20% increase in memory usage for our imaging pipeline.

### 4.3. Improving the Point-source Model

Our pipeline uses two modes of FHD—the full deconvolution mode to identify point sources and build a catalog, and a production mode, where we calibrate using the sky model and subtract it from the data without further fitting. Because full deconvolution on every observation from the MWA is computationally not feasible, we currently restrict ourselves to the golden data set to build our model. The process of building the catalog is presented in Carroll et al. (2016). They applied machine learning classification methods to select reliable detections from the full deconvolution FHD mode, culminating in the KGS catalog.[37] The work here uses an early iteration of the KGS catalog that was available at the time of analysis.

It has been shown that subtracting a foreground model strictly within the main lobe of the primary beam will not be sufficient to suppress the power spectrum wedge and unlock the EoR window (Thyagarajan et al. 2015a, 2015b; Pober et al. 2016). We are in the process of repeating the model building described above using additional MWA observations pointed away from our field of interest. Until that work is complete, we supplement the extent of our point-source model using additional catalogs. We accomplish this through a hierarchical catalog pulling from our early KGS catalog, the MWA Commissioning Survey[38] (MWACS, Hurley-Walker et al. 2014), the Culgoora catalog (Slee 1977), and the Molonglo Reference Catalog (MRC, Large et al. 1981). Source flux densities are prioritized in this order based on our confidence in their predicted flux density at 182 MHz. We first cluster the source lists to avoid redundant sources, using a 3.5 arcmin neighborhood radius, we then select the flux density from the highest priority catalog. Spectral indices were obtained from the MWACS and Culgoora catalogs when available, otherwise a two-point spectral index was estimated for Culgoora–MRC matches or MRC–SUMSS matches. All other sources were given a spectral index of −0.8, the previously determined median spectral index of sources below 1.4 GHz (Oort et al. 1988; Hunstead 1991; De Breuck et al. 2000; Mauch et al. 2003). The spectral index was used

[37] KGS is an abbreviation of KATALOGSS, the KDD (Knowledge Discovery in Databases) Astrometry, Trueness, and Apparent Luminosity of Galaxies in Snapshot Surveys.

[38] The more complete GLEAM Survey (Wayth et al. 2015; Hurley-Walker et al. 2017) was not available at the time of analysis.

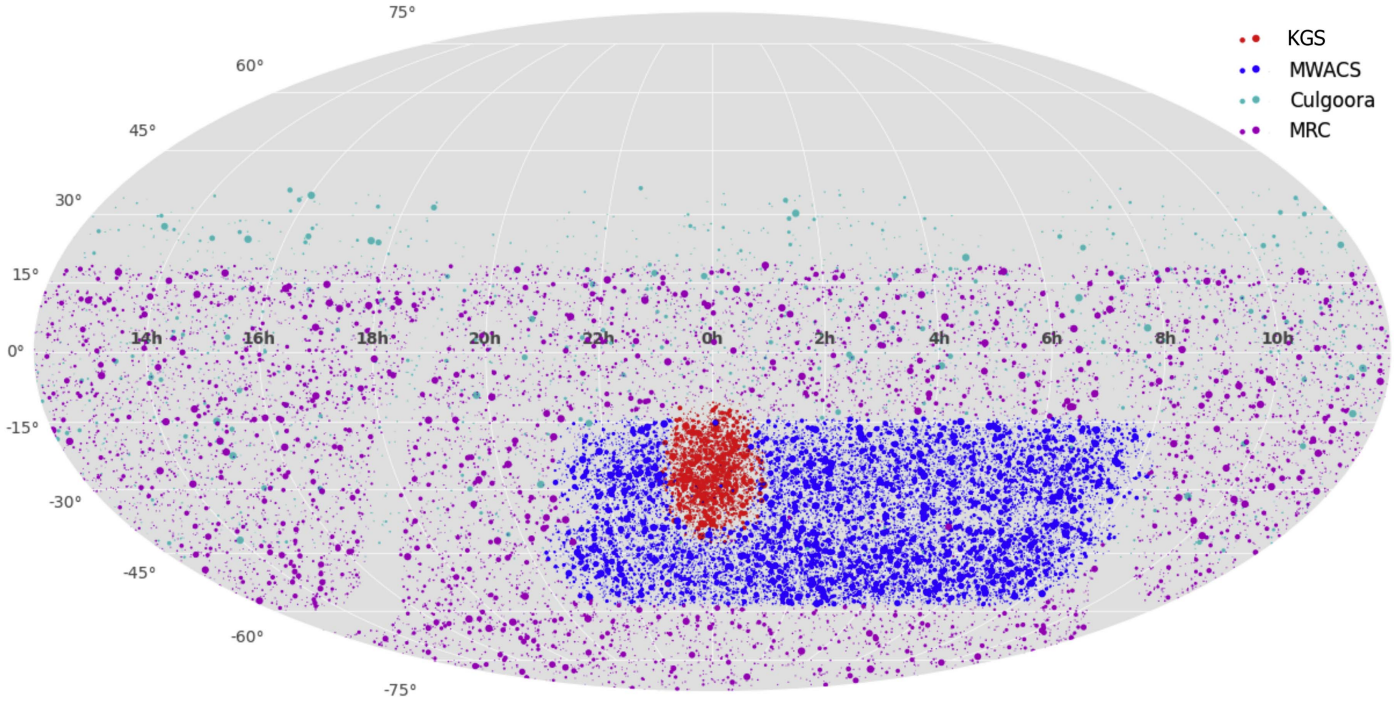


**Figure 6.** Hierarchical catalog used for foreground subtraction. This catalog combines the source lists from the KGS catalog, the MWACS catalog, Culgoora sources, and the MRC, prioritizing in that order. The EoR0 field corresponds to the red KGS patch, while we use the other catalogs to fill in the sidelobes of the MWA. The size of each dot is proportional to the 182 MHz flux density of the source, clipped at 20 Jy.

to extrapolate the flux density from the catalog frequency to 182 MHz, but a uniform spectral index of −0.8 was used within our band when forming model visibilities for calibration and foreground subtraction.

The resulting hierarchical catalog is shown in Figure 6. The EoR0 field corresponds to the red KGS source patch. All sources above the horizon and within a primary beam value greater than 1% maximum (including sidelobes) are included in both our calibration and foreground subtraction models.

To show the effect our hierarchical catalog has on the power spectrum FoM, we ran our entire analysis pipeline on the golden data set with two foreground models—first with only the MWACS catalog, and again with the full hierarchical catalog. We then compare the resulting 2D power spectra to inspect whether the new catalog results in more accurate calibration and foreground subtraction.

Because we use the input foreground model to calibrate our visibilities, the gain estimates for our two runs differed. In particular, we saw the overall flux scale of the calibrated visibilities was higher when using the full hierarchical catalog, owing to a more complete sky model requiring lower amplitude gains to describe the data. This difference is on the order of a part in $10^3$ in $\mathrm{mK}^2$ units, but is largely amplified when observing the difference in the power spectra. In order to put both residual 2D power spectra onto the same scale for comparison, we first divide each ($k_\perp$, $k_\parallel$) cell by the corresponding dirty power spectrum pixel, then subtract to arrive at a ratio difference,

$$\texttt{ratio difference} = \frac{\texttt{residual}_1}{\texttt{dirty}_1} - \frac{\texttt{residual}_2}{\texttt{dirty}_2}. \quad (6)$$

Here we use subscript 1 to represent the MWACS catalog spectra, and 2 for the hierarchical catalog spectra. The ratio difference is shown in Figure 7. The entire wedge being positive (blue) demonstrates that using the new catalog resulted in successfully subtracting a higher fractional power. Not only does this indicate that we subtract more power, but it also confirms that our calibrated data is more closely matched to the model, meaning that our calibration solutions, in general, are more accurate.

Line et al. (2016) performed a similar analysis, but with an emphasis on the positional accuracy of catalog entries. By using simulations of visibilities, they were able to isolate the effect of using perfect versus offset source positions. In their simulations, source position offsets of 14% of the synthesized beam-width had a significant impact on foreground subtraction, confirming the importance of a complete and accurate point-source catalog in order to retain a clean EoR window.

### 4.4. Diffuse Foreground Model Subtraction

In addition to a point-source foreground model, we also introduce a diffuse emission model within the primary field of view of the MWA. Though the EoR0 field was chosen to be relatively devoid of Galactic emission, we find this diffuse structure still highly contaminating due to the very high sensitivity of the MWA at large scales. For computational purposes, we have divided the diffuse emission into two regimes—the faint clouds in our main lobe, and the bright plane and other structures in the sidelobes. While the latter has been studied by many people and a Global Sky Model (GSM) is readily available (de Oliveira-Costa et al. 2008), the computational obstacle of simulating the instrument response to a full-sky diffuse model is yet prohibitive, though under active pursuit (Thyagarajan et al. 2015b). We instead focus here on the diffuse structure within the primary field and leave the full-sky model to future iterations of analysis.

