

First limits on the 21 cm power spectrum during the Epoch of X-ray heating

A. Ewall-Wice,^{1,2★} Joshua S. Dillon,^{1,2,3} J. N. Hewitt,^{1,2} A. Loeb,⁴ A. Mesinger,⁵
A. R. Neben,^{1,2} A. R. Offringa,^{6,7} M. Tegmark,^{1,2} N. Barry,⁸ A. P. Beardsley,⁹
G. Bernardi,^{4,10,11} Judd D. Bowman,⁹ F. Briggs,^{7,12} R. J. Cappallo,¹³ P. Carroll,⁸
B. E. Corey,¹³ A. de Oliveira-Costa,² D. Emrich,¹⁴ L. Feng,^{1,2} B. M. Gaensler,^{7,15,16}
R. Goeke,² L. J. Greenhill,⁴ B. J. Hazelton,^{8,17} N. Hurley-Walker,¹⁴
M. Johnston-Hollitt,¹⁸ Daniel C. Jacobs,⁹ D. L. Kaplan,¹⁹ J. C. Kasper,^{4,20}
HS Kim,^{7,21} E. Kratzenberg,¹³ E. Lenc,^{7,15} J. Line,^{7,21} C. J. Lonsdale,¹³
M. J. Lynch,¹⁴ B. McKinley,^{7,21} S. R. McWhirter,¹³ D. A. Mitchell,^{7,22}
M. F. Morales,⁸ E. Morgan,² Nithyanandan Thyagarajan,⁹ D. Oberoi,²³ S. M. Ord,^{7,14}
S. Paul,²⁴ B. Pindor,^{7,21} J. C. Pober,^{8,25} T. Prabu,²⁴ P. Procopio,^{7,21} J. Riding,^{7,21}
A. E. E. Rogers,¹³ A. Roshi,²⁶ N. Udaya Shankar,²⁴ Shiv K. Sethi,²⁴ K. S. Srivani,²⁴
R. Subrahmanyam,^{7,24} I. S. Sullivan,⁸ S. J. Tingay,^{7,14} C. M. Trott,^{7,14}
M. Waterson,^{12,14} R. B. Wayth,^{7,14} R. L. Webster,^{7,21} A. R. Whitney,¹³
A. Williams,¹⁴ C. L. Williams,^{1,2} C. Wu²⁷ and J. S. B. Wyithe^{7,21}

Affiliations are listed at the end of the paper

Accepted 2016 April 28. Received 2016 April 26; in original form 2015 September 18

ABSTRACT

We present first results from radio observations with the Murchison Widefield Array seeking to constrain the power spectrum of 21 cm brightness temperature fluctuations between the redshifts of 11.6 and 17.9 (113 and 75 MHz). 3 h of observations were conducted over two nights with significantly different levels of ionospheric activity. We use these data to assess the impact of systematic errors at low frequency, including the ionosphere and radio-frequency interference, on a power spectrum measurement. We find that after the 1–3 h of integration presented here, our measurements at the Murchison Radio Observatory are not limited by RFI, even within the FM band, and that the ionosphere does not appear to affect the level of power in the modes that we expect to be sensitive to cosmology. Power spectrum detections, inconsistent with noise, due to fine spectral structure imprinted on the foregrounds by reflections in the signal-chain, occupy the spatial Fourier modes where we would otherwise be most sensitive to the cosmological signal. We are able to reduce this contamination using calibration solutions derived from autocorrelations so that we achieve a sensitivity of 10^4 mK on comoving scales $k \lesssim 0.5 h \text{ Mpc}^{-1}$. This represents the first upper limits on the 21 cm power spectrum fluctuations at redshifts $12 \lesssim z \lesssim 18$ but is still limited by calibration systematics. While calibration improvements may allow us to further remove this contamination, our results emphasize that future experiments should consider carefully the existence of and their ability to calibrate out any spectral structure within the EoR window.

* E-mail: aaronew@MIT.EDU

Key words: techniques: interferometric – dark ages, reionization, first stars – radio lines: general – X-rays: galaxies.

1 INTRODUCTION

Mapping the 21 cm transition of neutral hydrogen at high redshift promises to revolutionize our knowledge on the first generations of stars and galaxies and to provide a unique probe of the ‘Dark Ages’ preceding this first generation of luminous objects (see Barkana & Loeb 2001; Furlanetto, Oh & Briggs 2006; Morales & Wyithe 2010 for reviews). Planned instruments such as the Square Kilometre Array (SKA; Koopmans et al. 2015) and the Hydrogen Epoch of Reionization Array (HERA; Pober et al. 2014) are expected to elucidate the formation of the first luminous structures and to place strict constraints on the properties of the sources that reionized the intergalactic medium (Pober et al. 2014; Greig & Mesinger 2015). A number of experiments including the Giant Metrewave Telescope (GMRT; Paciga et al. 2013), the Low-Frequency Array (LOFAR; van Haarlem et al. 2013), the Murchison Widefield Array (MWA; Bowman et al. 2013; Tingay et al. 2013a), and the Precision Array for Probing the Epoch of Reionization (PAPER; Parsons et al. 2010) are already underway to explore the challenges of separating the faint cosmological signal from bright foregrounds and to attempt a first detection of the power spectrum of the cosmological 21 cm emission line.

Thus far, these experiments have targeted redshifts between 6 and 12. During this Epoch of Reionization (EoR), ultraviolet photons from the first generations of luminous sources transformed the intergalactic medium (IGM) from predominantly neutral to ionized. Over the past several years, deep integrations have placed significant upper limits on the power spectrum during reionization (Paciga et al. 2013; Dillon et al. 2014; Parsons et al. 2014; Ali et al. 2015; Dillon et al. 2015b; Trott et al. 2016). The best upper limits of 502 mK^2 at $z \sim 8.4$ (Ali et al. 2015) have begun to rule out scenarios where the neutral IGM experiences little or no heating (Pober et al. 2015). Integrations at comparatively high redshifts have also been carried out. Dillon et al. (2014) put an upper limit on the power spectrum at $z = 11.7$ using the 32-tile MWA pathfinder. A much deeper integration at redshift 10.3 was performed with PAPER’s 32-element configuration (Jacobs et al. 2015), though it was limited by residual foregrounds at the edge of the instrumental bandpass.

While observations of the power spectrum during the EoR alone will shed light on the sources and astrophysics that drove reionization, it is only the final milestone in the evolution of the neutral IGM. Before reionization, the gas was heated, most likely by the first generations of high mass X-ray binaries (HMXB; Mirabel et al. 2011) and/or hot interstellar medium (ISM; Pacucci et al. 2014). Brightness temperature fluctuations from inhomogeneous heating at these early times can yield power spectrum amplitudes that are over an order of magnitude larger than those expected during reionization (Pritchard & Furlanetto 2007; Mesinger, Ferrara & Spiegel 2013). Even before the X-ray heating, fluctuations in the brightness temperature were likely sourced by fluctuations in the Lyman α flux field from the first stars (Barkana & Loeb 2005; Pritchard & Furlanetto 2006).

The ultimate goal of 21 cm cosmology is a three-dimensional map of the entire IGM between $z \approx 200$ and reionization since, at least in principle, the 21 cm line is a cosmological observable accessible all the way back through the dark ages to the decoupling of the spin temperature from the cosmic microwave background (CMB; Furlanetto et al. 2006). Even though the ionosphere obscures

extraterrestrial radio signals below about 30 MHz (Jester & Falcke 2009), it is expedient to use ground-based experiments to cover as great a redshift span as possible. Because foreground amplitudes and ionospheric effects grow progressively at lower frequency, the most reasonable next step after reionization in our march into the dark ages is the Epoch of X-ray heating (EoX). The exact redshift range for the EoX depends on the astrophysical model (see for example Mesinger, Ferrara & Spiegel 2013; Mesinger, Ewall-Wice & Hewitt 2014; Pacucci et al. 2014), but a reasonable range, targeted in this work, is $z = 11.6$ (113 MHz) to $z = 17.9$ (75 MHz).

First experimental 21 cm constraints on the thermal history of the IGM come from Pober et al. (2015) who used upper limits on the power spectrum at reionization redshifts to rule out inefficient heating histories. These constraints arise from the fact that the observable brightness temperature difference from the CMB depends on the spin temperature as

$$\Delta T_b \propto \left(1 - \frac{T_\gamma}{T_s}\right), \quad (1)$$

where T_γ is the temperature of the CMB and T_s is the spin temperature of the gas which is expected to be closely coupled to the gas kinetic temperature before substantial heating takes place (Furlanetto et al. 2006). For a cold IGM, $1 - T_\gamma/T_s$ is large and negative leading to large amplitude contrasts between neutral and ionized regions.

However, assuming that the number of X-rays per baryon involved in star formation is the same as what is observed in nearby star-forming galaxies (Mineo, Gilfanov & Sunyaev 2012), the H I spin temperature is expected to be heated well above the CMB by the time reionization begins, saturating the effect of heating on equation (1) (Furlanetto 2006b). Hence, direct measurements of the 21 cm line during the EoX will be necessary if we want to learn about the detailed properties of the thermal history and the astrophysical phenomena that influence it. Recent work has shown that if X-ray heating proceeds inefficiently, the current generation of interferometers will be sensitive enough to detect the power spectrum sourced by spin temperature fluctuations at $z \approx 12$ (Christian & Loeb 2013). Next generation of 21 cm observatories will detect the heating power spectrum for a wide range of heating scenarios out to redshifts as high as 20 (Mesinger et al. 2014) and place percent level constraints on the properties of the earliest X-ray sources (Ewall-Wice et al. 2016a). While pre-reionization measurements are expected to shed light on the first stellar mass black holes or the hot ISM, they may also offer us insights into other astrophysical processes. It is possible for dark matter annihilation (Valdés et al. 2013) and the existence of warm dark matter (Sitwell et al. 2014) to create observable impacts on the IGM thermal history. Finally, the IGM is especially cool and optically thick during the beginning of the heating process, making it ideal for 21 cm forest (Carilli, Gnedin & Owen 2002; Furlanetto & Loeb 2002; Furlanetto 2006a; Mack & Wyithe 2012; Ciardi et al. 2013) studies should any radio loud sources exist at those redshifts. It is also possible to constrain the source population itself by detecting its signature in the 21 cm power spectrum (Ewall-Wice et al. 2014).

Complementary observations of the sky-averaged (the ‘global’) 21 cm signal with a single dipole can also explore the reionization and pre-reionization epochs and experiments such as EDGES (Bowman & Rogers 2010), LEDA (Greenhill & Bernardi 2012),

DARE (Burns et al. 2012), SARAS (Patra et al. 2013), SciHI (Voytek et al. 2014), and BIGHORNS (Sokolowski et al. 2015) are beginning to take data. While demanding much greater sensitivity than global signal experiments, power spectrum measurements with an interferometer probe fine frequency scales while foregrounds occupy a limited region of Fourier-space, known as the ‘wedge’ (Datta, Bowman & Carilli 2010; Morales et al. 2012; Parsons et al. 2012; Trott, Wayth & Tingay 2012; Vedantham, Udaya Shankar & Subrahmanyan 2012; Hazelton, Morales & Sullivan 2013; Thyagarajan et al. 2013; Liu, Parsons & Trott 2014a,b; Thyagarajan et al. 2015a,b). The region of Fourier space outside of the wedge, in principle free of foregrounds and therefore having greater sensitivity to brightness temperature fluctuations, is known as the ‘EoR window’ (henceforth ‘window’). foreground modelling and calibration errors that are smooth in frequency should have limited impact within the window.

In this paper we assess the levels of systematic errors that are especially severe at the lower EoX frequencies (relative to those typical of EoR studies) including the ionosphere, radio-frequency interference (RFI), and the enhanced noise and foregrounds from a sky that is both intrinsically brighter at lower frequencies and observed with a larger primary beam. In Section 2 we describe the MWA, our observations, and data reduction. In Section 3 we address the systematic errors that are especially challenging below EoR frequencies and our efforts to mitigate them. The limiting systematic error that we encounter is fine frequency structure in the instrumental bandpass due to standing wave reflections in the cables between the MWA’s beamformers and receivers. After making a reasonable assumption about the relationship between our autocorrelations and the gain amplitudes, we achieve notable improvement in calibration but we are still left with significant foreground contamination.

Power spectrum upper limits are derived (Section 4) which are broadly consistent with thermal noise except in several regions of Fourier space corresponding to the light travel time delays of the reflections. We expect that refined calibration techniques employing better foreground models (Caroll et al., in preparation) and redundant baselines (Wieringa 1992; Liu et al. 2010; Zheng et al. 2014) can improve the removal of this contamination. In order to avoid signal loss and the introduction of spurious spectral structure, we have been conservative in the number of free parameters allowed in our gain solutions; increasing these may also resolve this problem. Reduced cable lengths expected in upcoming experiments such as HERA and the SKA will ameliorate the problem of reflections.

2 OBSERVING AND INITIAL DATA REDUCTION

We begin our discussion with an overview of our observations and our data reduction procedure. Our analysis yields two different image products with 112 s cadence: high resolution continuum images created from bandwidth multifrequency synthesis (MFS) (with ≈ 6 arcmin resolution) where baselines across all fine frequencies are combined into a single image, and naturally weighted multifrequency data cubes, where each fine frequency channel is imaged separately and integrated over 3 h. We use the MFS images to evaluate ionospheric conditions, and we use the multifrequency data cubes in our power spectrum analysis. We note that with 112 s averaging, we are performing significant averaging over fine time-scale ionospheric effects which for the MWA baselines have a typical coherence time of ≈ 10 –44 s (Vedantham & Koopmans 2015, hereafter V15). After outlining the instrument and our observing strategy (Section 2.1), we discuss our initial calibration procedure

(Section 2.2) and finish with the production MFS images (Section 2.3) and data cubes (Section 2.4) which serve as the input to our power spectrum pipeline.

2.1 Observations with the MWA

The MWA (Lonsdale et al. 2009; Tingay et al. 2013a) is a 128 antenna interferometer located at the Murchison Radio Observatory (MRO) in Western Australia (26.70°S, 116.67° E) with an analog bandpass of 80–300 MHz. Each correlated antenna tile consists of 16 dual polarization dipole elements arranged in a four-by-four grid. The phased output of the dipoles on each tile is summed together in an analog beam-former and delivered to one of sixteen different receiver units where 30.72 MHz of bandwidth is digitized before correlation in an on-site building. We refer the reader to Prabu et al. (2015) and Ord et al. (2015) for a detailed discussion of the MWA’s receivers and correlator, respectively. The instrument is designed to achieve a diverse set of science goals (Bowman et al. 2013) including a first detection of the 21 cm power spectrum during the EoR, detecting and monitoring transients, pulsars (Tremblay et al. 2015), solar and heliospheric science (Tingay et al. 2013b), and a low-frequency survey of the sky below DEC = +25° (Wayth et al. 2015).

Observations of a field centred at R.A.(J2000) = $4^h0^m0^s$ and decl.(J2000) = $-30^\circ0'0''$ were carried out for 4.13 h over two nights on 2013 September 5 and 6, with primary beams of the 128 MWA antenna elements (‘tiles’) formed at five different altitude/azimuth pointings each night to track the field. After flagging for RFI and anomalous behaviour that we will discuss in detail in Section 3.1, our total observation time for our power spectrum upper limit is 3.08 h. In Fig. 1 we show the relative integration time on the

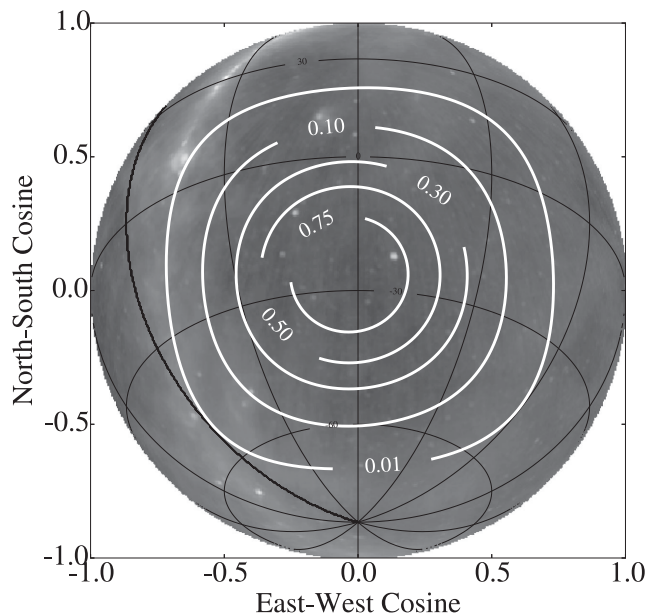


Figure 1. A radio map at 408 MHz (Haslam et al. 1982) sin-projected over the region of the sky observed in this paper. Cyan through magenta contours indicate the total fraction of observation time weighted by our primary beam gain for our 3 h of observation at 83 MHz. Red contours indicate R.A.-decl. lines. Observation tracked the position (R.A.(J2000) = $4^h0^m0^s$, decl.(J2000) = $-30^\circ0'0''$) on a region of the sky with relatively little galactic contamination and dominated by the resolved sources Fornax A and Pictor A. The galactic anticenter and bright diffuse sources, such as the Gum Nebula, are below 1 per cent bore-site gain.