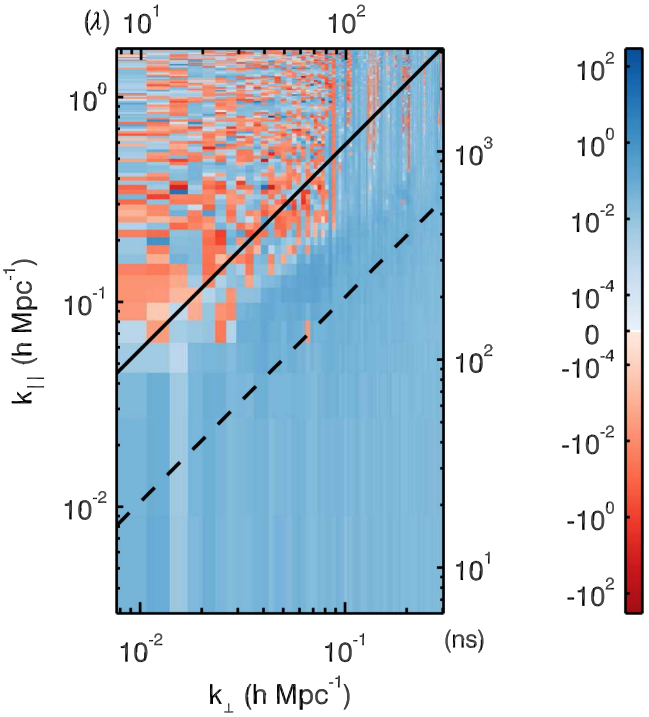


**Figure 7.** Power spectrum ratio difference, according to Equation (6). This metric shows the difference in the fractional E–W residual 2D power spectrum, comparing the MWACS catalog and our hierarchical catalog. By dividing the residual spectrum by the corresponding dirty spectrum before subtracting, we remove the effect of the overall flux scale change due to differing calibration solutions. We see that the wedge is completely positive (higher fractional residual when using the MWACS catalog), confirming that the hierarchical catalog subtracts more power from the data. This also reassures that the calibration using the new catalog is more accurate.

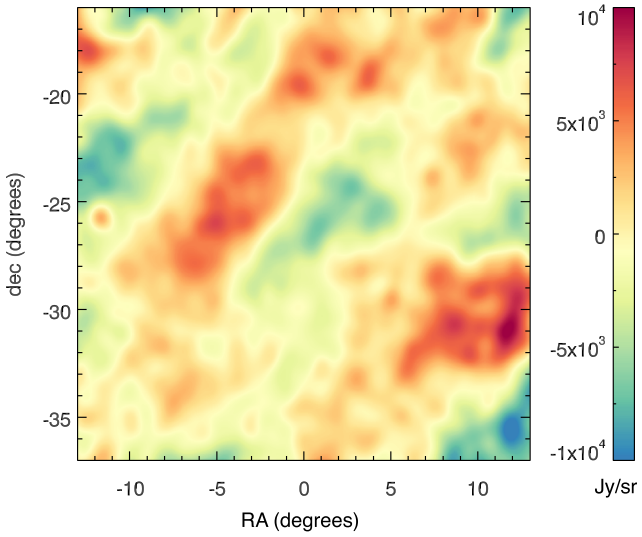


**Figure 8.** Diffuse foreground model within the EoR0 field used for foreground subtraction. This model was created using residual images from the golden data set. The image shown is smoothed to degree scales to emphasize the large-scale structure and to approximate an MWA snapshot natural weighting. Note that the negative brightness values are a consequence of not sampling the zero-spacing part of the $(u, v)$ plane.

This first iteration of modeling the diffuse structure within our primary field was done by simply using output point-source subtracted residual images of FHD from the three-hour golden data set. We combine the images from three hours to leverage the rotation of the Earth to improve the PSF of the instrument. This integrating of images is done with uniform weighted data to minimize the effect of double instrument convolution (once when the data are taken, and again when we use the output as a model). We then form a pseudo Stokes I image by adding beam-weighted E–W and N–S polarizations. We also integrate in frequency to form a single continuum image for our model. By forcing our model to contain no spectral structure, we mitigate the risk of subtracting a cosmological signal.

Future iterations of this model will contain a spectral index and multiple polarization components. Lenc et al. (2016) demonstrated the presence of strong polarized large-scale structures in the EoR0 field at 154 MHz with varying Faraday depths. They observed rotations in the Stokes Q–U plane due to ionospheric conditions, which will need to be taken into account in our future diffuse models.

Ultimately, we need model visibilities to subtract from our data. However, rather than storing a model for all observations (different pointings and phase centers), we find it simplest to treat the model as a single HEALPix image that can be used to create model visibilities for each independent observation. FHD treats each pixel in the image as a point source at the pixel location with the total flux density equal to the surface brightness of the diffuse image times the area of the pixel, and it creates model visibilities for each snapshot in the same way that it imports a catalog of point sources. This is similar to the strategy employed by Thyagarajan et al. (2015b) to model diffuse structure.

The diffuse model used for this work is shown in Figure 8. While the actual model image is uniform weighted with resolution ∼6 arcmin, we show it smoothed to degree scales to emphasize the large-scale structure and to approximate what the MWA instrument observes with a natural weighting.

For this iteration of analysis, we only use the diffuse model for subtraction, and omit it for calibration. At the time of writing, short baselines were not producing reliable results in the calibration loop. Instead, we chose to omit the diffuse model and mask baselines shorter than 50 wavelengths for calibration purposes. However, we add the diffuse model and unmask short baselines when performing foreground subtraction. We compare the total residual power in the images by squaring and summing the residual image cube from before and after the diffuse subtraction. This includes all scales measured by the instrument, but, because the image cubes are naturally weighted, the $\boldsymbol{k}$ space weighting is the inverse of Figure 3, meaning that $k_\perp \lesssim 0.05\,h$ Mpc$^{-1}$ dominates the average. We saw a 70% reduction in residual power when including our diffuse model—a strong indication that our diffuse model improves our subtraction. Future iterations will incorporate the model in calibration and gradually improve the model itself.

We demonstrate the impact of our diffuse model again by running the golden data set through our entire analysis pipeline with and without using the diffuse model. Because the foreground models used for calibration in the two runs are identical (diffuse is omitted for calibration), the calibration solutions were identical and therefore the dirty power spectra were identical as well. We compare the model and residual power spectra by direct subtraction in Figure 9. The top panel shows our point-source foreground model power spectrum minus the model power spectrum when including diffuse. Because our diffuse model added power to the full foreground model, the entire plot is negative (red). Even the EoR window is red (albeit at a much lower level than the wedge) because of

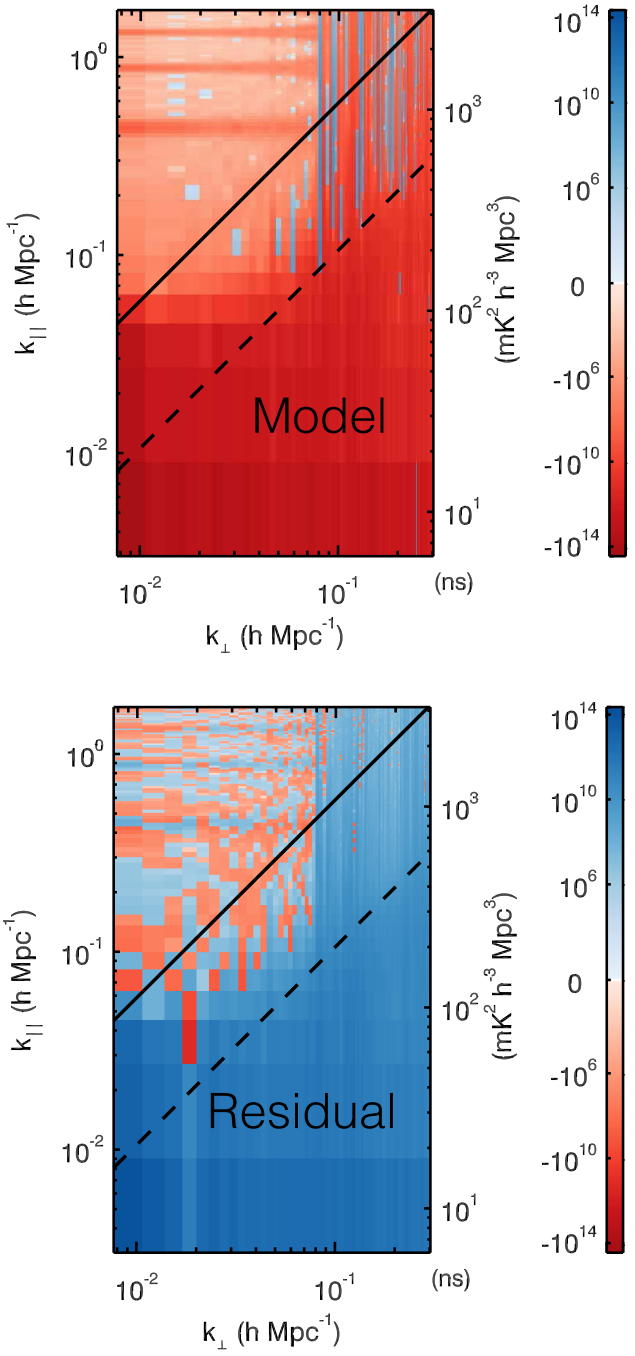


**Figure 9.** Single polarization (E–W) power spectrum differences from the golden data set showing the effect of subtracting a diffuse foreground model. The calibration is identical with and without the diffuse model, so the dirty power spectra are identical. In the top panel, the model power spectrum difference is almost entirely negative (red), indicating that the diffuse model added a large amount of power to the foreground model. The bottom panel shows the residual power spectrum difference is positive in the wedge (blue), demonstrating a successful subtraction. While this first attempt at a diffuse model was rudimentary, it successfully removed 70% of the residual power.

power leakage from non-uniform spectral sampling and from the dynamic range limit of the Blackman–Harris window function. The bottom panel is the difference in residual power. The wedge is completely positive (blue), indicating that the diffuse model successfully subtracted from the dirty visibilities. The differences in the window are positive and negative, indicating that the differences in the foreground model in this region are below the three-hour noise level.

## 5. A DEEP INTEGRATION

Up to this point we have only discussed testing and analysis on the three-hour golden data set. We next turn toward a deeper integration, incorporating the techniques described above. We start with all MWA observations taken on the EoR0 field at 182 MHz between 2013 August 23 and November 29. This includes 2780 snapshots, or about 86.5 hours of data. We first make data quality cuts, then form power spectra.

### 5.1. Data Selection

All 2780 snapshots in the data set are preprocessed, calibrated, and imaged using the pipeline described in the previous sections of this article. Similar to tests on the golden data set, we rely heavily on the 2D power spectrum as a diagnostic tool, this time to identify and excise poor-quality data.

We use the jackknife method to filter out bad data and detect patterns. This involves dividing the data into many subgroups and forming power spectra. The goal is to find observational parameters that affect the quality of the data, such as the day or time of night. A powerful grouping is to divide the data into the observation day and "pointing" (which direction the antennas were pointed while tracking the EoR0 field). We show an example of one day's worth of per-pointing power spectra in Figure 10. The pointings are labeled sequentially with −5 corresponding to five pointing steps prior to zenith transit, 0 corresponding to zenith, and +4 corresponding to four steps after zenith. This numbering scheme is shown graphically in Figure 11. Early in the night (−5 through −3) the bulk of the Galactic plane was still above the horizon, and, despite being very far from our primary field of view, highly contaminated our observations through the sidelobes of the instrument. We then have relatively well-behaved pointings, with the EoR window dominated by noise, until the final pointing of the night, when we can see evidence of the Galaxy rising again indicated by strong lines of power at the edge of the foreground wedge. We saw the same contamination on all days of observation and decided to cut all −5, −4, −3, and +4 pointings from our data set. In principle, it may be possible to model the Galactic plane well enough to account for its presence, but we leave this to future work.

Motivated by the manual classification done with power-spectrum jackknives, we developed another metric to quickly predict the power spectrum quality for each snapshot in our data set. We do this by forming a delay spectrum (Parsons et al. 2012) from the raw, uncalibrated visibilities. We then calculate an estimated total EoR window power by integrating the power above the horizon line and below the first coarse-band line. While we would ideally use calibrated visibilities for this metric, we have found that the uncalibrated power is strongly correlated to the calibrated power, and the cuts we make below are independent of calibration. This is encouraging because future analysis could use this metric before processing the snapshots, saving valuable computing resources.[39]

[39] Because calibration involves generating model visibilities for each snapshot, approximately half of the computational cost of the entire imaging pipeline is required to obtain calibration solutions.

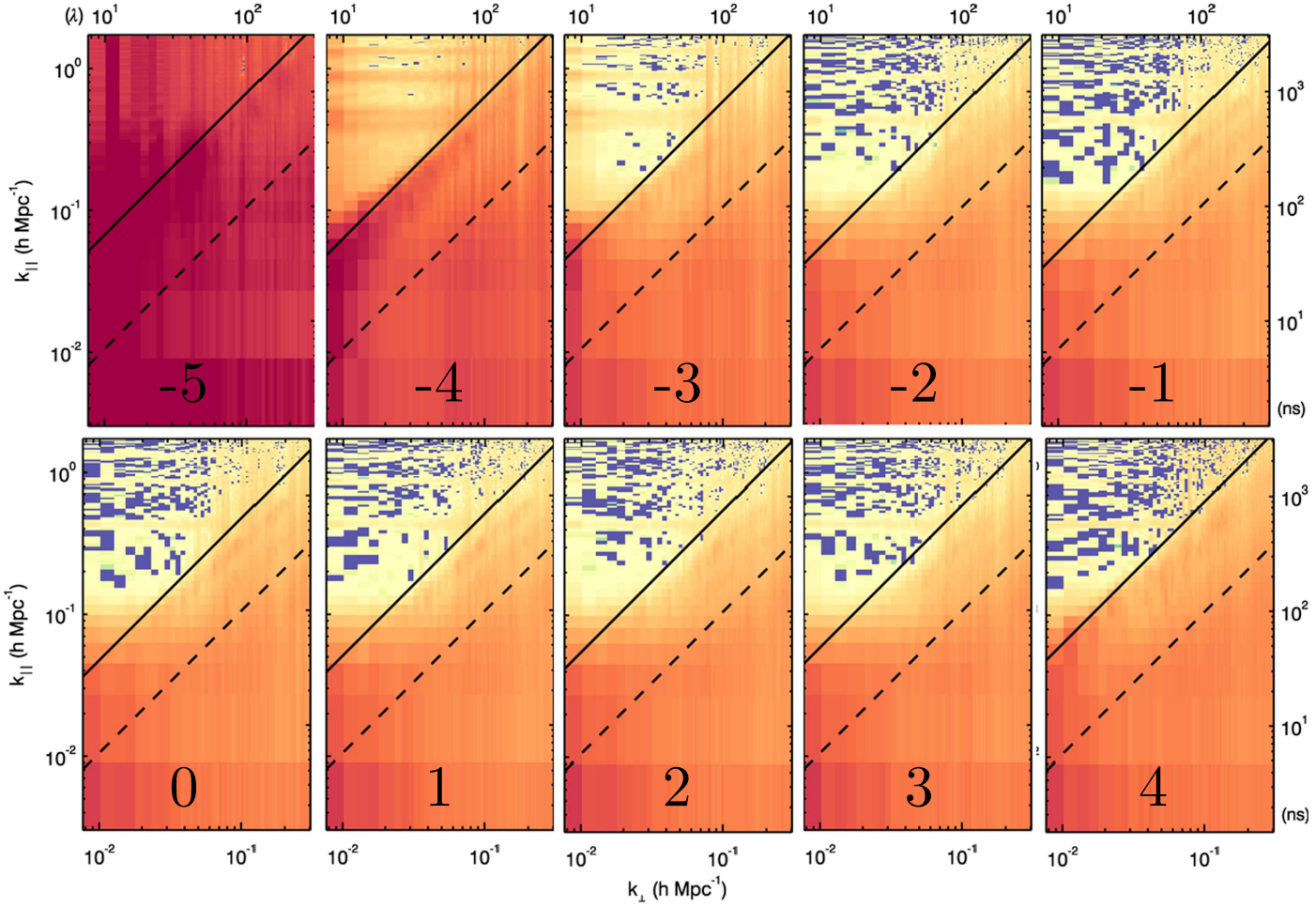


**Figure 10.** Example jackknife test. For this test, we divided the data into days and pointings. This is an example array of power spectra (residual E–W polarization) for a single day, 2013 August 26. The early pointings are heavily contaminated by the Galaxy in the sidelobes, the window becomes more clear near zenith, and we can see trace contamination at the end of the night (pointing +4) when the Galaxy has risen again.

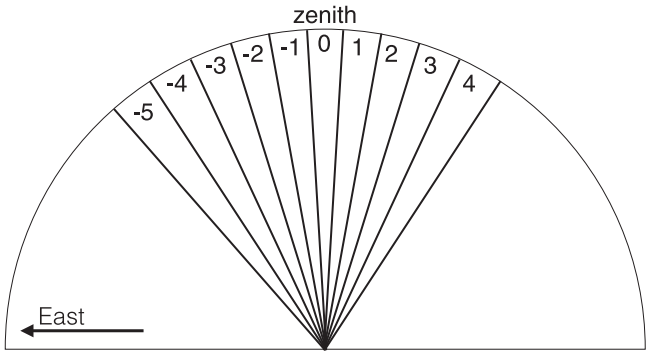


**Figure 11.** Cartoon depiction of the numbering scheme used to label pointings in Figures 10 and 12. Looking south, a field transits from the east (left), and the telescope begins observing about 5 pointings before zenith (−5). As the field drifts overhead, we periodically repoint the telescope to recenter the field. Each shift increments the pointing label by one, and the zenith pointing is defined as zero. While a + 5 pointing (and beyond) is possible with the MWA, this observing campaign did not contain any such data, because we instead switched to the EoR1 field at that time.

We plot the window power for each snapshot in our data set in Figure 12(a). We first note that our conclusion from the jackknife tests is confirmed here—early pointings contain strong contamination and have high window power. We also see many outlier snapshots with excess power. Inspecting power spectra from these individual snapshots confirmed poor data quality, even after calibration. The cause of these contaminants is yet unknown, but could easily be attributed to low-level RFI that was missed by the AOFLAGGER, or intermittent hardware failures. To remove the poor data, we made the cut shown with the black box, keeping only snapshots inside.

Next we compare the power in our two instrumental polarizations. When conducting our jackknife tests, we saw contamination could exist in one polarization and not the other —potentially due to RFI, or an instrumental failure. We plot the ratio of N–S power to E–W power in Figure 12(b), after applying the previous cut. We see that the ratio is generally flat, with the exception of some outliers and almost all of pointing +3. We suspect that the excess N–S power in the later pointing is due to the Galaxy leaking back into the sidelobe of the N–S polarization, but not in the other. Again, we make the cut shown with a black box.