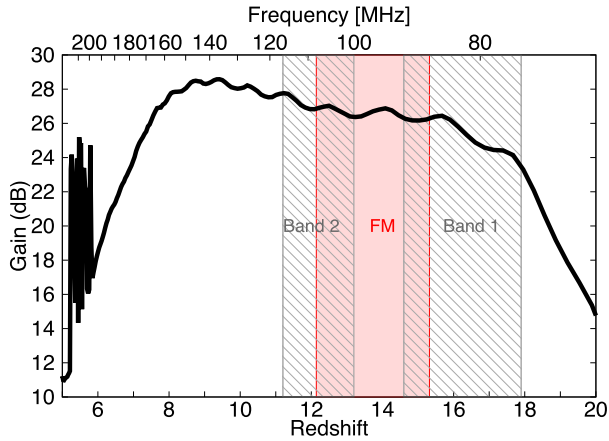


Figure 2. The autocorrelation spectrum of a single tile, showing the MWA bandpass, is plotted here (solid black line) along with the frequency ranges over which our data was taken (grey striped rectangles). Observations were performed simultaneously on two non-contiguous bands located on either side of the 88–108 MHz FM band (red shaded region) to both assess conditions within the FM band and to preserve some usable bandwidth should it have proven overly contaminated by RFI.

sky weighted by the primary beam over all observations. We observed with 40 kHz spectral resolution simultaneously over two contiguous bands; a 16.64 MHz interval between 75.52 MHz and 90.88 MHz (Band 1) and a 14.08 MHz band between 98.84 MHz and 112.64 MHz (Band 2). Both of these sub-bands overlap with the FM band (88–108 MHz). In Fig. 2 we show our observed bands superimposed on the autocorrelation spectrum of a single MWA tile. Observations in Band 1 took place right on the edge of the analogue cutoff of the MWA, making its shape relatively complicated to model in calibration (our calibration is direction independent and described in Section 2.2 and Section 3.4). Band 2, while in a flatter region of the bandpass, has a larger overlap with the FM band. Observations were divided into 112 s snapshot with data averaged into 0.5 s integration intervals by the correlator.

2.2 RFI flagging and initial calibration

The data were first flagged for RFI contamination. An optimized version of `aoflagger` (Offringa, van de Gronde & Roerdink 2012), called `cotter` (Offringa et al. 2015), was run on each snapshot with automatic RFI identification performed only on the visibility cross-correlations. Additional flags were applied to the centre and 40 kHz edges of each 1.28 MHz resolution coarse spectral channel to remove digital artefacts arising from the two stage channelization scheme used in the MWA. After flagging, visibilities were averaged in time from 0.5 to 2 s and in frequency from 40 to 80 kHz. The averaged visibilities were then converted to Common Astronomy Software Application (CASA) measurement sets (McMullin et al. 2007) which served as the inputs to all ensuing steps in our pipeline. The percentage of all data flagged by `cotter` was approximately 0.75 per cent for Band 1 and 2 per cent for Band 2. As described in Section 3.1, we also implement additional (and highly conservative) flagging of observations by inspecting autocorrelations which leads us to discard 25 per cent of our data. Our initial calibration was divided into three steps: A preliminary complex antenna gain solution using an approximate sky model, one iteration of self calibration, and polynomial fitting to reduce noise and limit fine frequency scale systematics that might arise from our incomplete sky model.

For the first step, our sky model combined a list of point sources with images of the two bright resolved sources in our field: Fornax A and Pictor A. For the point source model we included the 200 brightest sources at our frequencies based on extrapolated power-law fits to data from the Coolgoora survey (Slee 1995), the Molongolo Reference Catalogue (Large et al. 1981) and more recent measurements by PAPER at 145 MHz (Jacobs et al. 2011). Fornax A is the brightest extended source in our field and is highly resolved, so we modelled it with a VLA image taken by Fomalont et al. (1989) at 1.4 GHz, scaled to match the flux density and spectral index measured by Bernardi et al. (2013) at 180 MHz, and extrapolated to our band with a spectral index of -0.88 . For Pictor A, we used a VLA image at 333 MHz by Perley, Roser & Meisenheimer (1997) and extrapolated to our band with a spectral index of -0.71 . The model components for our initial calibration extended down to an apparent flux density of ≈ 5 Jy which is comparable to the flux uncertainties in the brightest sources in the initial catalog. Due to the high uncertainty in the sky at our frequencies, this model was updated by a round of self calibration which we describe shortly.

The CASA bandpass function was used to obtain a first set of best-fitting calibration gains, averaging over 32 fine channels for each solution. Since our starting model was uncertain at the 10 per cent level, an iteration of self calibration was performed by MFS imaging and deconvolving 10^4 components with `WSCLEAN` (Offringa et al. 2014). The CLEAN components were used as a model for a second run of `bandpass` where we relaxed channel averaging so that the complex gain for each 80 kHz fine channel was found independently. In Fig. 3 we show the fractional change in calibration amplitude for one antenna tile over time intervals in which beamformer settings were held constant (pointings). Over our two nights we find that variations are on the order of a few per cent and dominated by uniform amplitude jumps across the entire frequency range that are strongly correlated between all antennas. Observations of the autocorrelations do not show such behaviour within each pointing so these gain jumps must arise from the calibration routine itself. Possible sources of these jumps could be variations in the overall amplitudes in self-calibration which occurs if the cleaning step does not recover all of the flux density on large scales, unmodelled sources moving through the sidelobes of the primary beam, or the result of a varying sky in the presence of beam modelling errors. Since these features do not introduce fine frequency structure, we do not think they are an impediment to the power spectrum analysis in this work.

We attempted to apply calibration solutions in which each frequency channel was allowed to vary freely but we found significant power was introduced on fine scales, contaminating the EoR window. This could be due to unmodelled sources adding spectral structure from long to short baselines or insufficient SNR on the solutions themselves. There is also possibility for signal loss given the large number of degrees of freedom. The degree to which calibration can remove signal and how unmodelled sources with enhanced spectral structure at long baselines can be mixed into short baselines are open questions that are beyond the scope of this analysis and are being investigated by Barry et al. (2016), Trott & Wayth (2016) and Ewall-Wice et al. (2016b). With these concerns in mind, we erred on the side of caution and fit each j th antenna gain with the product of three smooth functions in frequency, f ,

$$g_j(f) = P_j(f)R_{j,7m}(f)B(f), \quad (2)$$

where $P_j(f)$ is a third-order polynomial in amplitude and first-order polynomial in phase; $B(f)$ is a median bandpass that accounts for the course band shape determined by taking the median of

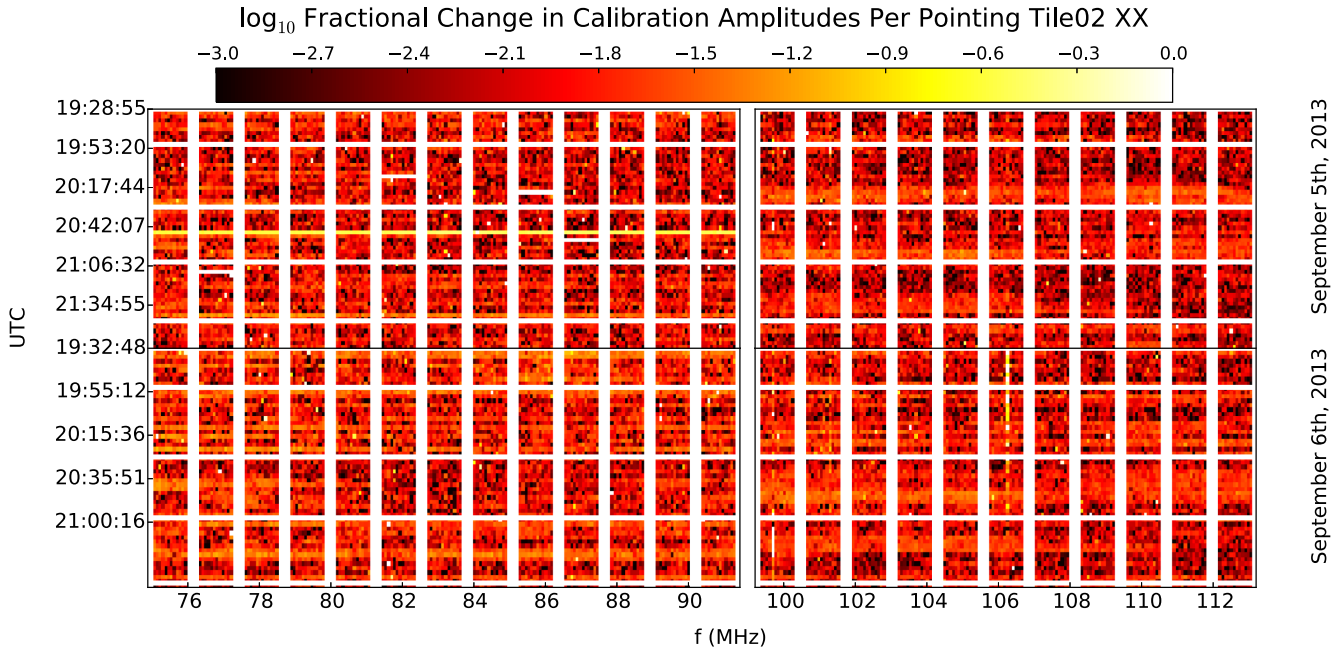


Figure 3. A false-colour plot of the fractional change in our calibration amplitudes over each pointing (in which beamformer delay settings are fixed). Pointing changes are marked by the solid white horizontal lines. We see that the calibration amplitudes vary within a pointing on the order of several percent with little systematic variation. There are several coarse channel scale jumps on September 5 that correspond to observations in which the number of sources identified within a snapshot image were reduced (see Fig. 9). We found that these jumps corresponded to excess flagging from `cotter`, indicating high levels of RFI or other bad data, and dropped them from our analysis. Vertical lines from RFI are visible at 98 MHz and 107 MHz, especially on September 6.

calibration amplitudes across all tiles and polarizations; and $R_{j, 7m}(f)$ is a reflection function that accounts for standing wave ripples in the bandpass arising from impedance mismatches in the 7-metre low noise amplifier (LNA) to beamformer cable connections. It is straightforward to show (see Appendix A) that a cable with length $L(j)$ introduces a multiplicative reflection term to the overall gain,

$$R_j(f) = \frac{1}{1 - r_j e^{i(2\pi f \tau_{L(j)} + \phi_j)}}, \quad (3)$$

where $r_j e^{i\phi_j}$ is a complex coefficient that is a function of the impedance of the cable and its connections at either end and $\tau_{L(j)}$ is the time for a signal to travel from one end of the cable and back. We note that this multiplicative term is derived from the infinite sum of reflections occurring from each n th round trip of the reflected wave and hence accounts for higher order reflections, not just the first round trip contribution.

When we formed power spectra from visibilities calibrated by this initial method, we found that our beamformer to receiver cables introduced spectral structure at the $\lesssim 1$ per cent level into our instrumental bandpass which were not removed by this smooth model (we return to this issue in Section 3.3). The effect of this spectral structure on the MFS images, used to measure ionospheric refraction in Section 3.2, was negligible and we therefore employed this calibration for their production. A more refined calibration procedure, which we describe in Section 3.4, was used for the data cubes and our power spectra.

2.3 MFS imaging and flux scale corrections

To form MFS images, we averaged the antenna phases over each night and held them constant for every snapshot calibration solution to average over time-variability introduced by the ionosphere. This was done in order to ensure that short time-scale snapshot-to-

snapshot variations were not due to time variations in the calibration solutions caused by the ionosphere. A multifrequency synthesis image was created from each snapshot, band, and polarization. From each XX and YY polarization snapshot which we call I_{XX} and I_{YY} respectively, we created a Stokes I snapshot corrected for the primary beam using an analytic dipole model of the MWA primary beam (which we denote as B_{XX} and B_{YY}).

$$I_I(\theta, \phi) = \frac{I_{XX}(\theta, \phi)B_{XX}(\theta, \phi) + I_{YY}(\theta, \phi)B_{YY}(\theta, \phi)}{B_{XX}^2(\theta, \phi) + B_{YY}^2(\theta, \phi)}. \quad (4)$$

Sources were identified using the `Aegean` source finder (Hancock et al. 2012). An overall flux scale for each stokes I snapshot was set following the technique, described in Jacobs et al. (2013), in which the flux scale for all sources is simultaneously fit to catalog flux densities using a Markov chain Monte Carlo method. We used the ten highest signal-to-noise point sources in each field and catalog flux densities interpolated between 74 MHz measurements from the Very Large Array Low-Frequency Sky Survey Redux (Lane et al. 2014) and 80 and 160 MHz measurements from the Culgoora catalog (Slee 1995). Since a flux scale error has no frequency dependence and the errors themselves evolve slowly in time near the centre of the primary beam, we do not think that such mismodelling will result in frequency-dependent errors. The dominant uncertainty in our flux scale is the systematic uncertainty in the model source fluxes themselves which are on the order of ≈ 20 per cent (Jacobs et al. 2013) while uncertainties in the beam model contribute at the several percent level (Neben et al. 2015). On September 6, we observed systematically smaller source counts than on September 5 (Fig. 9). We attribute this difference to greater ionospheric turbulence on September 6; and discuss this result further in Section 3.2.

In Fig. 4 we show a deep, primary beam corrected, integration of a portion of our field formed from an MFS image of Band 1. Known

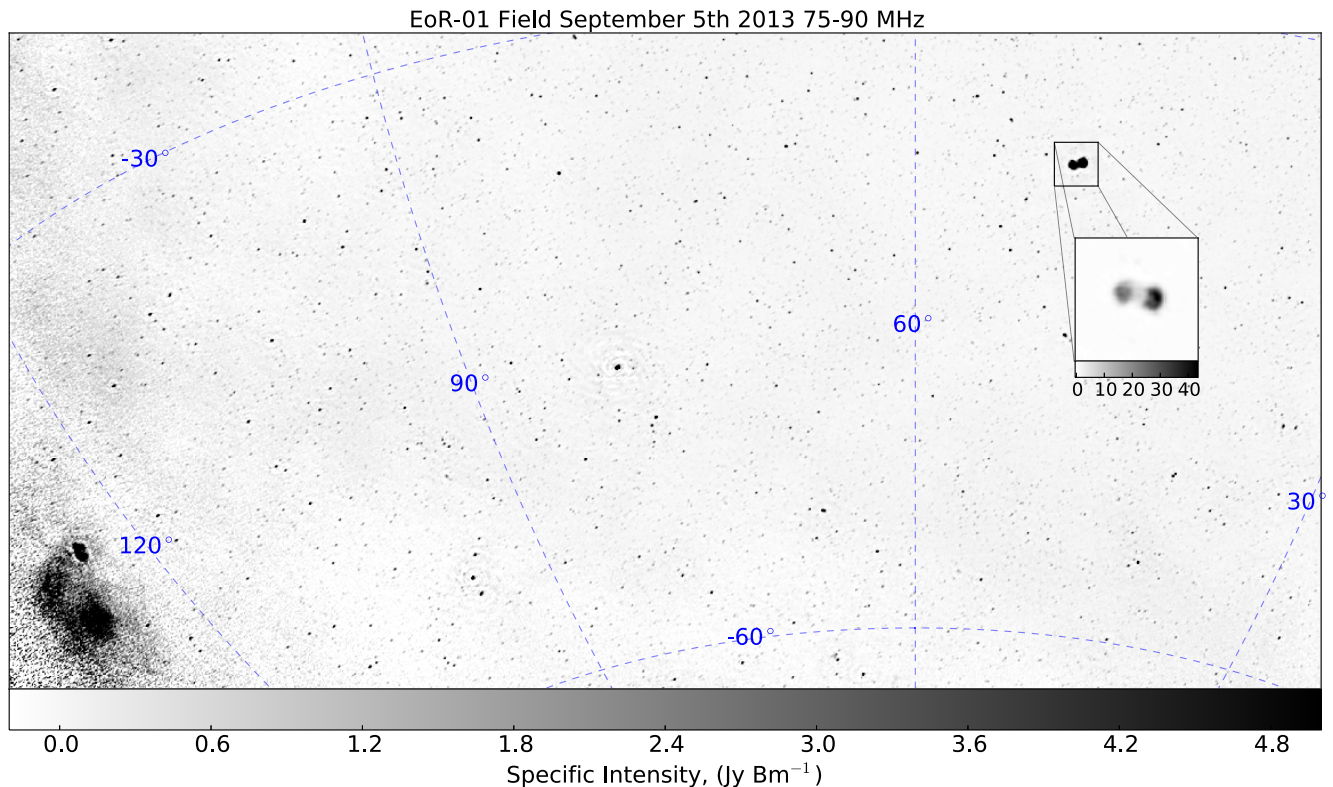


Figure 4. A deep image of the MWA ‘EoR-01’ field centred at (R.A.(J2000) = $4^h 0^m 0^s$ and decl.(J2000) = $-30^\circ 0' 0''$), derived by stacking restored multifrequency synthesis Stokes XX and YY images produced by `WSCLEAN` on Band 1. The dominant source in our field is Fornax A (detailed in the inset) whose structure is well recovered in imaging. Pictor A is also present at the centre of the image (at ~ 30 per cent beam) along with the diffuse Puppis and Vela supernova remnants on the left.

sources are well reproduced. The diffuse structures of the Vela and Puppis supernova remnants are clearly visible along with the fine scale structure of Fornax A.