Our final data cut was made when inspecting the snapshot residual continuum images output from FHD. We estimate a proxy for the residual flux density from foreground sources by calculating the rms of the fractional residual flux densities for all subtracted sources greater than 0.5 Jy within half-beam power. The fractional residual flux density is measured as the ratio of the residual image pixel value (Jy) at the source

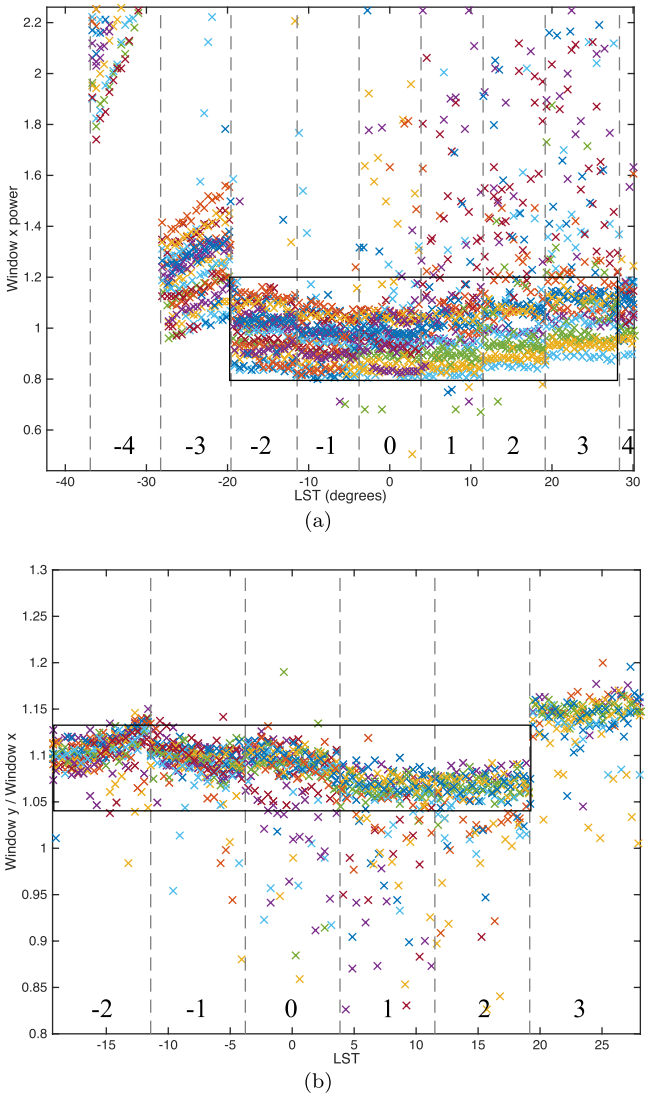


**Figure 12.** Top: window power for each observation in our data set. The window power is calculated from the delay spectrum of uncalibrated data as a fast quality metric. Because uncalibrated data is used, the units are arbitrary. Each color is a distinct day of observing, and the vertical dashed lines represent pointing shifts, with the pointing numbers indicated at the bottom of the plot above the horizontal axis. The black box indicates observations that passed this cut. Bottom: data cut based on window power polarization ratio. For each snapshot that passed the total power cut, we plot the ratio of window powers in the N–S and E–W polarizations. Clear outliers can be seen, including the whole of pointing +3. The black box again indicates the selected data.

position to the integrated flux density of the source. We found that most snapshots had residual flux density rms $<10\%$ with some outliers, which we cut from the data. This cut largely overlapped with previous cuts, but removed an additional 95 observations. The cause of the high-residual flux density rms is yet unknown and is not localized by observation day, LST, pointing, or any other jackknife we have performed.

After the cuts described above, we are left with 1029 snapshots, or just over 32 hours of data. This is an aggressive cut, with the aim of removing all poor-quality data and retaining a data set as clean as we can determine from our data quality checks. As mentioned earlier, more sophisticated analysis could allow us to recover data cut from this study in the future (e.g., modeling and removing the Galaxy in observations pointing far off zenith).

### 5.2. Results

In this section, we discuss the results after processing the remaining 32 hours of data through the $\epsilon$ppsilon pipeline, and place an upper limit on the EoR power spectrum.

The bandwidth of the MWA and our data set is 30.72 MHz, but in order to avoid the effects of cosmic evolution over the span of our measured redshift range, we limit observations to about 8 MHz, or $\Delta z \sim 0.3$ at our frequency. We do this by dividing the band into three overlapping sub-bands of 15.36 MHz each, which, after applying the Blackman–Harris window function, will result in effective bandwidths of 7.68 MHz. We label these sub-bands as "low" ($z \approx 7.1$), "mid" ($z \approx 6.8$), and "high" ($z \approx 6.5$).

The 32-hour integrated residual power spectra for our three sub-bands and both instrumental polarizations are shown in Figure 13. A number of features can be seen in these spectra. In all cases the foreground wedge is prominent, with extra power at large scales (low $k_\perp$). This is an indication that our diffuse model needs to improve for deeper subtraction.

The lowest region of the EoR window, above the horizon line and around $k_\parallel \approx 0.25\, h\,\mathrm{Mpc}^{-1}$, contains purely positive bins, indicating non-noise-like power. The cause for this seeming leakage is yet unknown, and it limits our integration in this analysis. One suspected origin of this contamination was individual snapshots with high contamination that slipped through our cuts. However, dividing the data into random subsets and comparing results indicated that there is not a small number of offending observations, instead, the leakage appears to exist at a low level in all of the data.

Another potential cause of the foreground leakage into the EoR window is insufficient calibration quality—particularly the spectral shape of the instrumental bandpass. Investigations are underway to combine snapshot observations for higher signal-to-noise calibration solutions, as well as more sophisticated parameterization of the gain model in Equation (4) to account for more physical effects, such as dependence on ambient temperature variation.

Moving up in the window, between the first two coarse-band harmonic lines we can see an additional faint line. This is the re-emergence of the 150 m cable reflections. While the method described in Section 3.2 sufficiently calibrated out this reflection line for the three-hour golden set integration, the lower noise level in this deeper integration shows it is not completely removed. Future analysis will require higher signal to noise on the reflection fitting, which may require combining snapshots for calibration solutions.

Between the coarse-band harmonic lines and the reflection line, we do see regions where our spectra appear noise-like (positive and negative values). This is encouraging, despite the leakage at lower $k_\parallel$, and motivates us to be selective when binning to 1D power spectra.

Figure 14 shows slices of power that help isolate the contamination. The two-dimensional power spectrum is the N–S, mid-band spectrum from Figure 13. We have drawn a horizontal solid blue line to indicate the slice used to plot power as a function of $k_\perp$ for a fixed $k_\parallel$ (top right), and a vertical dashed line to show the slice used to plot power as a function of $k_\parallel$ for a fixed $k_\perp$ (bottom right). In the $k_\perp$ power plot, we can see leakage at high $k_\perp$ from the residual of the vertical streaks due to poor sampling at large $(u, v)$. While the increased noise would mean that we will down weight the bins at large $k_\perp$, the contamination is even larger, meaning it would

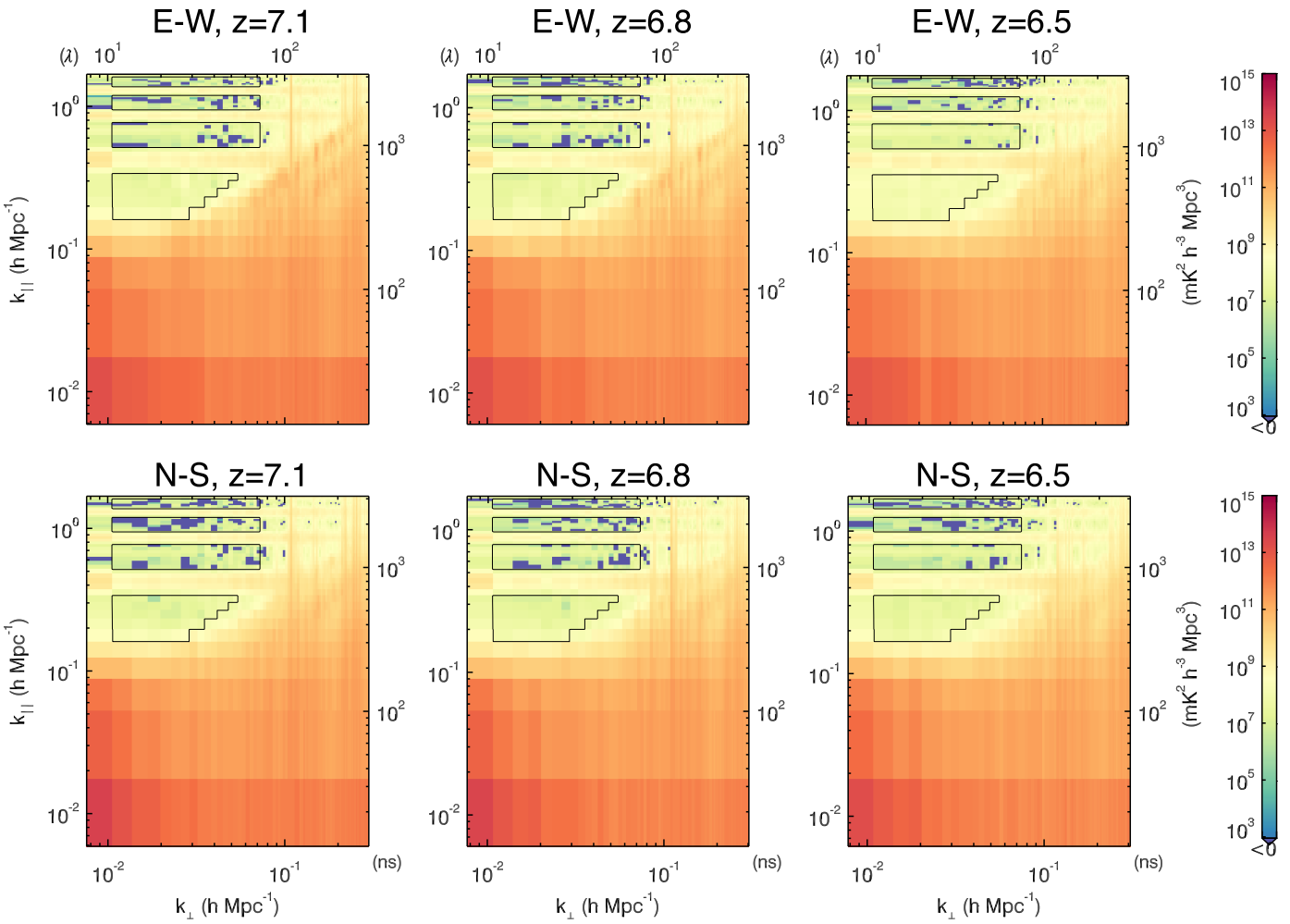


**Figure 13.** Residual two-dimensional power spectra for the three sub-bands used to place limits on the cosmological signal. The left panels show the low band, centered at 174.7 MHz, or a redshift of 7.1. The middle panels show the mid band, centered at 182.4 MHz, or a redshift of 6.8. The right panels show the high band, centered at 190.1 MHz, or a redshift of 6.5. The top row corresponds to east–west instrumental polarization, while the bottom row shows north–south polarization. The contours show the window used for averaging to one-dimension.

bias our estimate even though those bins are down weighted. The low end of $k_\perp$ contains exceptionally bright foregrounds at large scales, which, due to the width of the coarse-band harmonics, contaminate most of the $k_\parallel$ modes. With both of these regimes in consideration, we exclude bins with $u < 10\lambda$ and $u > 70\lambda$, shown with gray boxes. Between these two regimes are many bins consistent with the noise level, which we include in our 1D averages.

Similarly, we inspect the $k_\parallel$ power in the bottom right plot of Figure 14. Here, we see the huge contamination from foregrounds at low $k_\parallel$. While the leakage drops significantly, our measurements do not reach the noise level until after the first coarse-band harmonic. However, we chose to include bins below the first coarse band with $k_\parallel > 0.15\,h\,\mathrm{Mpc}^{-1}$ because the cosmological signal is expected to contain substantially higher power at large scales, and ultimately these low $k_\parallel$ bins provide our most competitive limits, despite being systematic limited. We exclude the coarse-band harmonics by masking out the harmonic bin itself and two bins on either side (total of five bins per harmonic).

The final mask we apply is the wedge. We found that a small buffer beyond the horizon was necessary to completely mask out the wedge, consistent with Dillon et al. (2015a). We implement the buffer by increasing the slope of the horizon line by 14%. This line is shown as the dashed diagonal line in the left panel of Figure 14.

The various masking described above is summarized with the black contours shown on the two-dimensional power spectra in Figure 13. The contours show the cuts in two-dimensional space, but the masking is actually performed directly in the 3D power spectrum cube and averaged directly to one-dimension.

The resulting 1D power spectra are shown in Figure 15, where the measured power is shown with solid blue, the $1\sigma$ noise level is shown with thin red, and the $2\sigma$ upper limit for each bin is shown with magenta. Where our unbiased estimator is negative, the absolute value is shown, but with a dotted blue line. For consistency with the literature, we plot our 1D power spectra as $\Delta^2(k) = k^3 P_{21}(k)/(2\pi^2)$, which has units of $\mathrm{mK}^2$.

In all bands and polarizations, we are heavily signal dominated in our most sensitive region (low $k$). This is not surprising based on the 2D spectra we examined earlier. The gaps in the data are due to the excised coarse-band harmonics. We can see the cable reflection line between the first two gaps in all polarizations and sub-bands. Between the coarse-band harmonic lines, especially at larger $k$, are bins that approach the

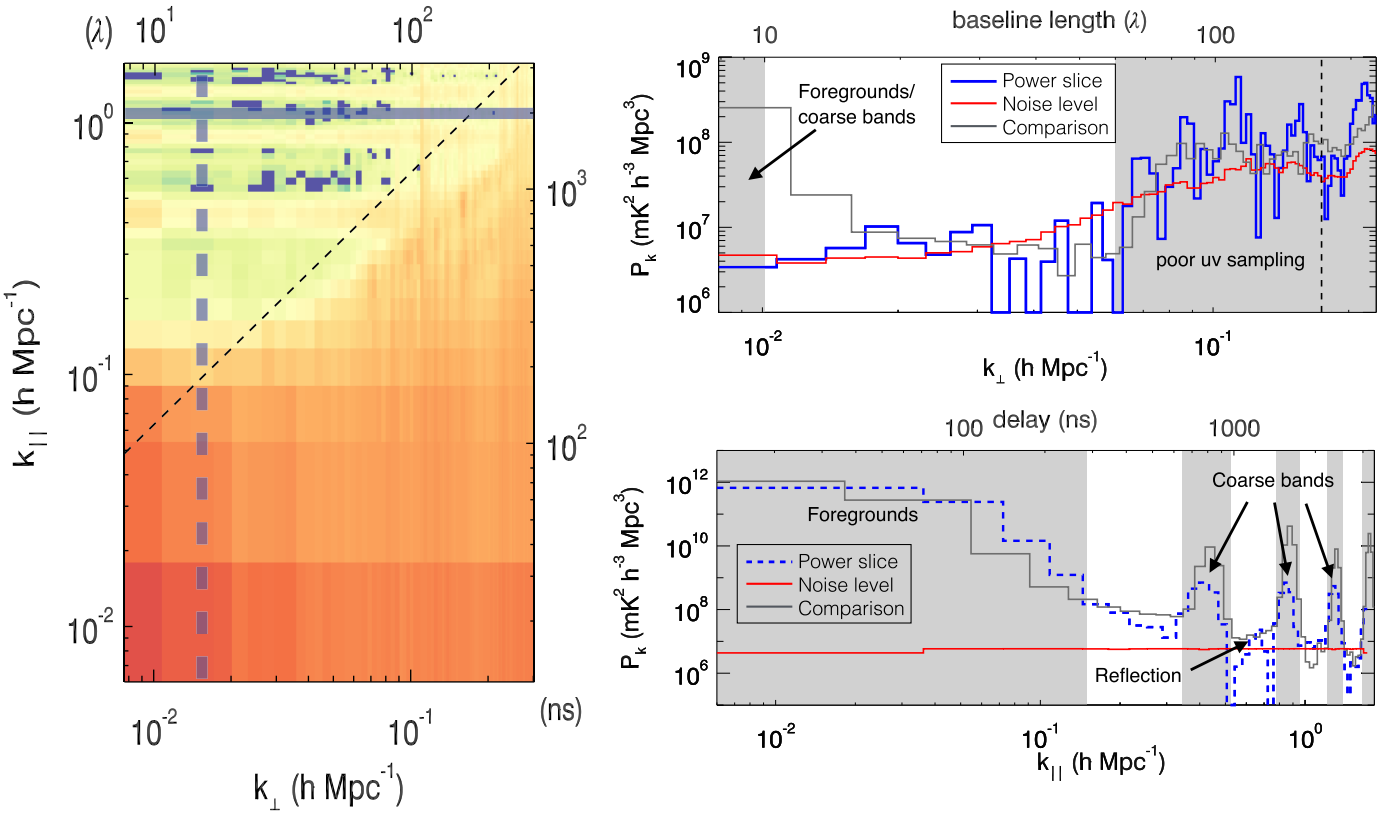


**Figure 14.** Left: the N–S, $z = 6.8$ two-dimensional power spectrum repeated from Figure 13. Here we have superimposed blue lines to show slices in $k_\perp$ (solid) and $k_\parallel$ (dashed). Top right: residual power as a function of $k_\perp$ for fixed $k_\parallel$, averaged in rings from the full 3D power cube (blue). The gray boxes show regions that will be excluded from the 1D averages. The red line shows the $1\sigma$ noise level, and the vertical black dashed line shows the intersection of this slice with the wedge. Bottom right: residual power as a function of $k_\parallel$ for fixed $k_\perp$, again averaged in rings from full 3D power cube (blue). The strong foreground contamination is evident at low $k_\parallel$, and the coarse-band harmonic lines appear as expected. The thin gray line in both right panels is the equivalent power slice from the RTS+CHIPS comparison pipeline, which is discussed in Section 5.3.

noise level and are consistent with zero (notably the N–S polarization for the low and mid bands). The bin just above the first coarse-band harmonic is also consistent with zero in the low band. These regions are encouraging because they have the potential to continue integrating down with more data without further analysis improvements. Further suppression of the coarse-band harmonic and cable reflection lines will also improve the limits presented here.

Our best upper limits for a cosmological signal are at low $k$, despite the strong leakage. Because any foreground leakage should not correlate with the EoR signal, we can assume that the power of the sum of leakage and cosmological signal is greater than the power of the EoR signal alone, allowing us to set an upper limit. We quote the best $2\sigma$ upper limits, $\Delta^2_{\rm UL}$, for each polarization and band in Table 1. These limits and their context in the field are further discussed in Section 6.

### 5.3. Comparison with Reference Pipeline

As demonstrated in Jacobs et al. (2016), comparison between independent analysis pipelines is essential to evaluate the validity of algorithms under active development. Such comparison provides insight into the strengths and weaknesses of the pipelines, as well as confirmation of results. The analysis discussed so far has been based on the FHD-to-ϵppsilon pipeline, and here we compare with the Real Time System (RTS, Mitchell et al. 2008; Ord et al. 2010) to Cosmological H I Power Spectrum (CHIPS, Trott et al. 2016) pipeline. Below, we present a brief overview of the reference pipeline and highlight the lessons learned from the comparison.

The RTS uses a fundamentally different calibration and foreground subtraction approach than FHD, as well as a different input point-source model. The primary differences lie in the use of ionospheric corrections, the bandpass calibration, and the lack of a diffuse model. RTS accomplishes the interferometric calibration in two steps: a direction-independent calibration to a single compound calibrator created by combining the 1000 apparently brightest sources in the sky, followed by an individual calibration and subtraction of those sources inside the Calibration Measurement Loop (CML, Mitchell et al. 2008). The CML estimates individual ionospheric corrections for all sources and performs a full direction-dependent calibration for the 5 very brightest ones. These tasks are distributed over independent parallel threads for each 1.28 MHz coarse channel.

The bandpass calibration is also fundamentally different to FHD. The bandpass for each antenna is fitted to a third-order polynomial within each 1.28 MHz coarse band and each direction and polarization individually. The delay of the 150 m cable reflection discussed earlier is comparable to the scale of the polynomial fit, so we expect it to naturally be fit out of the gain solutions. Indeed, the RTS has not yet seen evidence of the reflection in the processed data (Jacobs et al. 2016).