Similar to previous upper limits on the power spectrum (Dillon et al. 2014; Parsons et al. 2014; Ali et al. 2015; Dillon et al. 2015b; Trott et al. 2016), our analysis does not consider the cross polarization products from the interferometer, which would require an additional calibration step to solve for the arbitrary phase difference between the X and Y polarized arrays (Cotton 2012; Moore et al. 2015).

2.4 Data cubes for power spectrum analysis

The inputs for our power spectrum analysis are multifrequency data cubes further calibrated with autocorrelation data from each tile. We describe the autocorrelation calibration technique in Section 3.4 where we address systematic errors revealed by our first look at data cubes and power spectra. In this section we describe our procedure for building the data cubes. While the MWA’s long baselines, extending out to 2864 m, are useful for gauging ionospheric conditions and other systematics, the uv plane is completely filled out to $\lesssim 160$ m after ≈ 3 h of rotation synthesis. Since forming image power spectra using data with incomplete uv coverage leads to unwanted mode-mixing and spectral artefacts (Hazelton et al. 2013), we threw away the sparse regions of the uv plane and reprocessed the data for our power spectrum analysis at much lower angular resolution than our MFS images. For each 112 s snapshot, we divided the data into even and odd time step visibility sets formed from every other 2 s integration step, which we later cross multiply

to form power spectra without noise bias (Dillon et al. 2014). A naturally weighted, dirty snapshot cube was produced for each set with 80 kHz spectral resolution, 80 pixels on a side and $1^\circ 0' 75''$ resolution for Band 1 (2). To form a power spectrum we will have to uniformly weight the sum of these cubes, hence the array point spread function (PSF), $\mathbf{s}_{j, \text{even/odd}}$, which is the 2 d Fourier transform of the number of samples within each uv cell; and the analytic primary beam matrix, \mathbf{B}_j , were also saved for each snapshot, polarization and interleaved time step. The power spectrum formalism of (Dillon, Liu & Tegmark 2013) assumes the flat sky approximation and requires a sufficiently small field of view to be computationally tractable. Hence, before creating a uniformly weighted data cube by stacking the naturally weighted images in the uv plane and dividing by their cumulative sampling function, we cropped each snapshot from 80×80 to 24×24 pixels.

We created even and odd, uniformly weighted temperature cubes by summing all naturally weighted snapshots across both the XX and YY polarizations and dividing by the sum of their cumulative sampling function in the uv plane,

$$\hat{\mathbf{x}}_{\text{even/odd}} = \frac{\lambda^2}{2k_b \Omega_{\text{pix}}} \mathbf{F}_2^{-1} \left[\frac{\sum_j \mathbf{F}_2 \mathbf{B}_j \hat{\mathbf{x}}_{j, \text{even/odd}}}{\sum_j \mathbf{F}_2 \mathbf{B}_j^2 \mathbf{s}_j} \right], \quad (5)$$

where the division of the two sums is to be understood as element-wise division. \mathbf{F}_2 denotes the two dimensional Fourier transform matrix. Indexing by angle cosines, ℓ, m , their duals, u and v , and frequency, f , we may write \mathbf{F}_2 and its inverse as,

$$[\mathbf{F}_2]_{uvf' \ell mf} = \Omega_{\text{pix}} e^{-2\pi i(\ell u + m v)} \delta_{f' f} \quad (6)$$

$$[\mathbf{F}_2^{-1}]_{\ell m f uv f'} = \frac{1}{N_\ell N_m} \Omega_{\text{pix}} e^{2\pi i(\ell u + m v)} \delta_{ff'}, \quad (7)$$

where Ω_{pix} is the solid angle of each pixel. Note that we are dividing by the convolution of the sampling function with the beam squared in the uv plane so we do not have to worry about dividing by the beam nulls in image space. The prefactor at the front of equation (5) converts from brightness to temperature units.

By dividing by the point spread function in the uv plane, equation (5) is essentially the application of uniform weighting to our data with some additional factors of the primary beam which warrant explanation. An additional factor of the primary beam was included in the sum to upweight regions of the field to which the MWA has the greatest gain, and is equivalent to optimal mapmaking (Tegmark 1997b; Dillon et al. 2015a). Other 21 cm pipelines, notably Fast Holographic Deconvolution (Sullivan et al. 2012) perform a similar upweighting by gridding visibilities with the primary beam while the fringe rate filtering procedure (Parsons et al. 2016) weights without gridding or imaging at all.

The impact of multiplying by the two factors of the beam in equation (5) has the effect of convolving the true visibilities the uv space beam convolved with its complex conjugate. Since our uv cells are quiet large, this is well approximated by multiplication of the visibilities by the convolution of the uv beam with its complex conjugate. We therefore also include the factor \mathbf{B}^2 in the denominator to correctly normalize the data in the uv plane.

Our $\hat{\mathbf{x}}$ estimate of \mathbf{I} in equation (5) is similar to the Stokes-visibility \mathbf{I} approximated in previous power spectrum analyses (Dillon et al. 2014, 2015b; Parsons et al. 2014; Ali et al. 2015). Visibility stokes \mathbf{I} is susceptible to leakage from visibility stokes $\mathbf{Q} \equiv \frac{1}{2}(XX - YY)$ due to beam ellipticity which has the potential to introduce fine frequency structure, caused by Faraday rotation, into the EoR window (Jelić et al. 2010; Geil, Gaensler & Wyithe 2011; Moore et al. 2013; Jelić et al. 2014; Asad et al. 2015; Moore et al. 2015). We can obtain an upper limit on polarization leakage in our power spectrum estimate by considering the best upper limits to date of the polarized \mathbf{Q} , \mathbf{U} , and \mathbf{V} visibility power spectra that might leak into our \mathbf{I} visibilities, measured by Moore et al. (2015) over the large field of view of PAPER. In this analysis the authors place limits of $\approx 5 \times 10^4 \text{ mK}^2$ on \mathbf{Q} , \mathbf{U} , and \mathbf{V} at $\approx 120 \text{ MHz}$. Scaling by the frequency dependence of the sky, the polarization power spectrum levels at our frequencies should be below $\approx 5 \times 10^4 \text{ mK}^2 (80 \text{ MHz}/120 \text{ MHz})^5 \approx 3.7 \times 10^5 \text{ mK}^2$. The leakage from \mathbf{Q}/\mathbf{U} to \mathbf{I} is given by equations 15 and 16 in Moore et al. (2015), and is equal to the product of the polarized power spectrum and the ratio between the integral of the differences of the X and Y polarized beams squared and the integral of the sum of the polarized beams squared. Using a short dipole model of our primary beam, we find that this ratio for the MWA beam is $\approx 5 \times 10^{-3}$. We therefore estimate an upper limit on the stokes \mathbf{Q} , \mathbf{U} to \mathbf{I} leakage in our power spectra to be $\approx 5 \times 10^{-3} \times 3.7 \times 10^5 \text{ mK}^2 \approx 1.5 \times 10^3 \text{ mK}^2$. This is slightly larger than the EoX power spectrum which is anticipated to be several hundred mK^2 (Pritchard & Furlanetto 2007; Santos et al. 2008; Baek et al. 2010; Mesinger et al. 2013) so it is still possible that polarized leakage may limit a detection unless direction-dependent polarization corrections are applied. However, this number is an upper limit and the actual leakage is probably lower. As of now, the most sensitive limits on the EoR power spectrum formed from XX and YY visibilities (Ali et al. 2015) limit (\mathbf{Q} , \mathbf{U}) $\rightarrow \mathbf{I}$ leakage from the similarly elliptical PAPER beam to below the level of $\approx 500 \text{ mK}^2$ between $k_\perp \approx 0.2\text{--}0.5 \text{ h Mpc}^{-1}$ at 150 MHz while Asad et al. (2015) predict stokes polarized power

spectrum from observations of the 3C196 field at $\approx 142 \text{ MHz}$ to be at the level of only $10^2\text{--}10^3 \text{ mK}^2$ at $k_\perp \lesssim 0.1 \text{ h Mpc}^{-1}$ (see their fig. 12, panel a). Scaling this up by $(140 \text{ MHz}/80 \text{ MHz})^5 \approx 16$ to account for the increasing sky temperature and applying our ellipticity factor of 5×10^{-3} gives a polarization leakage of $\approx 8\text{--}80 \text{ mK}^2$ which is still several times smaller than the predicted amplitude of the EoX power spectrum.

In order to reduce artefacts from aliasing at the edges of the coarse channels due to the two-stage polyphase filter bank, we flagged 240 kHz at each side. Finally, uv cells with poor sampling were flagged since sampling and noise in these cells can change rapidly with frequency, leading to fine frequency artefacts (Hazelton et al. 2013).

To check our flux scale, we estimate the level of thermal noise and the system temperature from the difference of our even/odd interleaved cubes, $\hat{\mathbf{x}}_{\text{even}} - \hat{\mathbf{x}}_{\text{odd}}$. Since the PSF is virtually identical between 2 s time steps, each pixel of $\hat{\mathbf{x}}_{\text{even}} - \hat{\mathbf{x}}_{\text{odd}}$ in the (u, v, f) basis has zero mean and a variance of (Thompson, Moran & Swenson 1986)

$$\langle |[\mathbf{F}_2 \hat{\mathbf{x}}]_{uvf}|^2 \rangle = \frac{\lambda^4 T_{\text{sys}}^2(f)}{2A_e^2(f)t(u, v, f)df}, \quad (8)$$

where df is the fine channel frequency width, T_{sys} is the system temperature, and $t(u, v, f)$ is the total integration time in the uv cell equal to the sampling function in the uv plane multiplied by the 2 s integration time step dt ,

$$t(u, v, f) = \sum_j [\mathbf{F}_2 \mathbf{B}_j^2 \mathbf{s}_j]_{uvf} dt. \quad (9)$$

$A_e(f)$ is the effective area of the MWA tile at frequency f computed from an analytic dipole model. We may determine T_{sys} by taking the ratio of the variance across uv cells at each frequency in $\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1$ and the average across uv cells of our model variance predicted by equation (8) without the T_{sys} factor. We find that $T_{\text{sys}}(f) = 2091 \text{ K}$ (1139 K) at 83 (106) MHz with an error of ≈ 20 per cent which is dominated by the systematic uncertainty in the fluxes of the sources used to set our flux scale. Assuming a spectral index of -2.6 (Rogers & Bowman 2008; Fixsen et al. 2011), these values imply a system temperature of $T_{\text{sys}} \approx 470 \text{ K}$ at 150 MHz , consistent with what is found at higher frequencies in Dillon et al. (2015b).

In Fig. 5 we show the standard deviation across all uv cells at each frequency in our Band 1 difference cube along with the square root of the mean of our model variances at each frequency assuming $T_{\text{sys}} \propto f^{-2.6}$ and normalized to the centre channel. We see that they are in good agreement. We also find that the standard deviation across frequency in each uv cell is consistent with the square root of the mean across frequency of our model variances.

An interesting question is whether or not our determination of T_{sys} is contaminated by ionospheric scintillation noise. For baselines within the Fresnel radius, $r_F = \sqrt{\lambda h / (2\pi)}$, of an ionospheric plasma screen of height h (which is the case for the MWA core), V15 determine the coherence time for scintillation noise to be set by the length of time it takes for overhead plasma, travelling at velocity v , to cross the Fresnel radius, $2r_F/v$. For typical plasma velocities of $\approx 100\text{--}500 \text{ km s}^{-1}$ and $h \approx 600 \text{ km}$, consistent with measurements (Loi et al. 2015a), we obtain coherence times between 10 and 44 s at 83 MHz , which is significantly greater than the 2 s interleaving of our data cubes. Hence, ionospheric fluctuations are likely subtracted away in our differencing on 2 s intervals. For a much lower altitude of 100 km , the correlation time is still $\approx 4 \text{ s}$ at 83 MHz . Even if there was still some variation between the time slices, we can put an upper limit on the variation relative to thermal noise by

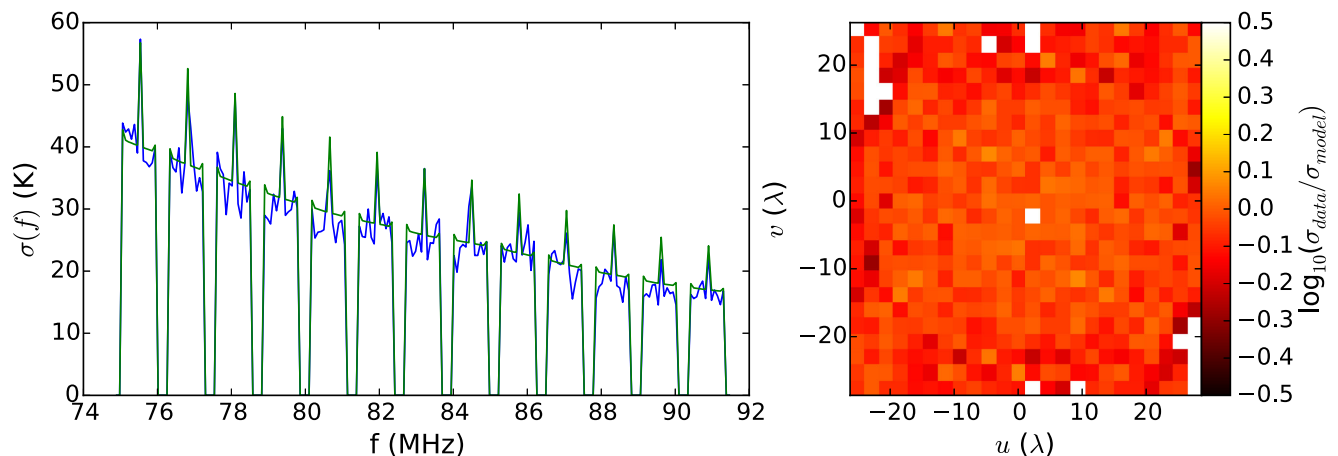


Figure 5. Left: the standard deviation over uv cells as a function of frequency for an even/odd time step difference cube after 3 h of integration. The blue line is derived from data, while the green line is a model with a system temperature of 470 K at 150 MHz and a spectral index of -2.6 . Spikes in the standard deviation are present at the centre of each coarse channel since the centre channel has one half of the data due to flagging the centre channel which is contaminated by a digital artefact before averaging from 40 to 80 kHz. Right: the ratio of variance taken over frequency in each uv cell and our variance model using the same system temperature as on the left. The ratio between our model and observed variance is close to unity across the uv plane. White cells indicate uv voxels that were flagged at all frequencies due to poor sampling. All data in this figure is from Band 1.

comparing the amplitude of scintillation noise we might expect given our primary beam and the parameters of the phase power spectrum of the ionospheric fluctuations that we determine in Section 3.2. We find that the level of scintillation noise on a single visibility in a 2 s snapshot (Appendix C) is only $\lesssim 2$ per cent the thermal noise level. Further suppression of the scintillations comes from the fact that we approximate T_{sys} at each frequency by taking the standard deviation across the uv plane within the Fresnel zone, in which the noise is expected to be coherent and would not contribute to such a standard deviation. The coherence in frequency of the ionospheric fluctuations (V15,V16) would also suppress their contribution to the standard deviation across frequency in each uv cell. For these reasons, we expect scintillation noise to have a very small contribution to our determination of T_{sys} at or below the 1 per cent level.

3 ADDRESSING THE CHALLENGES OF LOW FREQUENCY OBSERVING

A number of systematics associated with observing the EoR become dramatically more challenging as one moves to higher redshift. Because the Epoch of X-ray heating (EoX) spans the FM band, we expect enhanced RFI contamination. The ionosphere’s influence on electromagnetic wave propagation increases with wavelength, though its smooth evolution in frequency should cause its impact on source mis-subtraction and calibration to be contained within the wedge (Trott & Tingay 2015; Vedantham & Koopmans 2016, hereafter V16). Moving down in frequency, the larger primary beam extends foreground emission to higher delays and hence larger k_{\parallel} while the foregrounds increase rapidly in brightness temperature as $\approx f^{-2.6}$ (Rogers & Bowman 2008; Fixsen et al. 2011). Finally, spectral structure in our gains at fixed delays move down in k_{\parallel} at higher redshift and, due to the increased primary beam width, occupy a greater extent in k -space as well (Thyagarajan et al. 2015a). In this section we determine the impact (if any) of each these obstacles on our power spectrum analysis and describe our strategies for mitigating them. We deal with RFI, ionosphere, and spectral structure in Sections 3.1, 3.2, and 3.3/3.4, respectively.

3.1 Radio-frequency interference

As explained above, automated RFI detection and flagging was performed using `cotter` on the 0.5 s, 40 kHz resolution cross-correlations before time and frequency averaging. To illustrate the time-frequency structure of RFI contamination, we plot the fraction of visibilities flagged at each fine frequency channel and 112 s snapshot interval in Fig. 6. One can see that the majority of Band 1 is clear of RFI. Even within the region overlapping with the FM, events are sparse in time and frequency. In Band 2 we see significantly greater interference, especially in the two lowest coarse channels. There are clearly a greater number of events contained within the FM band however they only appear intermittently with the exception of a handful of 40 kHz fine-channels. The existence of intermittent FM signals, even in a radio quiet site such as the MRO is possibly due to signals from over the horizon, reflected off of the bottom of the ionosphere

It is possible for interference that is present for extended periods of time but weak enough to remain below the 112 s noise floor over which RFI is flagged to contaminate our data. RFI can also make it past flagging through calibration solutions which are derived from the autocorrelations on which `cotter` does not perform flagging (Section 3.4). When we first created integrated data cubes there were clear signs in Band 2 that some low level RFI contamination remained in the form of ripples in the power spectrum and spikes in the frequency domain of our gridded visibility cubes. We identified and discarded observations that appeared to contain increased flagging for a wide range of frequencies and completely flagged any channels that contained spikes in our final data cubes.

The lower two coarse channels in Band 2 were contaminated for a wide range of times (Fig. 6) and were thus excluded from our power spectrum cubes entirely. In addition, we flagged a total of 9, 80 kHz channels (7 per cent of our data) which appeared to be contaminated by RFI at contiguous intervals over a significant number of observations; 104.08, 104.48, 106.08, 106.24, 106.32, 107.44, 107.60, and 107.84 MHz. The rest of the FM band appears clean after routine `cotter` flagging is applied. After these channels are discarded, we see no evidence that our 3 h power spectrum results are limited by RFI (see the end of Section 4.1 for further discussion).

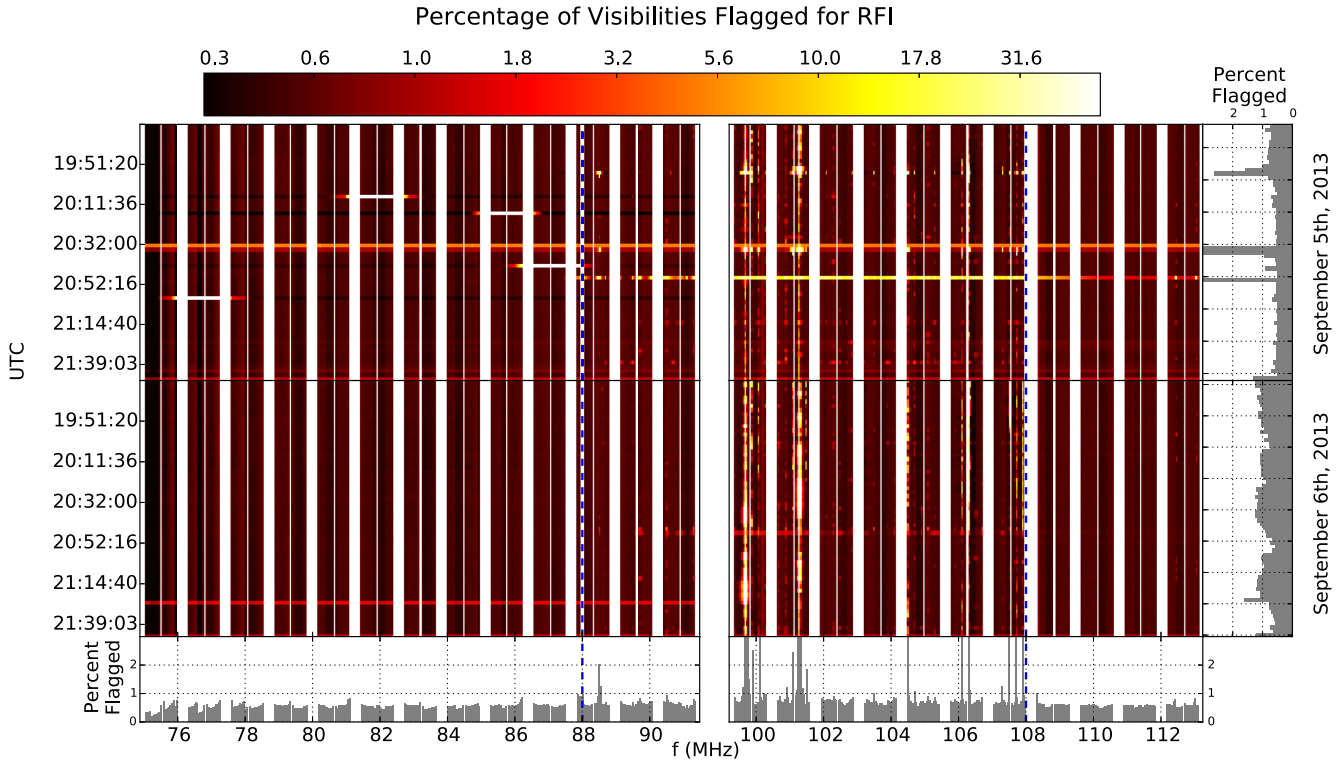


Figure 6. The percentage of visibilities flagged for RFI by `cotter` as a function of time and frequency. White regions indicate missing data including the coarse band edges and blue-dashed vertical lines indicate the edge of the FM band. While Band 1 remains predominantly clear with a few sparse events within the lower end of the FM band, contamination is significantly greater over Band 2. Even in the FM band, RFI events are either isolated in frequency or time, allowing us to flag them. A handful of observations in Band 1 on September 5 are missing entire coarse channels which we also discard. Bar plots on the bottom and right show the averages of the RFI flagging fraction over time and frequency, respectively.

Table 1. Bulk ionospheric and solar weather properties on the two nights of observing presented here. $\langle \text{TEC} \rangle$ indicates the mean total electron content over the entire night. We also show the diffractive scale calculated from all source separations on both nights which differ by a factor of 2.

Night	K_p	$\langle \text{TEC} \rangle$ (TECU)	F10.7 (sfu)	$\langle r_{\text{diff}} \rangle$ (km)
September 5	2	10.6	10.9	$10.4 \pm_{0.2}^{0.4}$
September 6	2	11.4	11.0	$5.20 \pm_{0.05}^{0.04}$

3.2 Ionospheric contamination

The refraction induced by gradients in the total electron content (TEC) of the ionosphere scales as λ^2 and is therefore expected to be more severe at our frequencies than those associated with the EoR. In this section we quantify the severity of ionospheric conditions over our observations by measuring the differential refraction of point sources relative to known catalog positions. We find that the ionospheric gradients change considerably over the duration of our observations, despite nearly constant and mild solar weather indicators (Table 1). However, when we form power spectra from data with nearly a factor of 2 difference in the observed gradients, we see no effect on the power spectrum within the EoR window (Fig. 22). The analysis presented here is meant to assess the level of refraction and, in Section 4.3, its impact on the power spectrum. Readers who are interested in a more detailed analysis of TEC gradients over the MRO and an interpretation of their physical origin or their impact on time domain astrophysics should consult Loi et al. (2015a,b).

Radiation passing through a plasma of electrons with spatial density $N_e(r)$ acquires a phase given by (Rybicki & Lightman 1979)

$$\phi \approx \frac{e^2}{cm_e f} \int N_e(r) dr, \quad (10)$$

where c is the speed of light, e is the electron charge, and m_e is the electron mass. If we assume the ionosphere is a flat screen of plasma at height h and that the TEC changes linearly over scales comparable to the separation between antenna pairs in an interferometer, $\mathbf{x}_i - \mathbf{x}_j = \mathbf{b}_{ij}$, the visibilities formed by cross multiplying and averaging the electric field measured by two antennas is given by (see V15 for a derivation)

$$V_{ij} = \langle E(\mathbf{x}_i) E^*(\mathbf{x}_j) \rangle \\ \approx \int d^2s I(\mathbf{s}) e^{i\nabla\phi(\mathbf{s}) \cdot \mathbf{b}_{ij}} e^{-2\pi i \mathbf{s} \cdot \mathbf{b}_{ij} f/c}, \quad (11)$$

This is the standard equation for radio interferometric visibilities (Thompson et al. 1986) where the angular positions \mathbf{s} of the sources with intensity $I(\mathbf{s})$ have been modified to be

$$\mathbf{s} \rightarrow \mathbf{s}' = \mathbf{s} + \frac{c}{2\pi f} \nabla\phi(\mathbf{s}). \quad (12)$$

where $\nabla\phi(\mathbf{s})$ is the gradient of the phase screen with respect to the E-W and N-S directions. Hence one can measure gradients in the TEC by observing offsets in the positions of point sources. Note that the gradients themselves are a function of position $\nabla\phi(\mathbf{s})$. While we focus on refractive effects as a proxy for ionospheric gradients, we also note that a significant number of the MWA baselines are within the Fresnel radius at these frequencies and experience significant

fluctuations in the source amplitudes as well, but we do not attempt a detailed analysis of these amplitude fluctuations in this work.

Cohen & Röttgering (2009) have observed ionospheric TEC gradients with the VLA using the differential refraction statistic and Helmboldt & Intema (2012) have measured the 2 d power spectrum of spatial and temporal fluctuations in TEC over the VLA. A similar power spectrum analysis exploiting the MWA's much larger field of view was recently carried out by Loi et al. (2015a,b). Because the ionosphere is curved and the derivation of equation (11) relies on the Fresnel approximation this model is only strictly valid for small fields of view so we only measure source shifts within 15 degrees of the phase centre. To obtain a global picture of ionospheric conditions, we turn to the differential refraction statistic employed in Cohen & Röttgering (2009) which we now briefly describe.

For an ensemble of source pairs with an angular separation of θ , the one dimensional differential refraction statistic, $D(\theta)$, is defined as

$$D(\theta) = \langle |\Delta\theta_1 - \Delta\theta_2|^2 \rangle, \quad (13)$$

where $\Delta\theta_{1/2}$ is the measured offset of source (1/2) from its known catalog position.

If $\Delta\theta = \Delta\theta_1 - \Delta\theta_2 = (\Delta\alpha, \Delta\delta)$ is a two-dimensional vector with each component distributed with standard deviation σ , then the probability density function of the amplitude square is given by an exponential distribution. We compute an estimate of $D(\theta)$, $\hat{D}(\theta)$, within each angular separation bin by fitting a histogram of the lower 80 per cent of source separations to an exponential distribution in order to eliminate potentially spurious outliers. $D(\theta)$, as will be explained, is directly related to the power spectrum of phase fluctuations whose properties we will determine below.

Each $\Delta\theta$ is the sum of an ionospheric offset and a noise term arising from random errors in determining the position of the source,

$$\Delta\theta_{1/2} = \Delta\mathbf{I}_{1/2} + \Delta\mathbf{N}_{1/2}, \quad (14)$$

where $\Delta\mathbf{I}$ is the contribution to position offset due to TEC gradients and $\Delta\mathbf{N}$ is the contribution from position errors. Expanding equation (13),

$$D(\theta) = \langle |\Delta\mathbf{I}_1|^2 \rangle + \langle |\Delta\mathbf{I}_2|^2 \rangle - 2\langle \Delta\mathbf{I}_1 \cdot \Delta\mathbf{I}_2 \rangle \\ + \langle |\Delta\mathbf{N}_1|^2 \rangle + \langle |\Delta\mathbf{N}_2|^2 \rangle - 2\langle \Delta\mathbf{N}_1 \cdot \Delta\mathbf{N}_2 \rangle. \quad (15)$$

For separations greater than several times the width of the synthesized beam, the background noise is uncorrelated. In our analysis we only consider separations that are greater than $\approx 8.4'$ while our synthesized beam has a diameter of 4.2 arcmin. In this regime, we can ignore the cross term in equation (15). Furthermore, it is roughly stationary over the centre of the primary beam lobe so that the noise terms add a θ independent offset to equation (15). If θ is small enough that both sources fall behind a single isoplanatic ionosphere patch but large enough such that the synthesized beams do not significantly overlap, $\langle |\Delta\mathbf{I}_1|^2 \rangle = \langle |\Delta\mathbf{I}_2|^2 \rangle = \langle \Delta\mathbf{I}_1 \cdot \Delta\mathbf{I}_2 \rangle$ so the ionospheric terms in equation (15) cancel out and we are left with only the noise bias terms. We may therefore determine the noise bias from smallest non-zero separation bin and subtract it. Our estimate of the square root of the structure function of the ionospheric fluctuations is

$$\hat{R}(\theta) = \sqrt{\hat{D}(\theta) - \hat{D}(\epsilon)}, \quad (16)$$

where ϵ is the median angular separation of our smallest bin which is 30 arcmin.

In Fig. 7 we show the differential refraction computed from all differential source pair separations over 30 min on 2013 September

5 and 6 for both of our observing bands. We see that the level of fluctuations recorded in both bands scales as λ^2 indicating that it indeed originates from ionospheric effects. Kassim et al. (1993) use a similar comparison to confirm ionospheric refraction as a source of variation in visibility phases. On September 6, the levels of refraction are approximately a factor of 2 greater than those observed on September 5, peaking at the end of the night.