The source catalog used with RTS was a combined cross-match created with the Positional Update and Matching Algorithm (PUMA, Line et al. 2016). Full details of PUMA

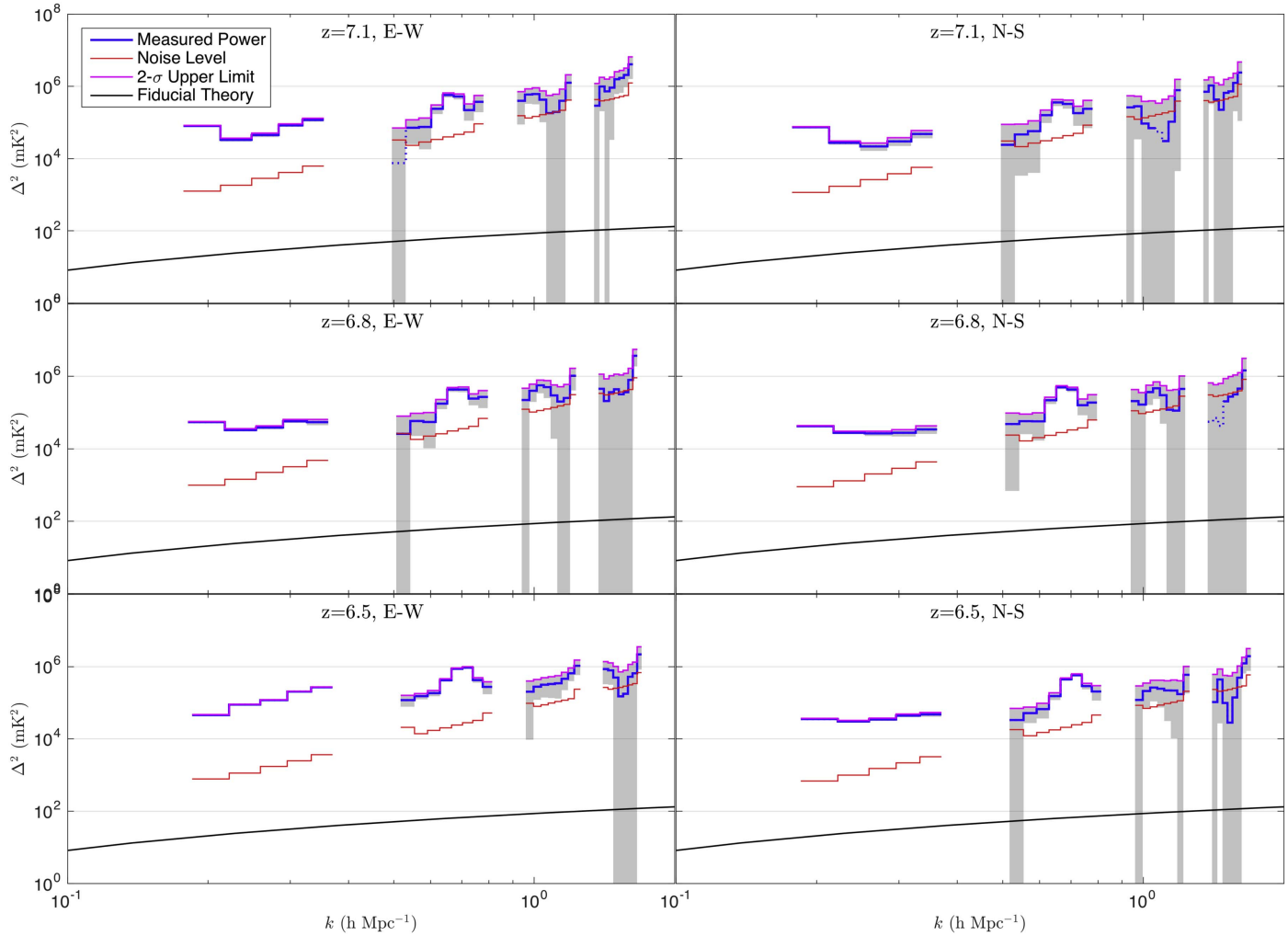


**Figure 15.** 1D power spectra for our three sub-bands and both instrumental polarizations. The solid blue line shows the measured power spectrum with step widths corresponding to the bin size used in the average. Where the measured signal is negative, we plot the absolute value with a dotted line. The gray boxes show the $\pm 2\sigma$ error bars on the measured power spectrum. Where the boxes meet the horizontal axis we are consistent with zero. The thin red line is the $1\sigma$ noise level, and the magenta line is the $2\sigma$ upper limit for each $k$ bin. A fiducial theoretical model for a fully neutral IGM from Furlanetto et al. (2006) is shown in black for reference.

**Table 1**
Upper Limits on the EoR Power Spectrum for Our Three Sub-bands and Two Polarizations

| | | | FHD+ϵppsilon | | RTS+CHIPS | |
|---|---|---|---|---|---|---|
| Sub-band | $z_0$ | Pol | $k$ | $\Delta^2_{\rm UL}$ | $k$ | $\Delta^2_{\rm UL}$ |
| Low | 7.1 | E–W | 0.231 | $3.67 \times 10^4$ | | |
| Low | 7.1 | N–S | 0.27 | $2.70 \times 10^4$ | 0.16 | $3.2 \times 10^4$ |
| Mid | 6.8 | E–W | 0.24 | $3.56 \times 10^4$ | | |
| Mid | 6.8 | N–S | 0.24 | $3.02 \times 10^4$ | 0.14 | $2.6 \times 10^4$ |
| High | 6.5 | E–W | 0.20 | $4.70 \times 10^4$ | | |
| High | 6.5 | N–S | 0.24 | $3.22 \times 10^4$ | 0.14 | $2.5 \times 10^4$ |

**Note.** Upper limits, $\Delta^2_{\rm UL}$, are at the 97.7% confidence level. Cosmological wave numbers, $k$, are in Units of $h$ Mpc$^{-1}$ and upper limits are in units of mK$^2$. The last two columns, RTS+CHIPS, are produced from the reference pipeline, which is discussed in Section 5.3.

and this cross-matched catalog can be found in Line et al. (2016), but we include a brief description here. PUMA is designed to cross-match low radio-frequency ($\lesssim$1 GHz) catalogs. It utilizes a Bayesian probabilistic positional cross-match approach based on Budavári & Szalay (2008), combined with criteria based on catalog resolution and fitting the spectral energy distribution to a power-law. These criteria allow for surveys with differing resolutions, which can introduce confusing sources into the matched process. Primarily, the source catalog was based on the MWACS (Hurley-Walker et al. 2014), which was cross-matched to the 74 MHz Very Large Array Low Frequency Sky Survey redux (VLSSr, Lane et al. 2012), the MRC, (Large et al. 1981), the 843 MHz Sydney University Molonglo Sky Survey (SUMSS, Mauch et al. 2003), and the 1.4 GHz NRAO VLA Sky Survey (NVSS, Condon et al. 1998). MWACS covers approximately $20\overset{\rm h}{.}5 < {\rm R.A.} < 8\overset{\rm h}{.}5$, $-58° < \delta < -14°$, so to complete the sky coverage needed, MRC was used as a base catalog outside of the MWACS coverage and was matched to VLSSr, SUMSS and NVSS.

Since RTS has the added step of individual source peeling, there is a greater possibility for numerical calibration failure. Before passing to CHIPS, a simple quality assurance step is considered, which is based upon the variance of residual visibilities in each 1.28 MHz coarse band. Observations with divergent or unstable calibrations were readily identified by variances far outside the otherwise observed distribution of values. Of the 1029 observations calibrated and used by the

FHD pipeline, the RTS+CHIPS pipeline successfully calibrated and included 1003. The remainder failed to calibrate successfully, potentially due to adverse ionospheric conditions.

The CHIPS power spectrum estimator applies a full maximum-likelihood estimator to the data. In its full form, the data are weighted by the thermal noise, and a model for the residual point-source signal, within a full frequency-dependent description of the instrument. For this comparison, the foreground weighting is not used, in order for a direct comparison with the weighting scheme used by ϵppsilon. The data were divided into the same three redshift bins, and the 2D and 1D power spectra formed.

Table 1 displays comparison $2\sigma$ upper limits, demonstrating broad consistency between the pipelines. The RTS+CHIPS best limits appear at a lower $k$ bin than FHD+ϵppsilon, owing to the different structure of the foreground leakage for the two pipelines. Figure 14, which shows the resulting FHD+ϵppsilon slices in $k_{\parallel}$ and $k_{\perp}$, also displays the equivalent power profiles for RTS+CHIPS. The comparison pipeline shows significantly more power at low $k_{\perp}$, consistent with no diffuse model being subtracted from the data. Aside from this area, the two pipelines show consistent results. Figure 16 compares the 1D power from both pipelines for the three central redshifts studied in this work. The magenta line reproduces the FHD+ϵppsilon results from Figure 15, while the gray line plots the equivalent RTS+CHIPS results (with coarse-band harmonics included). Despite the differences in the two pipelines, they produce consistent power across a range of scales, supporting the robustness of the results.

## 6. DISCUSSION

Inspecting our 32-hour integrated power spectra we see two places where we are limited. First, there is a region that is leakage-dominated at low $k$. The cause of this leakage is yet unknown, but likely due to imperfect calibration, beam models, and/or foreground models. In order to improve on our limit in this regime, future analysis will require calibration that better accounts for the instrumental response, perhaps by increasing signal to noise through multiple-snapshot solutions or refined antenna response models. Through simulation, Thyagarajan et al. (2016) recently demonstrated the necessity of precise foregrounds and instrumental models in order to retain a clean EoR window. Studies of the beam are underway (e.g., Neben et al. 2015; Sutinjo et al. 2015), and will likely help to further isolate the foreground contaminates in our power spectra. The foreground model is a constant area of investigation, but improved spectral dependence in both the point-source catalog and the diffuse model, as well as the addition of an all-sky Galactic model, will continue to improve the foreground subtraction and further unlock the EoR window.

The second limited region in our 1D power spectra is between the coarse-band harmonics where our measured signal approaches the noise level. Additional data will likely reduce the noise, and therefore our upper limit, in this regime. However, until the leakage at low $k$ is understood, progress at higher $k$ is susceptible to running into the same systematic with longer integrations.

Finally, we place our best upper limits in context with the 21 cm EoR field as a whole. A direct comparison between measurements from different instruments is difficult due to the varying methods, redshifts, and scales probed. However, if we approximate the power spectrum in units of $\mathrm{mK}^2$ to be roughly

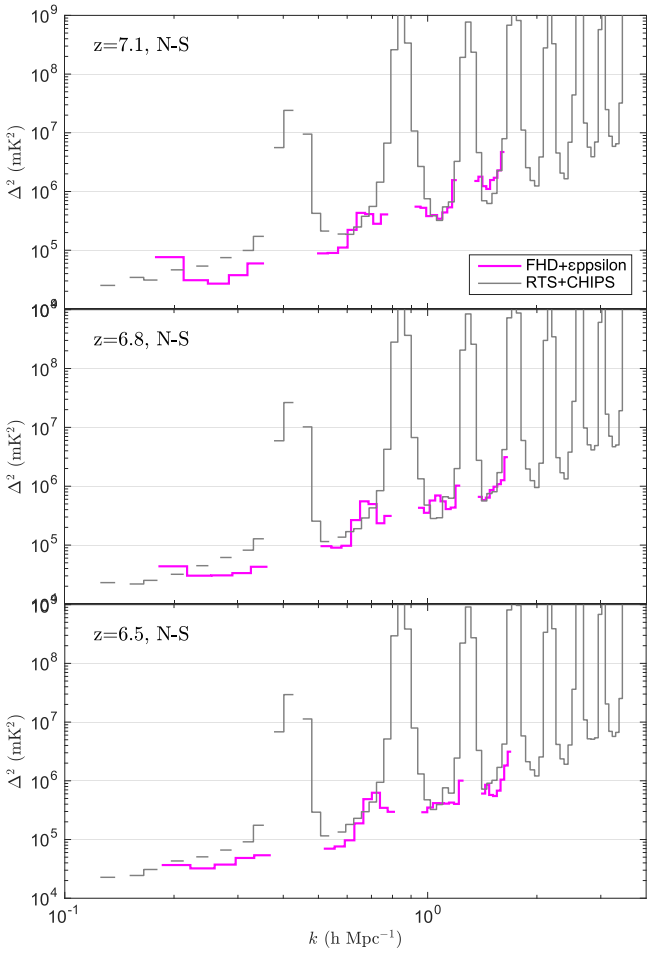


**Figure 16.** Here we compare the upper limits from the FHD+ϵppsilon pipeline with those from the RTS+CHIPS pipeline. The magenta lines are repeated from Figure 15, and the gray lines are the power plus $2\sigma$ thermal error bars lines generated by CHIPS. Due to the logarithmic binning used by CHIPS, some bins are empty, resulting in gaps in the spectra. CHIPS does not excise the coarse-band harmonic lines, and so we see their strong effect where there are gaps in the magenta lines. The RTS+CHIPS pipeline accesses larger $k$ modes because the data is processed at 80 kHz, whereas the FHD+ϵppsilon pipeline averages to 160 kHz when gridding. The most notable differences are the ability of the RTS+CHIPS pipe to access lower $k$ modes, while the FHD+ϵppsilon pipe suppresses more power before and after the first coarse-band harmonic. While the different analysis methods result in slightly different upper limits, they are generally in agreement.

flat over the scales probed by the MWA, PAPER, and GMRT, we can glean a view of the current state of the field. The best result from this analysis is an upper limit on the power spectrum of $\Delta^2 \leqslant 2.7 \times 10^4$ $\mathrm{mK}^2$ at $k = 0.27\ h$ $\mathrm{Mpc}^{-1}$ and $z = 7.1$. This is a modest improvement over the previous best MWA results (Dillon et al. 2015a), which is not surprising - given that this integration contains about 10 times more data, but is systematics limited. Our results are also consistent with the 40 hr GMRT limit of $\Delta^2 \leqslant 6.15 \times 10^4$ $\mathrm{mK}^2$ at $k = 0.5\ h$ $\mathrm{Mpc}^{-1}$ and $z = 8.6$ (Paciga et al. 2013), though they probe significantly different periods of the EoR. The front runner in terms of an upper limit is the PAPER experiment (Parsons et al. 2014; Ali et al. 2015; Jacobs et al. 2015). The PAPER-64 limit of $\Delta^2 \leqslant 502$ $\mathrm{mK}^2$ at $z = 8.4$ from a 135-day observing campaign is about one and a half orders of magnitude lower than our best limit in power units, though again probing a significantly different redshift. The MWA analysis is rapidly improving as we uncover and mitigate systematics to realize

the full sensitivity of the telescope. The red lines in Figure 15 indicate the noise level possible to reach with these data if systematics can be overcome.

### 6.1. Summary of Planned Analysis Improvements

Through the reduction of the first season of MWA high-band EoR observations, we have identified several opportunities to improve our analysis. Here, we compile a list of future improvements and their potential impact on the power-spectrum results. Broadly speaking, these improvements emphasize a need to better understand the foregrounds and the instrument, and are likely to be relevant to imaging based analyses in general.

*Extended point-source model*. The hierarchical catalog used here was a definite improvement over the MWACS catalog alone. This point-source model will continue to improve with the release of the GLEAM Survey, as well as dedicated MWA observations to target sources in the sidelobes of the EoR fields. Improving the foreground catalog in our sidelobes will be crucial in controlling the foreground leakage into the EoR window (Pober et al. 2016). Incorporating these improved catalogs into our analysis will improve both the foreground subtraction and the calibration.

*Diffuse emission model*. The diffuse model here contained no spectral information and represented only the Stokes I polarization. Nevertheless, subtracting this model removed 70% of the residual power in the foreground wedge. Adding a spectral index and multiple polarization components to the model will improve this subtraction even further. As Lenc et al. (2016) demonstrated, substantial polarized diffuse structure exists in the EoR0 field. The contributions to Stokes Q and U depend on the ionospheric conditions, which will be accounted for in future analyses to appropriately model this emission. In addition, the diffuse model will be extended to an all-sky model. It is evident from the jackknives shown in Figure 10 that strong emission near the horizon is leaking into our power spectra during off-zenith observations. By modeling and subtracting this emission we may be able to recover these observations in future analyses.

*Primary beam model*. In order to properly model the instrument response to the sky we also need an accurate primary beam model. This is especially important for EoR power spectrum measurements because the (often bright) near-horizon emission resides at the edge of the foreground wedge (Thyagarajan et al. 2015a, 2015b; Pober et al. 2016), but the low elevation beam can be difficult to model at the necessary precision (Thyagarajan et al. 2016). Asad et al. (2016) demonstrated the need for a polarized primary beam model in order to mitigate the risk of diffuse polarized emission leaking into Stokes I, conflating with a potential cosmological measurement. The MWA primary beam model is continuously improving to incorporate mutual coupling between the phased dipoles, and embedded element patterns for each of the 32 dipoles on each antenna (Sutinjo et al. 2015). Other methods are also being pursued to directly measure the full beam response in situ using the ORBCOMM satellite constellation (Neben et al. 2015), and drone-flown calibrator sources (D. C. Jacobs et al. 2016, in preparation). The latter method is similar to the strategy Virone et al. (2014) used to verify SKA beam patterns.

*Calibration*. The most limiting factor in our analysis is likely to be insufficiently precise calibration solutions, which can cause foreground power to leak into the EoR window, as we see in the integrated 2D power spectra (Figure 13). Because our calibration is based on sky and instrument models, the improvements described above will improve our calibration. We also plan to combine many snapshot observations when estimating gains and achieving higher signal to noise, which in turn will enable more sophisticated parameterizations of the antenna bandpasses. This will require an overhaul of the data flow depicted in Figure 1 because snapshots will need to be combined in the calibration step. Preliminary investigations have shown that the gains are stable enough to be fit across several hours of observation.

Between 2013 and the time of writing, the MWA has collected over 2,000 hours of data targeting the EoR fields. These observations include the three designated sky fields and two frequency bands. In addition, observations are underway to further characterize the sources in the instrument sidelobes. The second phase of the MWA will be commissioned in the latter half of 2016, and will include two sets of highly redundant cores of antennas. This hybrid configuration will enable a unique opportunity to study the calibration techniques used in imaging and delay spectrum analyses, providing another view of the systematics limiting our current results.

While the analysis presented here has not reached full potential, it represents the deepest power spectrum integration to date produced by an imaging pipeline. Imaging involves many difficulties (e.g., efficient gridding and mapmaking, foreground modeling), but if systematics can be overcome, it has the potential to compete with other more targeted experiments and analysis styles. More broadly, imaging analyses will be necessary to perform cross-correlation studies with complimentary probes such as galaxy surveys or other intensity mapping experiments (e.g., Lidz et al. 2009; Doré et al. 2014; Beardsley et al. 2015; Silva et al. 2015; DeBoer et al. 2016; Vrbanec et al. 2016), which will ultimately unlock the full potential of 21 cm observations. Here, we have identified a crucial region of power spectrum space that is currently contaminated, with suggestions for improvement in future analysis. With improved calibration techniques, primary beam models, and understanding of foregrounds, the MWA and other imaging analyses will be able to quickly approach the most competitive upper limits in the field.