We can relate our differential refraction measurements to the underlying power spectrum of phase fluctuations. In Appendix B, we derive the relationship

$$D(\theta) = \frac{2}{2\pi} \left(\frac{c}{2\pi f} \right)^2 \int dk k^3 (1 - J_0(kh\theta)) P(k) \quad (17)$$

where h is the height of the plasma screen, \mathbf{k} is the 2 d wave vector, and J_0 is the bessel function of the first kind. We parametrize $P(k)$ as a generalization of the form given in V15 which describes fluctuations that level out at some outer scale $r_0 = 2\pi/k_0$.

$$P(k) = \phi_0^2 \frac{4\pi(n-1)}{k_0^2} \left[\left(\frac{k}{k_0} \right)^2 + 1 \right]^{-n}. \quad (18)$$

Substituting this form of the power spectrum into equation (17), we obtain

$$D(\theta) = 4(n-1) \left(\frac{c}{2\pi f} \right)^2 \phi_0^2 k_0^2 F_n(k_0 h\theta), \quad (19)$$

where $F_n(x)$ is a dimensionless integral.

$$F_n(x) = \int dq q^3 (1 - J_0(qx)) [q^2 + 1]^{-n}. \quad (20)$$

In Fig. 8 we show several examples of $F_n(x)$, noting that it exhibits power-law behaviour for small values of x and levels off towards $x = 1$. Assuming a plasma height of $h = 600$ km and fitting our structure functions to equation (19), we obtain values for the power spectrum normalization, ϕ_0 , the spectral index, n , and the outer energy injection scale, r_0 . We show our fits to our 83 MHz band 1 as dashed black lines in Fig. 7. Fitted values for r_0 were on the order of several hundred kilometres, n ranged between 2.3 and 2.7, and ϕ_0 between 4 rad on the quietest times and 45 rad during the most severe refraction at the end of September 6. The reported n values are somewhat steeper than $n = 11/6 \approx 1.8$ for Kolmogorov turbulence. However, the spectral index of ionospheric fluctuations has been found to vary significantly (Rufenach 1972; Cohen & Röttgering 2009). We check the slopes of $F_n(x)$ for small x for our fitted indices and find that they lie within the slopes of the power-law fits derived from various time ranges by Cohen & Röttgering (2009). We note that there are small systematic departures from the smooth behaviour described by $F_n(x)$ in most of our intervals. These are possibly due to anisotropies in the phase fluctuation fields and departure from turbulent behaviour over the short 30-min time-scales due to transient phenomena such as travelling ionospheric disturbances.

Given our fitted values, we can also compute the diffractive scale of the ionospheric fluctuations r_{diff} which gives the scale at which the structure function of the phase field, ϕ reaches unity. Phase fluctuations that have large amplitudes and decorrelate rapidly with separation give a smaller diffractive scale, allowing it to serve as a single number indicator of the severity of the fluctuations. We determine r_{diff} for each 30-min interval by computing the structure function from each fitted power spectrum and numerically solving for r_{diff} . We obtain error bars by computing r_{diff} for 1000 instances of (r_0, n, ϕ_0) drawn from a multivariate Gaussian distribution whose covariance is the estimate of the fitted parameter covariances given

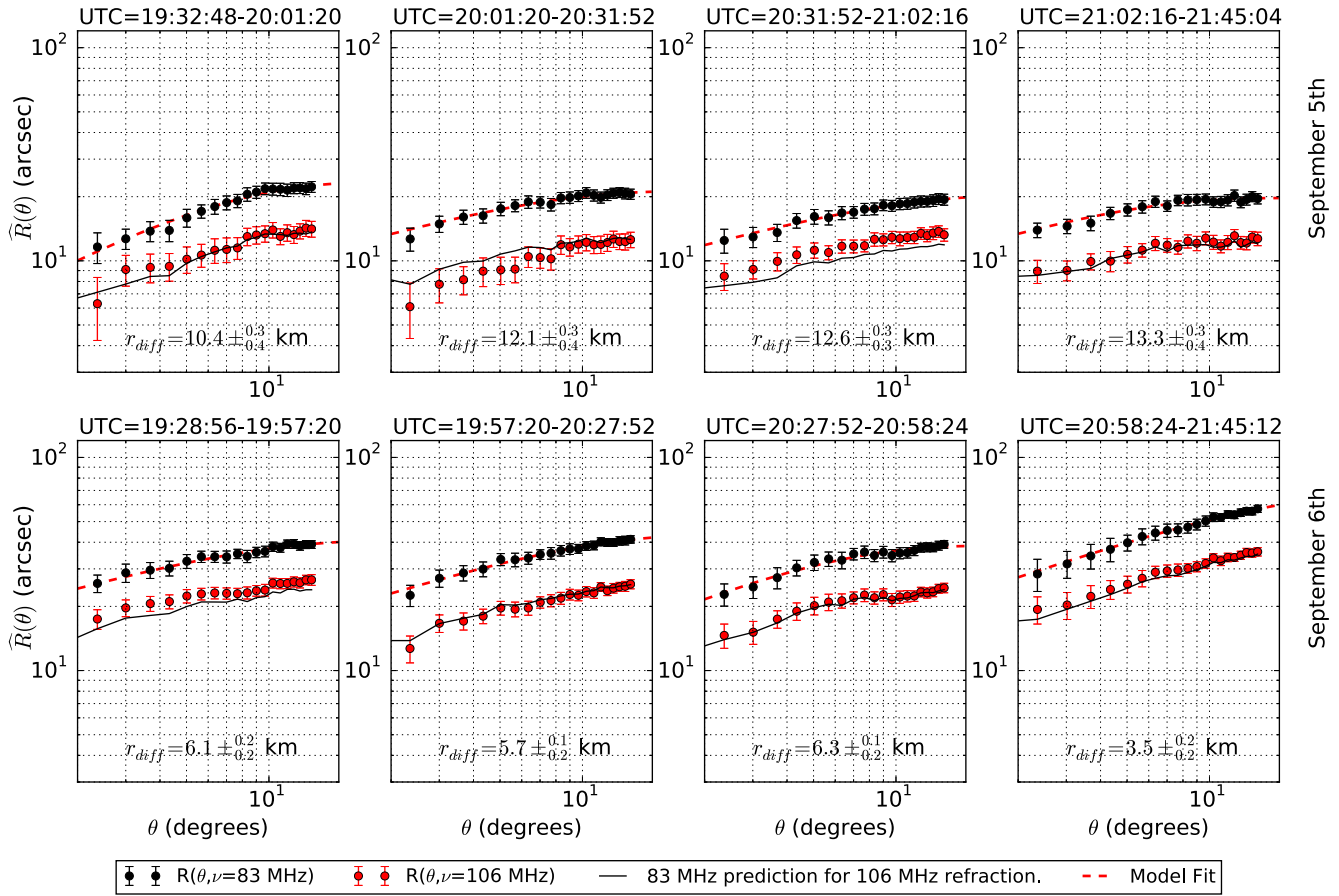


Figure 7. Differential refraction derived from source pairs within 30 min bins on 2013 September 5 (top row) and 2013 September 6 (bottom row). Band 1 (black points) scaled by the ratio of the band centre frequencies square (solid black line) very nicely predicts the differential refraction in Band 2 (red points), indicating that the refraction we are measuring here is indeed due to ionospheric fluctuations. The magnitude of ionospheric activity differs significantly between September 5 and 6 and peaks over the last observations taken on the 6. We also show fits to an isotropic power spectrum model of differential refraction at 83 MHz (dashed black line) and print the inferred diffractive scale.

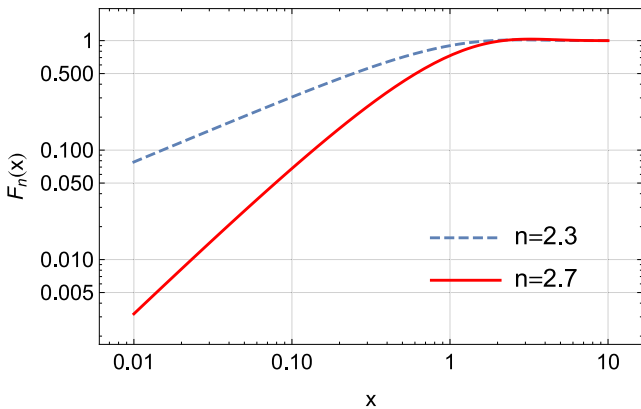


Figure 8. The dimensionless integral, $F_n(x)$ normalized to unity at $F_n(10)$, given in equation (20). For small values of x , $F_n(x)$ is well approximated by a power law, but flattens out towards $x = 1$. Hence, the structure function of observed source offsets levels out at the outer energy injection scale of the turbulence.

by the Levenberg–Marquardt method as implemented in `scipy`.¹ We use the 16 and 84 per cent percentile values of the resulting

distributions to obtain 1σ upper and lower bounds. On September 5, the median r_{diff} , at 83 MHz, ranged between ≈ 11 and 13 km while on September 6, it ranged between ≈ 4 and 6 km. We report these values and their errors in Fig. 7. We also show the values of r_{diff} derived from all source pairs on each night in table 1. The frequency scaling of the diffractive scale depends on the spectral slope of $P(k)$, n . The n -values in our analyses typically fall between 2.3 and 2.7 which yields a $r_{\text{diff}} \propto f$ scaling.² Hence, our measurements imply diffractive scales of ≈ 20 km and ≈ 10 km at 150 MHz on September 5 and 6, respectively. These values are within the range of typical scales (5–50 km) described in V15.

Changes in the refractive index over a source on time-scales shorter than the snapshot integration can cause source smearing in image space, resulting in a reduction in the observed peak brightness and reducing the number of source detections (Kassim et al. 2007). In Fig. 9 we compare the number of sources identified in each snapshot over September 5 and 6. On September 5, when the level of refraction is significantly lower, we observe that the number of sources increases as the field approaches transit, corresponding to the pointing at which the beam has maximal gain and then turns over. On September 6, the night in which significantly

² Kolmogorov turbulence, with a spectral index of $n = 11/6$ in the V15 parametrization scales in a very similar way; $r_{\text{diff}} \propto f/5 \approx 1$.

¹ <https://github.com/scipy/scipy>

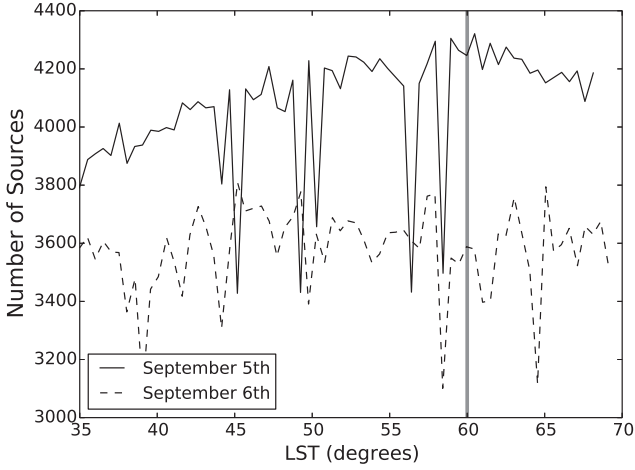


Figure 9. The number of sources identified in 112 s multifrequency synthesis images of Band 1 as a function of time over both nights of observing. On September 5, the source counts increase with primary beam gain, until transit (vertical grey line) before decreasing. On September 6, when more severe ionospheric refraction was observed, the source counts remain significantly lower. Fewer sources were identified in a handful of September 5 snapshots corresponding to observations in which calibration and flagging anomalies were observed (Figs 3 and 6). We exclude these snapshots from our analysis.

greater refraction was observed, the number of sources identified stays relatively constant and significantly lower than any of the source counts observed on September 5. There are also noticeable drops in the source counts on a handful of observations on September 5 which we found to correspond to flagging events in which an entire coarse band was eliminated with significant RFI detections on the edges (Fig. 6). We drop these snapshots from the rest of our analysis.

While refraction varied significantly over both nights and within each night, the bulk solar weather and geomagnetic conditions are nearly identical. In Table 1 we also list several bulk statistics such as K_p index, 10.7 cm flux (F10.7),³ and mean TEC at the MRO.⁴ The solar flux at 10.7 cm is an often used index of solar activity and is known to correlate with the emission of the UV photons responsible for generating the free electrons that impact ionospheric radio propagation (Yeh & Flaherty 1966; da Rosa et al. 1973; Titheridge 1973) with values ranging between 50 and 300 SFU (1 SFU = 10^{-22} W m⁻²). The K_p index (Bartels, Heck & Johnston 1939), quantifies the severity of geomagnetic activity and combines many local measurements of the maximal horizontal displacement of the Earth’s magnetic field. Values for K_p range from 0 to 9 with any values above 5 indicating a geomagnetic storm. We find that the bulk values are very similar between nights, indicating that ionospheric indicators derived from the observations themselves are a much better way of assessing data quality. Indeed, Helmboldt, Lane & Cotton (2012) also found that the levels of ionospheric turbulence and the incidence of travelling ionospheric disturbances did not appear to correlate strongly with bulk ionosphere and solar weather statistics. A systematic study of ionospheric conditions at the MRO site and correlations with bulk ionosphere statistics at higher frequencies is

³ K_p and 10.7 cm flux values were obtained from <http://spaceweather.com/archive.php>.

⁴ Values obtained by averaging global TEC maps downloaded from the MIT Haystack Madrigal data base (Rideout & Coster 2006).

currently underway (Loi et al., in preparation). Studies incorporating information from ancillary probes such as GPS stations are also being carried out (Arora et al. 2015).

Having established that one of our observing nights had twice the level of refraction than the other, we can compare power spectra from each night to gauge the impact of ionospheric fluctuations on the 21 cm power spectrum. In Section 4.3, we find that the two power spectra appear indistinguishable.

3.3 Instrumental spectral structure

At low frequencies, the combination of wider primary beams and intrinsically brighter foregrounds causes leakage into the EoR window to corrupt a wider range of delays which are related to cosmological Fourier modes. In addition, the cosmological modes occupied by a fixed delay depends on redshift, causing features that contaminate low-sensitivity regions of k -space at one redshift to contaminate scientifically important regions at another.

When we form a power spectrum from visibilities calibrated by the techniques described in Section 2.2, we are immediately confronted with detections of striped artefacts in the EoR window at discrete delays (see the top left panel of Fig. 20). These delays correspond to the round-trip travel times on the lengths of cable that connect the MWA’s receivers and beamformers. We can get a rough understanding of how miscalibrated cable reflections affect the power spectrum by considering the effect of reflections with delay, τ_j and complex amplitude \tilde{r}_j on the j th tile. To first order in $|\tilde{r}_{i/j}|$, the effect of uncalibrated cable reflection is to multiply a visibility by a reflection factor.

$$V_{ij} \rightarrow V_{ij} (1 + \tilde{r}_i e^{2\pi i \tau_i f} + \tilde{r}_j^* e^{-2\pi i \tau_j f} + \mathcal{O}(\tilde{r}^2)) \quad (21)$$

The power spectrum is formed, roughly, by taking a Fourier Transform of the gridded visibilities in frequency and squaring. The square of the Fourier transform of a visibility with uncorrected reflections becomes,

$$\begin{aligned} |\tilde{V}_{ij}(\tau)|^2 &\rightarrow |\tilde{V}_{ij}(\tau)|^2 \\ &+ 2 \operatorname{Re} [\tilde{r}_i \tilde{V}_{ij}(\tau) \tilde{V}_{ij}^*(\tau - \tau_i)] \\ &+ 2 \operatorname{Re} [\tilde{r}_j^* \tilde{V}_{ij}(\tau + \tau_j) \tilde{V}_{ij}^*(\tau)] \\ &+ 2 \operatorname{Re} [\tilde{r}_i \tilde{r}_j^* \tilde{V}_{ij}(\tau - \tau_i) \tilde{V}_{ij}^*(\tau + \tau_j)] \\ &+ 2 \operatorname{Re} [\tilde{r}_i \tilde{r}_j^* \tilde{V}_{ij}(\tau - \tau_i + \tau_j) \tilde{V}_{ij}^*(\tau)] \\ &+ 2 \operatorname{Re} [\tilde{r}_i^2 \tilde{V}_{ij}(\tau - 2\tau_i) \tilde{V}_{ij}^*(\tau)] \\ &+ 2 \operatorname{Re} [\tilde{r}_j^{*2} \tilde{V}_{ij}(\tau + 2\tau_j) \tilde{V}_{ij}^*(\tau)] \\ &+ |\tilde{r}_i|^2 |\tilde{V}_{ij}(\tau - \tau_i)|^2 + |\tilde{r}_j|^2 |\tilde{V}_{ij}(\tau + \tau_j)|^2 + \mathcal{O}(\tilde{r}^3). \end{aligned} \quad (22)$$

All terms in equation (22) involve the cross multiplications of $\tilde{V}_{ij}(\tau + \Delta\tau) \tilde{V}_{ij}^*(\tau + \Delta\tau')$ and a coefficient on the order of \tilde{r}^n . The cable delays on the MWA are all $\gtrsim 90$ m, corresponding to round trip delays significantly beyond the wedge for the short baseline lengths considered in our power spectrum analysis. Hence, if $|\tilde{r}|$ is greater than the ratio between the signal and foreground amplitudes, R_{fg} , terms with $\Delta\tau = \Delta\tau'$ dominate equation (22). As a consequence, the first and last lines in equation (22) dominate all others. The first ($\mathcal{O}(0)$) is the foregrounds in the absence of reflections and exceeds the signal by a factor of 10^{10} but is also contained within the wedge.

The last line contaminates $\tau = \tau_i, -\tau_j$ at the level of $\tilde{r}^2 \times 10^{10}$ the signal level. All other lines in equation (22) exceed the signal level by $\tilde{r}^2 10^5$ but also contaminate a greater range of delays: $\tau = \tau_i, -\tau_j, 2\tau_i, -2\tau_j, \tau_i - \tau_j$.

As we will discuss below, we find that $\tilde{r} \lesssim 10^{-2}$ (Fig. 9), hence our analysis is only sensitive to the first and last lines arising from first-order reflections. The fourth, fifth and sixth terms in equation (22), which couple second-order reflections and beats will be above the level of the cosmological signal twice the round trip delay times and their differences. However, if the $\lesssim 1$ per cent reflections can be corrected to be below the level of $\lesssim 10^{-3}$, these cross terms will only appear at the 10^{-3} level of the signal and not impede a detection (though they may introduce some bias).

Since the $\mathcal{O}(\tilde{r}^2)$ terms in the last line of equation (22) dominate the others outside of the wedge, the lowest order effect of a reflection is to multiply our foregrounds by the reflection coefficient squared and translate them outside of the low-delay region in which they are usually confined (the wedge) to the round-trip delay of our cable which for the MWA is outside of the EoR window. Unless $|\tilde{r}|$ can be brought below the ratio between the foregrounds and the signal itself ($|\tilde{r}_i| \lesssim 10^{-5}$), the modes within a wedge translated to τ_i will remain unusable. Uncorrected reflections at the 10^{-2} level will also contaminate higher order harmonics and the differences between the delays.

For \tilde{r} significantly larger than R_{fg} (as is the case here), delays corresponding to higher order reflections will also be contaminated. These higher order terms are well below the sensitivity of this analysis but they can still potentially pose an obstacle to the detection of the signal. Hence, it is worth commenting on the level that \tilde{r}^3 and \tilde{r}^4 terms contaminate our data. First we address \tilde{r}^3 . Since every contribution of order $\mathcal{O}(\tilde{r}^3)$ in equation (22) involves the product of an $\mathcal{O}(\tilde{r}^3)$ coefficient with delay transformed visibilities translated to two different delays ($\Delta\tau \neq \Delta\tau'$), these terms will contribute at the level of $\tilde{r}^3 \times 10^5$ the level of the foregrounds. Uncorrected third-order terms with $\tilde{r} \lesssim 10^{-2}$ will contaminate our data at 10^{-1} the level of the signal and we do not consider them a serious issue, especially if the reflections are corrected to the $\approx 10^{-3}$ level.

Writing down all \tilde{r}^4 order terms in equation (22) is straightforward but not terribly enlightening. We can obtain the leading contributions to the $\mathcal{O}(\tilde{r}^4)$ terms in equation (22) by ignoring the products of delay visibilities translated by different amounts, $\Delta\tau \neq \Delta\tau'$. The $\mathcal{O}(\tilde{r}^4)$ contributions with $\Delta\tau = \Delta\tau'$ are

$$\begin{aligned} & \dots + |\tilde{r}_i|^4 |\tilde{V}_{ij}(\tau - 2\tau_i)|^2 + |\tilde{r}_j|^4 |\tilde{V}_{ij}(\tau + 2\tau_j)|^2 \\ & + |\tilde{r}_i|^2 |\tilde{r}_j|^2 |V_{ij}(\tau - \tau_i + \tau_j)|^2 + \dots \end{aligned} \quad (23)$$

We conclude that the sub-percent reflections observed in our data will introduce $\mathcal{O}(\tilde{r}^4)$ terms to the power spectrum at $\tilde{r}^4 \times 10^{10}$ times the amplitude of the signal at twice the fundamental delays and their differences. Second-order reflections at the $\lesssim 1$ per cent level will therefore dominate the signal by two orders of magnitude, impeding a detection. Fortunately, the \tilde{r}^4 dependence of these second-order terms greatly amplifies even modest improvements in correcting the reflections. For example, if the reflections are brought to below the 0.1 per cent level, the $\mathcal{O}(\tilde{r}^4)$ terms will be brought to below 10^{-2} the level of the 21 cm signal. We are able to bring the reflections down to ≈ 0.002 so they are not a problem in our data.

What cosmological wave-vectors in our measurements are contaminated by reflections? In cosmological coordinates, a fixed de-

Table 2. There are N cables of each length (L) between the MWA receivers and beamformers with associated round-trip delay times (τ_r). In the $f_{30\lambda}$ column, we list the percentage of baselines within 30λ at 83 MHz (where the majority of the MWA's power spectrum sensitivity lies) that are formed from at least one tile with the given cable length. We also list the k_{\parallel} of each delay given by equation (24) for three different redshifts. Cable reflections that are significantly above the k_{\parallel} values where we expect to obtain maximum sensitivity to the power spectrum at EoR redshifts ($z \approx 7$) move into the maximum sensitivity region at EoX redshifts ($z \approx 16$). Higher order reflections will also contaminate multiples of and differences between the delays and k_{\parallel} values listed in this table (thought at a lower level).

L (m)	N	τ_r (μ s)	$f_{30\lambda}$ (per cent)	$k_{\parallel}(z=7)$ ($h \text{ Mpc}^{-1}$)	$k_{\parallel}(z=12)$ ($h \text{ Mpc}^{-1}$)	$k_{\parallel}(z=16)$ ($h \text{ Mpc}^{-1}$)
90	19	0.74	40	0.42	0.31	0.27
150	31	1.2	80	0.70	0.53	0.50
230	23	1.9	29	1.1	0.83	0.70
320	8	2.6	0.4	1.5	1.2	1.0
400	17	3.3	1.7	1.9	1.4	1.3
524	30	4.3	1.9	2.5	1.9	1.7

lay corresponds to a line-of-sight wavenumber of k_{\parallel} (in units of $h \text{ Mpc}^{-1}$) of

$$k_{\parallel} \approx \frac{2\pi H_0 f_{21} E(z)}{ch(1+z)^2} \tau, \quad (24)$$

(Morales & Hewitt 2004) where c is the speed of light, $H_0/h = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, f_{21} is the hyperfine emission frequency, and $E(z) = \sqrt{\Omega_M(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda}$. For constant τ , the k_{\parallel} centre of the translated delay wedge will decrease with increasing redshift.

In Table 2 we list the lengths of cable between the MWA receivers and beamformers along with their round-trip delays, and their corresponding k_{\parallel} at the centre redshift of our two bands along with a redshift typical of an EoR measurement. Because $P(k)$ decreases rapidly with increasing k , interferometers are expected to have the highest signal to noise at the smallest delay that is uncontaminated by foregrounds (280 ns, corresponding to $k_{\parallel} \approx 0.1\text{--}0.2 h \text{ Mpc}^{-1}$ at EoX to EoR frequencies (Pober et al. 2013)). Assuming that reflections can be corrected to be below the 10^{-3} level so that only the last line in equation (22) is above the signal, they should be benign as long as they remain at sufficiently large or small k_{\parallel} that they do not leak into the region of maximal sensitivity. Because standing waves translate the entire wedge up to their delay, reflections located at the edge of the wedge will result in excess supra-horizon emission while cable reflections outside of the wedge will contaminate a finite chunk of k_{\parallel} , not just the delay of the reflection itself.

We see in Table 2 that the minimum k_{\parallel} associated with a cable ripple on the MWA at $z = 7$ is $0.42 h \text{ Mpc}^{-1}$ which, even if we allow for a delay width of 280 ns (the approximate width of the wedge with a supra-horizon emission buffer at $k_{\perp} = 0$), is above the region of maximal sensitivity. However, at $z = 16$, the approximate centre redshift of our Band 1, this cable ripple lies at $0.27 h \text{ Mpc}^{-1}$, which can leak a significant amount of power into the sensitivity sweet spot due to its finite k_{\parallel} width. This effect is illustrated in Fig. 10. At $z = 7$, the smallest reflection delay on the MWA introduces foreground contamination down to $k_{\parallel} \approx 0.3 h \text{ Mpc}^{-1}$, leaving a small foreground-free window in which cosmological measurements may be performed. On the other hand, at $z = 16$, this window becomes smaller with the other reflections filling the EoR window up to $k_{\parallel} \sim 0.8 h \text{ Mpc}^{-1}$. If the corrected reflections are several times larger than 10^{-3} , higher order terms below the 90 m delay may be comparable to the level of the 21 cm signal (shown as light grey

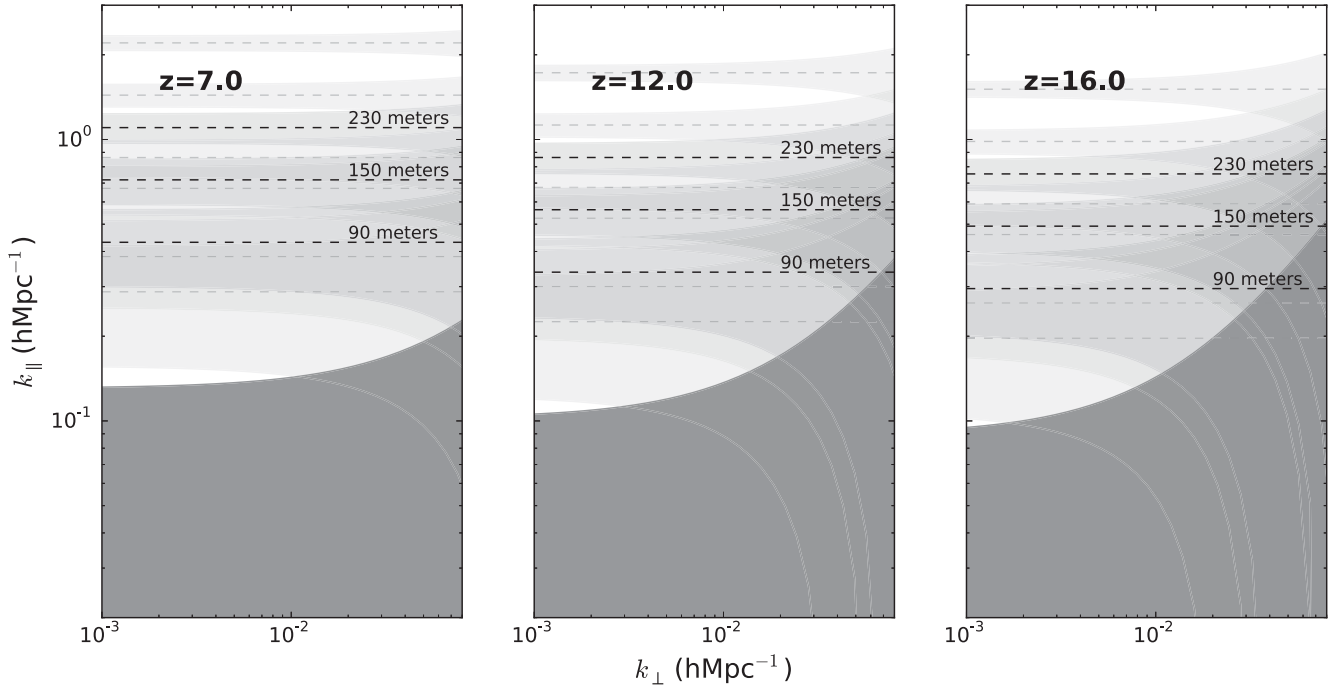


Figure 10. The regions of the EoR window contaminated by foregrounds due to uncalibrated cable reflections for several different redshifts. Dark grey regions denote contamination from first-order cable reflections assuming a wedge out to the first null of the primary beam plus the 0.15 h Mpc^{-1} at $z = 8.5$ buffer observed in Pober et al. (2013). Since the buffer is associated with the intrinsic spectral structure of foregrounds, it lives in delay space. Dark grey regions denote foreground contamination within the wedge which exists even without instrumental spectral structure. At $z = 7$, a representative EoR redshift, the contaminated regions remain at relatively high k_{\parallel} and have smaller widths due to the smaller primary beam and the scaling of k_{\perp} and k_{\parallel} with z . While regions exist between the first-order reflections that are somewhat wider at lower redshift, second-order reflections can still potentially contaminate nearly all of the EoR window in which interferometers are supposed to be sensitive (light grey shaded regions). Second-order reflections are below the sensitivity level of this study but will also pose an obstacle to longer integration unless calibrated out.

regions and centred on light-grey dashed lines). However, only the peak of this foreground power will be near the signal level and the broad wings caused by beam chromaticity and will be well below the signal.

This effect is purely geometrical in that while the mapping from instrumental delays to k -space varies, the number of measurements in (u, v, η) cells which are uncontaminated stays constant. However, as we go to higher redshift the increasing width of the primary beam increases foreground power at supra-horizon delays, effectively reducing the number of usable (u, v, η) cells. As mentioned above, the cosmological power spectrum decreases significantly with increasing k so the fact that smaller k are contaminated by cable reflections at higher redshifts hurts our sensitivity disproportionately. We also note that any partial reflections from kinks and bends within the cable itself can lead to contamination of additional delays below the round-trip travel time on the cable.

We make a first attempt to remove these reflections by fitting reflection functions (equation 3) to our self calibration solutions divided by the smooth fit in equation (2). Since our per snapshot calibration solutions are too noisy to obtain good fits for the beamformer-reflection, we fit these on calibration solutions that are averaged over each night of observing. We find that this method is of limited efficacy in removing the receiver to beamformer ripples (Fig. 20, top right).

Since, to avoid cosmological signal removal and spurious frequency structure due to mis-modelled foregrounds, we are attempting to model our bandpass with a small number of parameters we are unable to capture the full spectral structure of the cable reflections. As we see in Fig. 13 the reflection parameters are

frequency-dependent and we do not have a clear picture of their precise evolution. Calibration exploiting redundant baselines might be able to make headway on the problem since it is not sensitive to unmodelled signal except in for deriving an overall phase and amplitude scaling that averages over all baselines (Wieringa 1992; Liu et al. 2010; Zheng et al. 2014). The MWA’s baselines are designed for very low redundancy, making the technique unusable here. Future upgrades to the MWA are expected to include a significant number of redundant baselines (Tingay, private communication).

3.4 Calibration with autocorrelations

Confronted with the problem of reflections contaminating the power spectrum, we apply an alternative approach that uses tile autocorrelations to obtain calibration amplitude gains with sub-percent level accuracy.

What information is encoded in the autocorrelations? Let $I(\mathbf{s}, f)$ be the brightness distribution on the sky at frequency f and direction \mathbf{s} . Consider the sky signal entering the antenna and travelling through a signal chain in which the m th successive element applies a multiplicative complex gain $g_{j,m}$ and adds a zero mean noise component with variance $N_{j,m}^2$. Correlating the output at the M th gain element with itself to form an autocorrelation yields

$$V_{jj}(f) = |g_j|^2 \left[\sum_{m=0}^M \frac{N_{j,m}^2(f)}{\prod_{n=0}^{m-1} |g_{j,n}|^2} + \int d\Omega A_j(\mathbf{s}, f) I(\mathbf{s}, f) \right]. \quad (25)$$

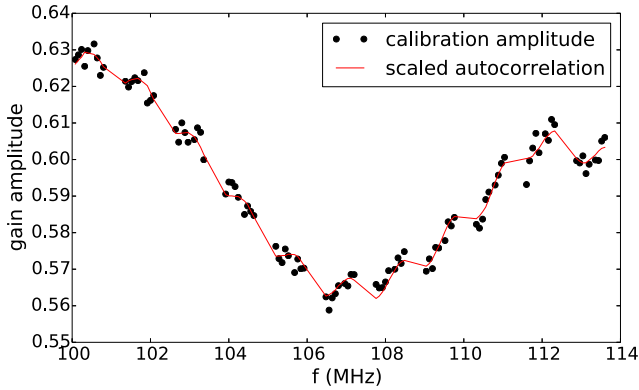


Figure 11. We show the amplitude of a calibration gain averaged over a 15 min pointing (black circles) along with the square root of our autocorrelations which have been scaled by a third-order polynomial and a single 7 m reflection to match the calibration solution (red line). After multiplying the autocorrelations by a smooth function, they are brought into good agreement with the calibration gains.