This work was supported by NSF grants AST-1410484 and AST-1206552. D.C.J. is supported by an NSF Astronomy and Astrophysics Postdoctoral Fellowship under award AST-1401708. This scientific work makes use of the Murchison Radio-astronomy Observatory, operated by CSIRO. We acknowledge the Wajarri Yamatji people as the traditional owners of the Observatory site. Support for the operation of the MWA is provided by the Australian Government (NCRIS), under a contract to Curtin University administered by Astronomy Australia Limited. We acknowledge the Pawsey Supercomputing Centre, which is supported by the Western Australian and Australian Governments.

## REFERENCES

Ali, Z. S., Parsons, A. R., Zheng, H., et al. 2015, ApJ, 809, 61
Asad, K. M. B., Koopmans, L. V. E., Jelić, V., et al. 2016, arXiv:1604.04534
Barry, N., Hazelton, B., Sullivan, I., Morales, M. F., & Pober, J. C. 2016, MNRAS, 461, 3135
Beardsley, A. P., Hazelton, B. J., Morales, M. F., et al. 2012, MNRAS, 425, 1781

Beardsley, A. P., Hazelton, B. J., Morales, M. F., et al. 2013, MNRAS Lett., 429, L5
Beardsley, A. P., Morales, M. F., Lidz, A., Malloy, M., & Sutter, P. M. 2015, ApJ, 800, 128
Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, ApJS, 208, 20
Bhatnagar, S., Cornwell, T. J., Golap, K., & Uson, J. M. 2008, A&A, 487, 419
Bouwens, R. J., Bradley, L., Zitrin, A., et al. 2014, ApJ, 795, 126
Bowman, J. D., Cairns, I., Kaplan, D. L., et al. 2013, PASA, 30, 31
Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2006, ApJ, 638, 20
Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2009, ApJ, 695, 183
Bowman, J. D., & Rogers, A. E. E. 2010, Natur, 468, 796
Budavári, T., & Szalay, A. S. 2008, ApJ, 679, 301
Carroll, P. A., Line, J., Morales, M. F., et al. 2016, MNRAS, 461, 4151
Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693
Datta, A., Bowman, J. D., & Carilli, C. L. 2010, ApJ, 724, 526
De Breuck, C., van Breugel, W., Röttgering, H. J. A., & Miley, G. 2000, A&AS, 143, 303
DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2016, arXiv:1606.07473
de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., et al. 2008, MNRAS, 388, 247
Dillon, J. S., Liu, A., & Tegmark, M. 2013, PhRvD, 87, 043005
Dillon, J. S., Liu, A., Williams, C. L., et al. 2014, PhRvD, 89, 023002
Dillon, J. S., Neben, A. R., Hewitt, J. N., et al. 2015a, PhRvD, 91, 123011
Dillon, J. S., Tegmark, M., Liu, A., et al. 2015b, PhRvD, 91, 023002
Doré, O., Bock, J., Ashby, M., et al. 2014, arXiv:1412.4872
Ewall-Wice, A., Dillon, J. S., Hewitt, J. N., et al. 2016, MNRAS, 460, 4320
Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, AJ, 132, 117
Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, PhR, 433, 181
Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, SSRv, 123, 485
Greig, B., Mesinger, A., & Pober, J. C. 2016, MNRAS, 455, 4295
Greisen, E. 2012, AIPS FITS File Format, AIPS Memo, 117
Hazelton, B. J., Morales, M. F., & Sullivan, I. S. 2013, ApJ, 770, 156
Hunstead, R. W. 1991, AuJPh, 44, 743
Hurley-Walker, N., Callingham, J. R., Hancock, P. J., et al. 2017, MNRAS, 464, 1146
Hurley-Walker, N., Morgan, J., Wayth, R. B., et al. 2014, PASA, 31, 45
Jacobs, D. C., Hazelton, B. J., Trott, C. M., et al. 2016, ApJ, 825, 114
Jacobs, D. C., Pober, J. C., Parsons, A. R., et al. 2015, ApJ, 801, 51
Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18
Lane, W. M., Cotton, W. D., Helmboldt, J. F., & Kassim, N. E. 2012, RaSc, 47 RS0K04
Large, M. I., Mills, B. Y., Little, A. G., Crawford, D. F., & Sutton, J. M. 1981, MNRAS, 194, 693
Lenc, E., Gaensler, B. M., Sun, X. H., et al. 2016, arXiv:1607.05779
Lidz, A., Zahn, O., Furlanetto, S. R., et al. 2009, ApJ, 690, 252
Line, J. L. B., Webster, R. L., Pindor, B., Mitchell, D. A., & Trott, C. M. 2016, PASA, in press, arXiv:1611.05534
Liu, A., Parsons, A. R., & Trott, C. M. 2014a, PhRvD, 90, 023018
Liu, A., Parsons, A. R., & Trott, C. M. 2014b, PhRvD, 90, 023019
Liu, A., & Tegmark, M. 2011, PhRvD, 83, 103006
Loeb, A., & Furlanetto, S. 2013, The First Galaxies in the Universe (Princeton, NJ: Princeton Univ. Press)
Malloy, M., & Lidz, A. 2013, ApJ, 767, 68
Mauch, T., Murphy, T., Buttery, H. J., et al. 2003, MNRAS, 342, 1117
McGreer, I. D., Mesinger, A., & D'Odorico, V. 2015, MNRAS, 447, 499
Mellema, G., Koopmans, L. V. E., Abdalla, F. A., et al. 2013, ExA, 36, 235
Mitchell, D., Greenhill, L., Wayth, R., et al. 2008, ISTSP, 2, 707
Morales, M., & Wyithe, J. 2010, ARA&A, 48, 127
Morales, M. F., Bowman, J. D., & Hewitt, J. N. 2006, ApJ, 648, 767
Morales, M. F., Hazelton, B., Sullivan, I., & Beardsley, A. 2012, ApJ, 752, 137
Morales, M. F., & Hewitt, J. N. 2004, ApJ, 615, 7
Morales, M. F., & Matejek, M. 2009, MNRAS, 400, 1814
Neben, A. R., Bradley, R. F., Hewitt, J. N., et al. 2015, RaSc, 50, 614
Offringa, A. R., de Bruyn, A. G., Biehl, M., et al. 2010, MNRAS, 405, 155
Offringa, A. R., Trott, C. M., Hurley-Walker, N., et al. 2016, MNRAS, 458, 1057
Offringa, A. R., Wayth, R. B., Hurley-Walker, N., et al. 2015, PASA, 32, 8
Oort, M. J. A., Steemers, W. J. G., & Windhorst, R. A. 1988, A&AS, 73, 103
Ord, S. M., Mitchell, D. A., Wayth, R. B., et al. 2010, PASP, 122, 1353
Paciga, G., Albert, J. G., Bandura, K., et al. 2013, MNRAS, 433, 639
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
Parsons, A. R., Pober, J. C., Aguirre, J. E., et al. 2012, ApJ, 756, 165
Patil, A. H., Yatawatta, S., Zaroubi, S., et al. 2016, arXiv:1605.07619
Patra, N., Subrahmanyan, R., Raghunathan, A., & Udaya Shankar, N. 2013, ExA, 36, 319
Planck Collaboration, Adam, R., Aghanim, N., et al. 2016, arXiv:1605.03507
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2015a, arXiv:1502.01589
Planck Collaboration, Aghanim, N., Arnaud, M., et al. 2015b, arXiv:1507.02704
Pober, J. C., Ali, Z. S., Parsons, A. R., et al. 2015, ApJ, 809, 62
Pober, J. C., Hazelton, B. J., Beardsley, A. P., et al. 2016, ApJ, 819, 8
Pober, J. C., Parsons, A. R., Aguirre, J. E., et al. 2013, ApJL, 768, L36
Prabu, T., Srivani, K., Roshi, D., et al. 2015, ExA, 39, 73
Pritchard, J. R., & Loeb, A. 2012, RPPh, 75, 086901
Salvini, S., & Wijnholds, S. J. 2014, A&A, 571, A97
Scargle, J. D. 1982, ApJ, 263, 835
Silva, M., Santos, M. G., Cooray, A., & Gong, Y. 2015, ApJ, 806, 209
Slee, O. B. 1977, AuJPA, 43, 1
Sokolowski, M., Tremblay, S. E., Wayth, R. B., et al. 2015, PASA, 32, e004
Sullivan, I. S., Morales, M. F., Hazelton, B. J., et al. 2012, ApJ, 759, 17
Sutinjo, A., O'Sullivan, J., Lenc, E., et al. 2015, RaSc, 50, 52
Thyagarajan, N., Jacobs, D. C., Bowman, J. D., et al. 2015a, ApJL, 807, L28
Thyagarajan, N., Jacobs, D. C., Bowman, J. D., et al. 2015b, ApJ, 804, 14
Thyagarajan, N., Parsons, A. R., DeBoer, D. R., et al. 2016, ApJ, 825, 9
Thyagarajan, N., Shankar, N. U., 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. 2014, PASA, 31, 26
Trott, C. M., Pindor, B., Procopio, P., et al. 2016, ApJ, 818, 139
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., Shankar, N. U., & Subrahmanyan, R. 2012, ApJ, 745, 176
Virone, G., Lingua, A., Piras, M., et al. 2014, IAWPL, 13, 169
Voytek, T. C., Natarajan, A., Jáuregui García, J. M., Peterson, J. B., & López-Cruz, O. 2014, ApJL, 782, L9
Vrbanec, D., Ciardi, B., Jelić, V., et al. 2016, MNRAS, 457, 666
Wayth, R. B., Lenc, E., Bell, M. E., et al. 2015, PASA, 32, e025
Yatawatta, S., de Bruyn, A. G., Brentjens, M. A., et al. 2013, A&A, 550, A136
Zaroubi, S. 2013, The First Galaxies, Vol. 396 ed. T. Wiklind, B. Mobasher, & V. Bromm, 45