Here, $g_j = \Pi_{m=0}^M g_{j,m}$ is the net gain and $A_j(s, f)$ is the antenna beam. Using simulations of diffuse emission and an analytic model of the MWA primary beam, we find that $\int A_j I(s, f) d\Omega$ is fit at the 10^{-5} – 10^{-6} level by a third-order polynomial while $N_{j,m}^2(f)$, which is due to noise in analogue electronics should also vary smoothly with frequency. Hence $\sqrt{V_{jj}}$ is well approximated by the product of $|g_j|$ multiplied by a smooth factor which may be modelled by a low-order polynomial.

To remove this multiplicative factor, we need a model for the ratio of the square root of each autocorrelation to the amplitude of the calibration solution. We use the product of a third-order polynomial and a 7 m cable reflection (modelling the LNA-beamformer cables) and fit it to the ratio, averaged over 112 s intervals, using our noisy initial calibration solutions.

We then use the square root of the autocorrelation divided by this polynomial as our calibration amplitude. In Fig. 11 we demonstrate the validity of this technique by comparing the smoothly corrected autocorrelations for a single snapshot with a calibration amplitude that has been averaged over a single pointing and see that they are in very good agreement.

While the autocorrelations can be used for our amplitude calibration, there still remains the problem of adding the reflection ripple to phase calibration. We can use autocorrelations to predict this

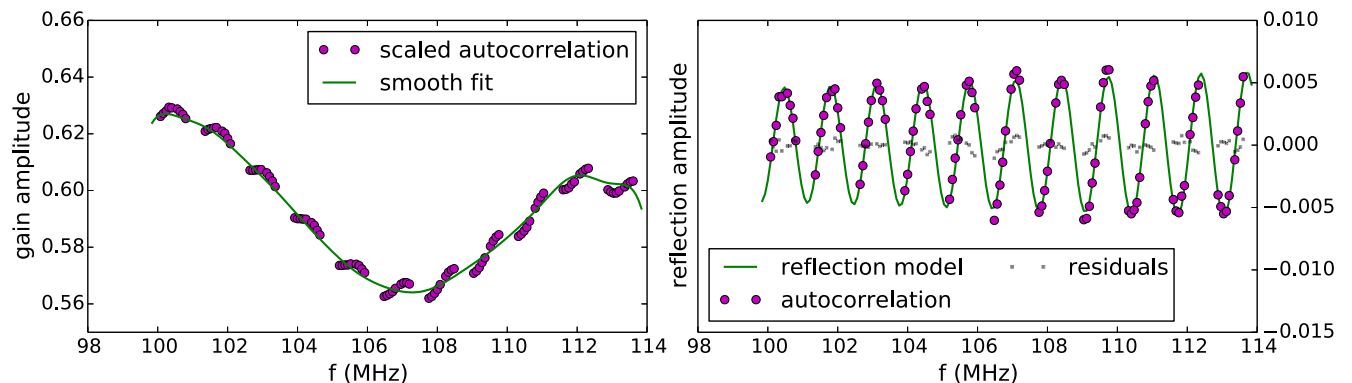


Figure 12. Left: in order to obtain reflection parameters, we divide our scaled autocorrelation (magenta circles) by a smooth function consisting of a third-order polynomial and large-scale reflections (green line). Right: we fit this ratio (magenta circles) to a reflection function (green line) and are left with ~ 10 per cent residuals (grey points).

ripple in the phases. We obtain its parameters following the same fitting procedure described in Section 2.2 except this time we fit the scaled autocorrelations, with a smooth polynomial divided out, to the amplitude of the reflection term in equation (3),

$$|R_j(f)| = \frac{1}{\sqrt{1 - 2r_j \cos(2\pi f \tau_{L(j)} + \phi_j) + r_j^2}}. \quad (26)$$

In Fig. 12 we illustrate the fitting procedure by showing the autocorrelations divided by the smooth fit along with the best-fitting model reflection. One can see that the residuals in the reflection fit tend to be on the order of 10 per cent hence there is some fine scale structure at the 10^{-3} level that we are still unable to model. Since our model includes the impact of higher order reflections, we think that these residuals arise from unmodelled frequency dependence in the reflection coefficients, sub-reflections in the cables, and digital artefacts present in the autocorrelations. Recalling our discussion in Section 3.3, 10^{-3} -level residuals will leave contamination in our power spectra at the level of 10^4 times the signal level due the $|\tilde{r}|^2$ terms in the last line of equation (22) but suppress all higher order reflections to below the signal level.

From $\tau_{L(j)}$, ϕ_j , and r_j , we add the reflection's additive contribution to the gain phase

$$\begin{aligned} \text{Arg}(g_j) &\rightarrow \text{Arg}(g'_j) \\ &= \text{Arg}(g_j) + \tan^{-1} \left[\frac{-r_j \sin(2\pi f \tau_{L(j)} + \phi_j)}{1 - r_j \cos(2\pi f \tau_{L(j)} + \phi_j)} \right]. \end{aligned} \quad (27)$$

Since tiles with 320, 400, and 524 m cables only contribute to ~ 4 per cent of our sensitive baselines, we discard them entirely. In Fig. 13 we show the distributions of the reflection amplitudes fitted from autocorrelations (averaged over the night of September 5) inferred for our 90, 150, and 230 m cables for both the high and low bands. One can see that the reflection coefficients are on the order of fractions of a percent and vary significantly from cable length to cable length. This is reasonable since the cable impedance, which determines the reflection amplitude, is a function of both its geometry and dielectric properties (with equal length cables likely formed from cable batches of similar dielectric properties). In addition, frequency evolution of the reflection amplitude is apparent by comparing the fit in Bands 1 and 2 implying that a single delay standing wave is not quite the correct model to use in our phases.

Autocorrelations are particularly susceptible to RFI and potential contamination due to cross talk and other artefacts. In Fig. 11, we saw that after flagging the channel edges, the spectral structure in

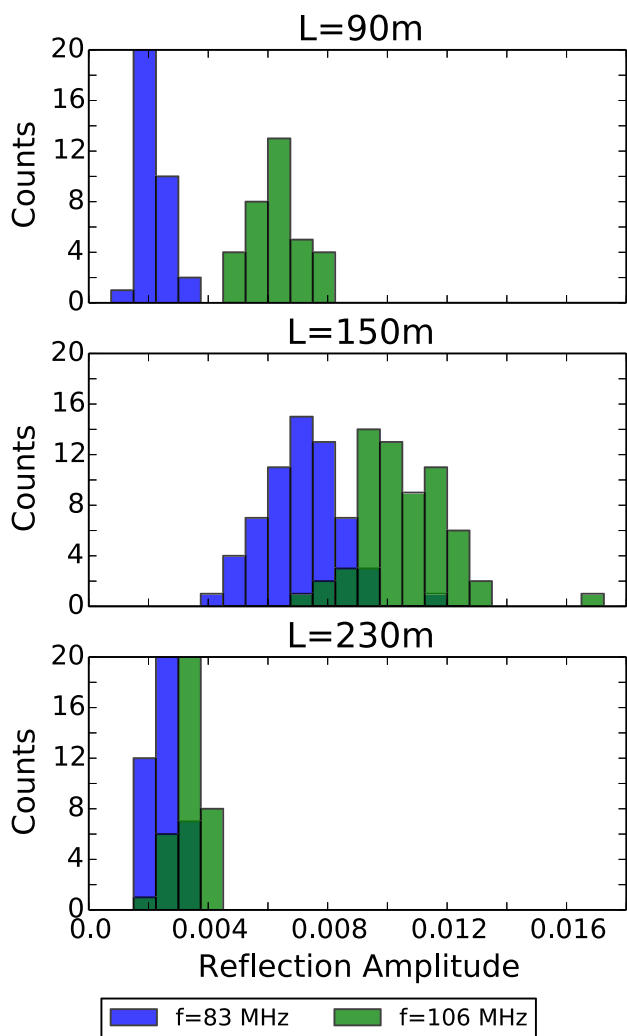


Figure 13. Histograms of fitted cable reflection amplitudes for Band 1 (blue) and Band 2 (green) obtained from fits to autocorrelations for three different cable lengths between MWA receivers and tiles. The reflection amplitudes range from 0.2 to 1 percent making them difficult to fit using the noisy self calibration solutions. Reflection amplitudes in Band 2 are systematically larger than Band 1 for all cable lengths, indicative of non-trivial frequency evolution in the reflection parameters.

the autocorrelations was consistent with our calibration solutions up to a smooth polynomial factor. In Fig. 14, we inspect for artefacts and RFI in a typical tile autocorrelation as a function of time. We see that RFI is present at similar times that were flagged in autocorrelation (Fig. 6). We also see that the time evolution of each autocorrelation is consistent between the two nights with rapid 10 per cent transitions occurring at ≈ 30 min intervals when the analogue beamformer settings are changed to track the sky. Ripples in frequency are also visible, corresponding to the structure in the standing wave reflections. It is difficult to pick out small artefacts in this dynamic spectrum view unless more large-scale smooth structure is fitted out, as is done in our calibration procedure. The residuals after this fitting give a better picture of what fine spectral features exist in the dynamic spectra of the autocorrelations which we discuss in the next section.

3.5 The time dependence of residual structure

We noted in Section 3.4 that our fits to reflections tended to have ~ 10 per cent residuals. Since we rely on these fits to predict the reflections in our gain phases, we expect residuals of these fits that are also present in the phases to contribute reflection power at a similar level. Our residuals could arise from thermal noise in the autocorrelations and calibration solutions. If this were the case we might expect them to average down with time. On the other hand, these residuals might also arise from mismodelling of the reflections themselves and would not average down with time. The result would be a systematic floor which can only be overcome by finding the correct model of the reflections or removing them from the signal path.

Plotting the fit residuals for two representative 90 m and 150 m tiles over the low band (Fig. 15), we find them to be at the $\sim 10^{-3}$ level. While there is some scatter in these residuals due to fitting noise, their frequency-dependent shape is relatively constant. As a consequence, the residuals average to a spectrum with frequency structure. These residuals are likely due to mismodelling of the frequency-dependent amplitude, phase, and period of the reflections but at a lower level may have some contributions from digital artefacts and cross-talk present in the autocorrelations. Because the component in these residuals that is sourced by reflections is also present in the phases which we are trying to model, there remains an uncorrected component to the gains that we are not calibrating out and does not average down with time.

While calibration with autocorrelations still appears to be limited by fine frequency artefacts arising from reflections, the high SNR of the reflections in the autocorrelations does offer significant improvement over fitting the reflections in the calibration solutions themselves. We provide a more quantitative look at the improvement achieved using calibration with auto-correlations in Section 4.2.

4 POWER SPECTRUM RESULTS

We can now present our power spectrum results and the first upper limits on the Epoch of X-ray heating power spectrum. We form cross power spectra of the even and odd timestep data cubes through the empirical covariance modelling technique developed in Dillon et al. (2015b, hereafter D15). In this procedure, the foreground residual model used in the inverse-covariance weighted quadratic power spectrum estimates and in the associated error statistics is derived from the data. It assumes that foreground residuals are correlated in frequency but uncorrelated in the uv plane and depend only on frequency and $|u|$. We refer the reader to D15 and its predecessors (Tegmark 1997a; Liu & Tegmark 2011; Dillon et al. 2013, 2014) for a thorough discussion of how this technique works. Along with estimates of the power spectrum amplitude, our pipeline outputs error bars and window functions which describe the mixing of the true power spectrum values into each estimate. We form 1 d power spectra by binning our 2 d power spectra using the optimal estimator formalism of Dillon et al. (2014) with the weights of all modes lying outside of the EoR window or with k_{\parallel} values showing consistent cable reflection contamination set to zero (D15).

First we will examine our two dimensional power spectra for Bands 1 and 2, derived from ≈ 15 MHz of bandwidth each, and comment on systematics (Section 4.1) and how well our calibration techniques mitigate them (Section 4.2). We finish by presenting our spherically binned 1 d power spectra, our most sensitive data product. We use our 1 d power spectra to compare foreground

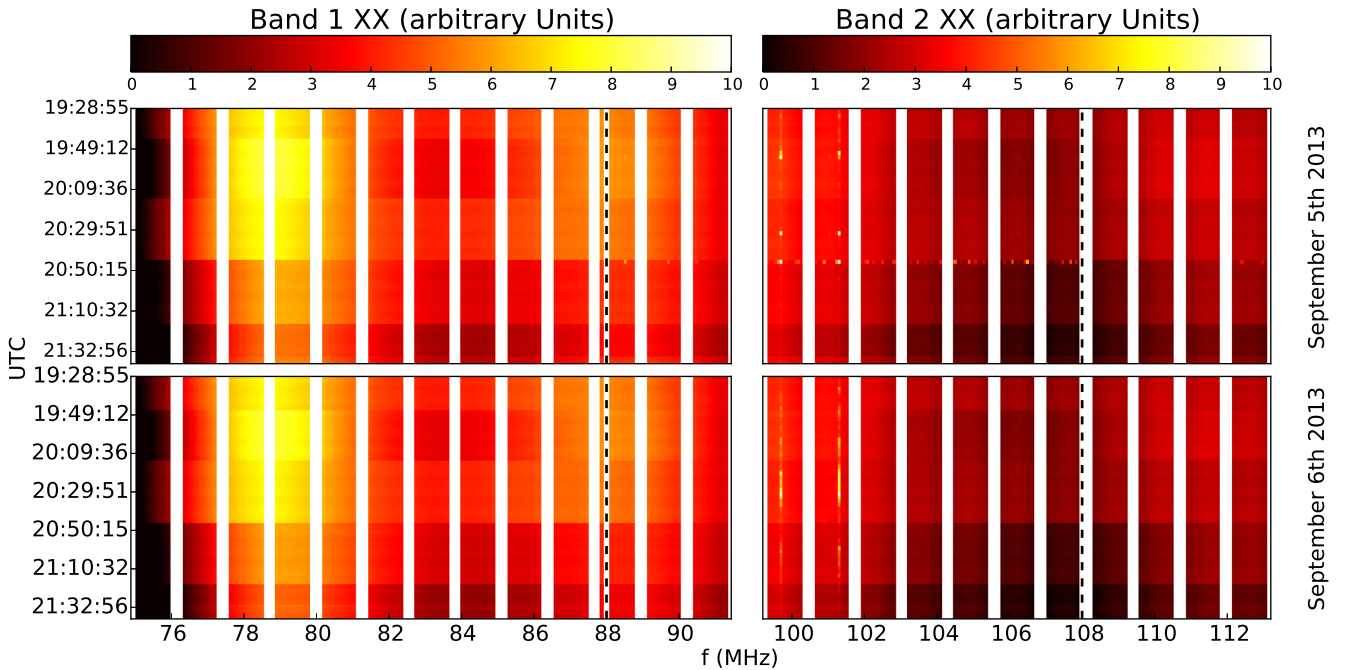


Figure 14. Dynamic spectra of the square root of a representative tile autocorrelation. Note the different colour bars for the two frequency bands since Band 1 evolves more steeply in frequency than Band 2. The autocorrelations exhibit repetitive structure in time from night to night with smooth time variations occurring as the sky rotates overhead and steep transitions occurring every ≈ 30 min due to changes in the analogue beamformer settings as the antennas track the sky. Limited RFI is plainly visible within the FM band, especially in Band 2, and the events are consistent with the flagging events identified by `cotter` shown in Fig. 6.

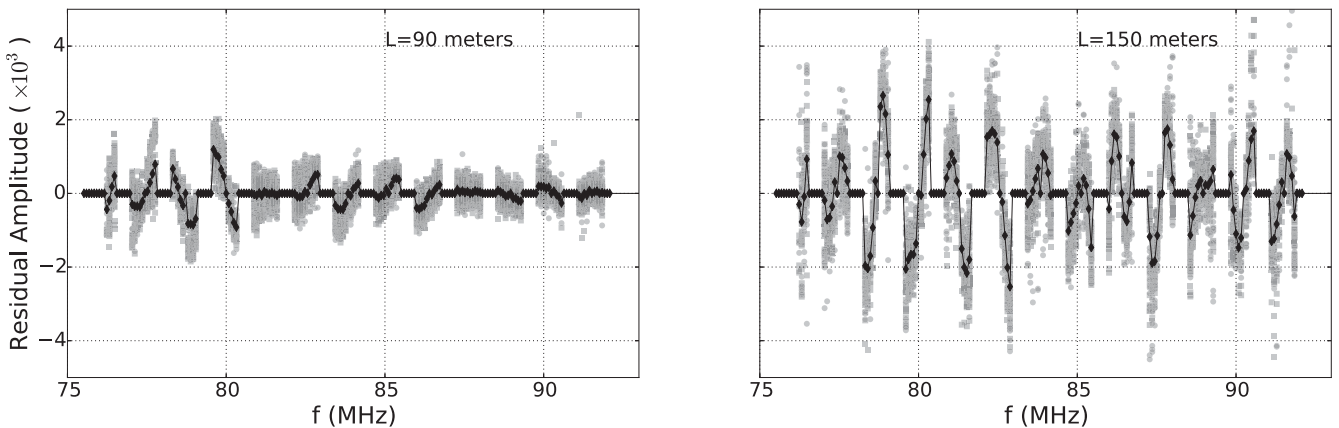


Figure 15. Left: the residuals to fitting reflection functions in our autocorrelations for all 2 min time-steps in our analysis for a representative tile with a 90 m beamformer to receiver connection (light grey points). While some scatter exists in the residuals due to fitting noise, they average to non-zero values on the order of $\sim 10^{-3}$ (black dots). These residuals are due to mismodelling the reflections and at a lower level potentially arise from digital artefacts. Right: the same as the left but for a 150 m cable whose reflection coefficient is ~ 2 as large as the 90 m cable, leading to larger residuals due to mismodelling.

contamination from ionospheric systematics on both nights (Section 4.3) and determine our best upper limits (Section 4.4).

4.1 Systematics in the 2d power spectrum

The absolute values of our two dimensional power spectrum estimates using data calibrated with auto-correlations are shown in Fig. 16 for both bands. The distinctive ‘wedge’ confines the majority of our foreground power with some supra-horizon emission clearly present out to $\sim 0.1 h \text{ Mpc}^{-1}$ as was found in observations of foreground contamination with a similarly large PAPER primary beam (Pofer et al. 2013). Smooth frequency calibration errors, arising from foreground mismodelling, may also contribute to

the supra-horizon emission along with intrinsic chromaticity in the primary beam itself. As we expected, the level of foregrounds and thermal noise is noticeably higher in our measurement of Band 1. We note that at the edge of our k_{\perp} range, there is a significant increase in power which is due to a rapid increase in thermal noise from the drop-off in the uv -coverage of our instrument. Though somewhat hard to see by eye, there are signs of coherent non-noise-like structures in both bands below $k_{\parallel} \approx 0.5 h \text{ Mpc}^{-1}$.

We confirm these faint $k_{\parallel} \lesssim 0.5 h \text{ Mpc}^{-1}$ structures as systematic contamination by inspecting the sign of our power spectrum estimate over the $k_{\perp}-k_{\parallel}$ plane. While the expected value of the even/odd cross power spectrum is always positive, k -bins that are dominated by noise have an equal probability of being positive or negative.

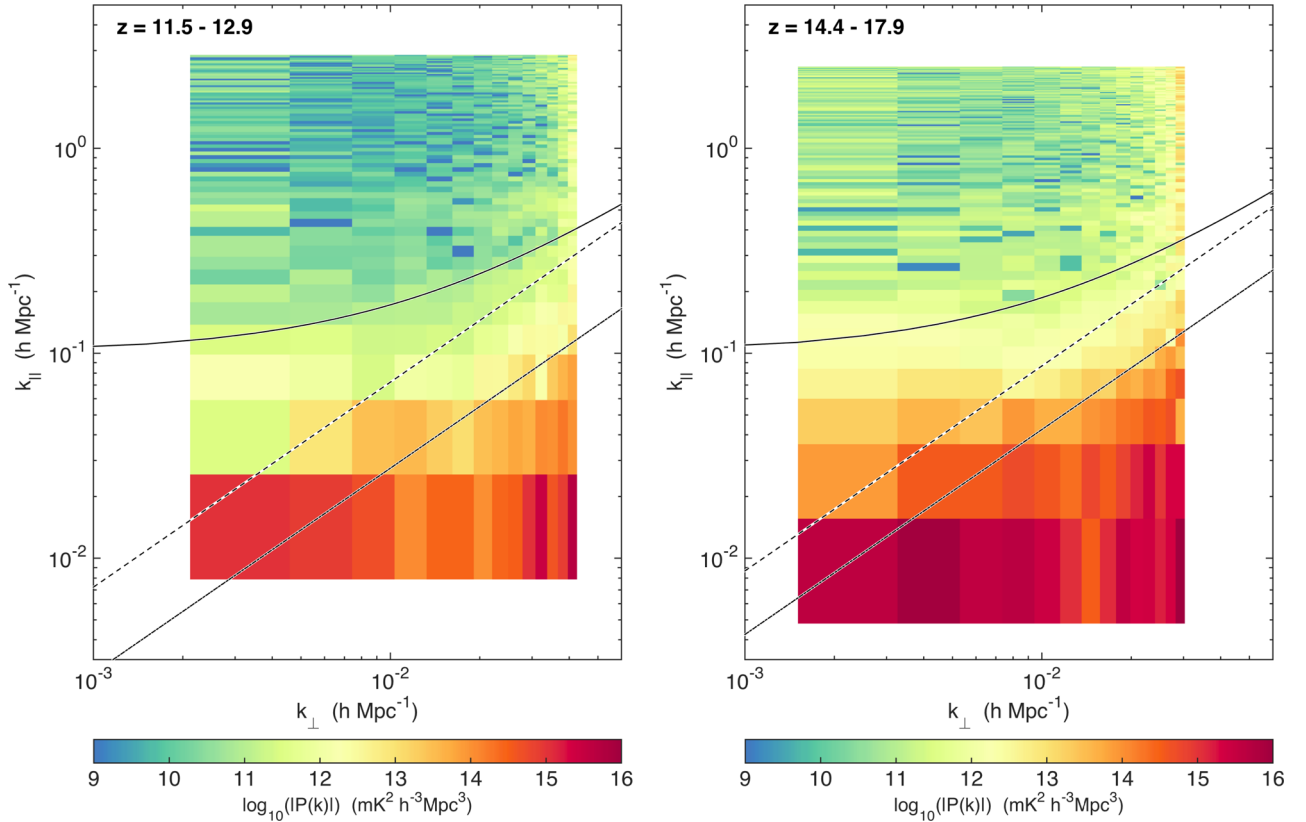


Figure 16. The absolute value of our cylindrical power spectrum estimate from our three nights of observing on Band 2 (left) and Band 1 (right). We overplot the locations of the primary beam (dash-dotted), horizon (dashed), and horizon plus a $0.1 h \text{ Mpc}^{-1}$ buffer (solid black) wedges. We see that the foregrounds are primarily contained within the wedge and that the EoR window is, for the most part, noise-like. There is some low SNR structure below $k_{\parallel} \approx 0.5 h \text{ Mpc}^{-1}$, corresponding to k_{\parallel} modes contaminated by cable reflections. The amplitude in power rises very quickly due to an increase in thermal noise which rises very quickly at large k_{\parallel} due to a rapid falloff in uv coverage beyond $k_{\perp} \sim 0.2 h \text{ Mpc}^{-1}$.

Regions in which band powers are predominately positive are detections of foregrounds or systematics. In Fig. 17 we show $P(k)$ from data calibrated with auto-correlations for both bands with an inverse hyperbolic sine colour scale to highlight regions of k -space that have positive or negative values. It is clear that the region of with $k_{\parallel} \lesssim 0.5 h \text{ Mpc}^{-1}$ is not well described by thermal noise.

Detections of foregrounds and systematics are especially visible in the ratio between the power spectrum and error bars predicted by our empirical covariance method (Fig. 18). In Fig. 19 we observe excess power at the $\sim 2\sigma$ level. While this is not a significant excess on a per cell basis, we detect this same power at high significance when we average in bins of constant $k \equiv \sqrt{k_{\perp}^2 + k_{\parallel}^2}$.

Because this excess power is present at similar levels over both of our observing subbands, one of which has a significantly greater overlap with the FM, we cannot attribute this excess to RFI. In our 1d power spectra we also find that excess power is detected in our highest redshift bin which is outside of the FM entirely. The best explanation we have for this leakage is the residual structure in the MWA’s bandpass caused by standing wave reflections on the beamformer to receiver cables. To demonstrate the plausibility of this explanation, we overlay the wedge translated to the k_{\parallel} modes corresponding to the delays of our 90 and 150 metre cables. For clarity, we do not show the 230 metre cable reflections in this overlay since their amplitudes and the number of tiles affected is comparatively small. We also observe this reflection in the 1d power spectrum which has higher signal to noise. We find that the region where one might expect contamination from

a cable reflection is in good agreement with the observed excess power.

4.2 Comparing calibration techniques

Having formed 2d power spectra and estimates of the vertical error bars, we are in a position to assess the performance of our calibration solution in removing systematics. By inspecting the signal to error ratio in the EoR window, we compare our different calibration techniques. In Fig. 20 we show the ratio of $P(k)$, binned over annuli, to the error bars in Band 1 for the calibration techniques discussed in this work. For all calibration methods, the majority of foreground detections are contained within the wedge with a $\sim 0.1 h \text{ Mpc}^{-1}$ buffer, indicating that all perform at a similar level in removing smooth gain structure within the wedge.

We first inspect a power spectrum derived from data calibrated using the initial method described in Section 2.2 in which coarse band structure is removed by averaging over tiles, the per tile amplitudes and phases of each antennas are fit to smooth polynomials, and no attempt is made to model the beamformer-receiver reflections (top left corner). Significant foreground power is visible beyond the wedge to $k \sim 0.5 h \text{ Mpc}^{-1}$ and is especially bright at the delays corresponding to the k_{\parallel} values of the cable reflections in Table 2. The fact that the 150 m delay dominates the others stems from the fact that most of our short baselines are formed from 150 m cables and that the amplitude of the reflections in the 150 m cables is larger compared to the 90 and 230 m cable lengths (Fig. 13). We next show

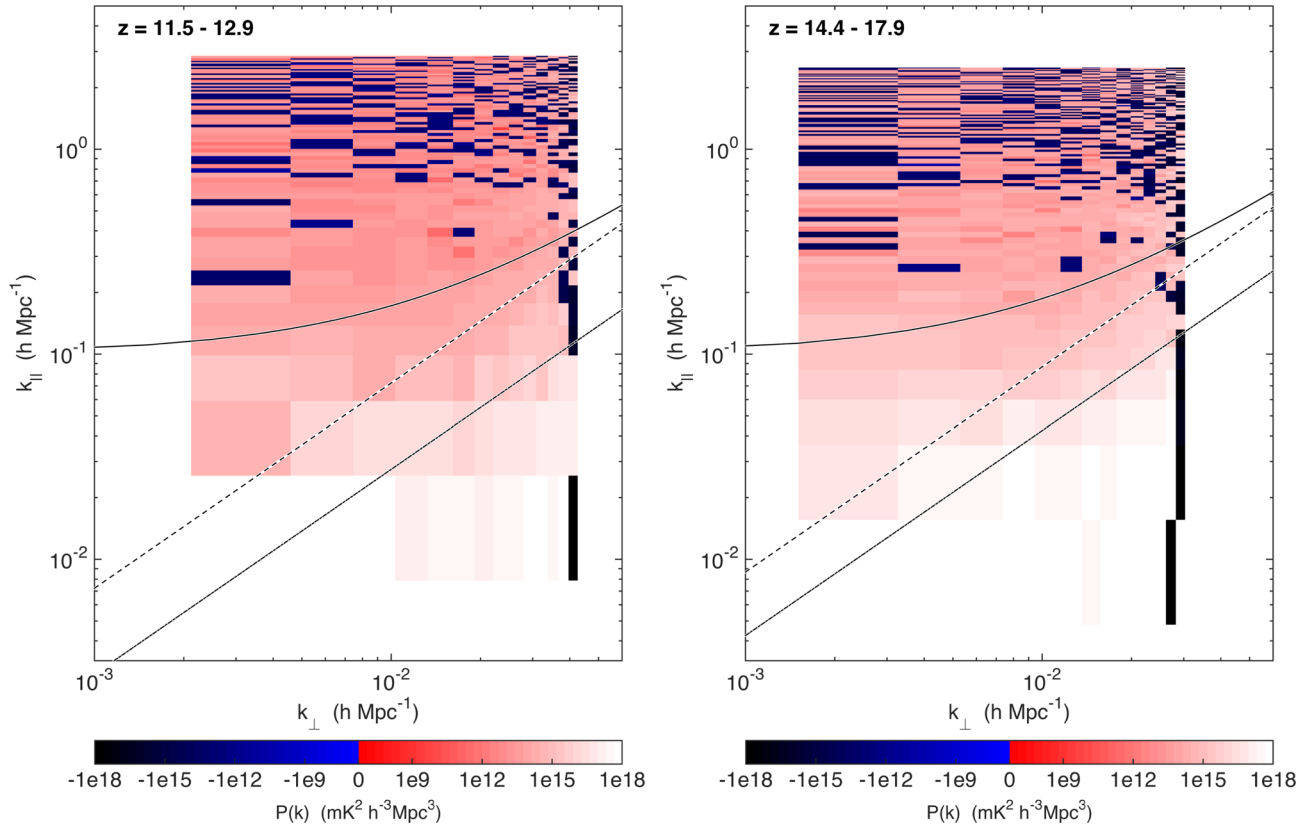


Figure 17. $P(k)$ over Band 2 (left) and Band 1 (right) with a colour scale that highlights cells with positive or negative values. We expect regions that are thermal noise dominated to contain an equal number of positive and negative estimates and regions that are dominated by foreground leakage to be entirely positive. We observe significant foreground contamination outside of the wedge up to $k_{\parallel} \approx 0.5 h \text{ Mpc}^{-1}$ in both bands.

a first attempt to fit out the reflections by averaging all calibration amplitudes in a night, dividing out a polynomial, and fitting equation (26). While the power in the bands is reduced significantly, residuals remain at the 2σ – 10σ level, especially in the reflection bands. Since our initial calibration solutions are so noisy, it makes sense that they are difficult to fit.

We finally inspect results from calibrations derived from the autocorrelations described in Section 3.4 (lower right). While there is significant reduction compared to the amplitude on the top right corner, there still exist residuals outside of the window at the $\sim 1\sigma$ – 2σ level. We think that these residuals arise from imperfect modelling of the reflection coefficients in the autocorrelation amplitudes, which will leave some reflection structure in the visibility phases. To demonstrate the impact of unmodelled reflection structure in the phases, we leave the phases of our auto-calibration solutions uncorrected for any fitted reflection coefficients (lower left) and find that significant power is reintroduced into the window.

We can get a more quantitative view of how much autocorrelations can improve calibration by taking a slice through the cylindrical power spectrum at the k_{\parallel} of our 150 m cable reflection (Fig. 21) where we see that fitting the calibration solutions was able to remove roughly an order of magnitude of the power in the reflection while AutoCal removes a factor of ~ 20 . Since the power spectrum is proportional to the square of the visibilities which are primarily contaminated by first-order reflection contributions, this corresponds to an accuracy of ≈ 20 per cent in removing the reflections in the visibilities and is consistent with the residuals observed in Fig. 15. Such inaccuracy likely arises from our inability to model the precise frequency dependence of the reflection parameters in

the phases and is on a similar order to the residuals observed in Fig. 12. Since the reflections are removed to this accuracy in the visibility, we can briefly comment on how the relative contribution of second-order reflections (which are below our noise floor even without any calibration). Since the second-order reflections appear in the data at the \tilde{r}^4 level and we have reduce their amplitude in the data from $\lesssim 0.01$ to $\lesssim 0.003$, they will enter the power spectrum at the level of $\lesssim 10 \times 10^{-1}$ the level of the 21 cm signal.

We attempted to better model the reflections by allowing for frequency evolution of the amplitudes but found little improvement in the power spectrum. We also found that we are able to obtain better fits of the autocorrelations by adding additional smooth reflection terms to equation (2) which could be important if unmodelled large scale structures bias our fits of small scale ones. However, using more complicated fits of the large scale structure, we did not observe significant improvement in power spectrum contamination. The solutions that we ultimately settled on in this analysis allow for a power-law evolution of the reflection amplitude and add an additional small delay reflection term to equation (2) which lies well within the wedge. In most of our autocorrelation fits, residuals remained at the ~ 10 per cent level, some of which may arise from secondary reflections in bent or kinked cables. While these residuals were clearly present at high SNR in the autocorrelations, we have not found a way to sufficiently model the contribution of these low level structures to our phases.

While using the autocorrelations has allowed us to characterize and subtract the fine spectral structure in the instrument better, it may not be a viable long term solution, even in the regions of the EoR window that currently appear foreground free. RFI contamination

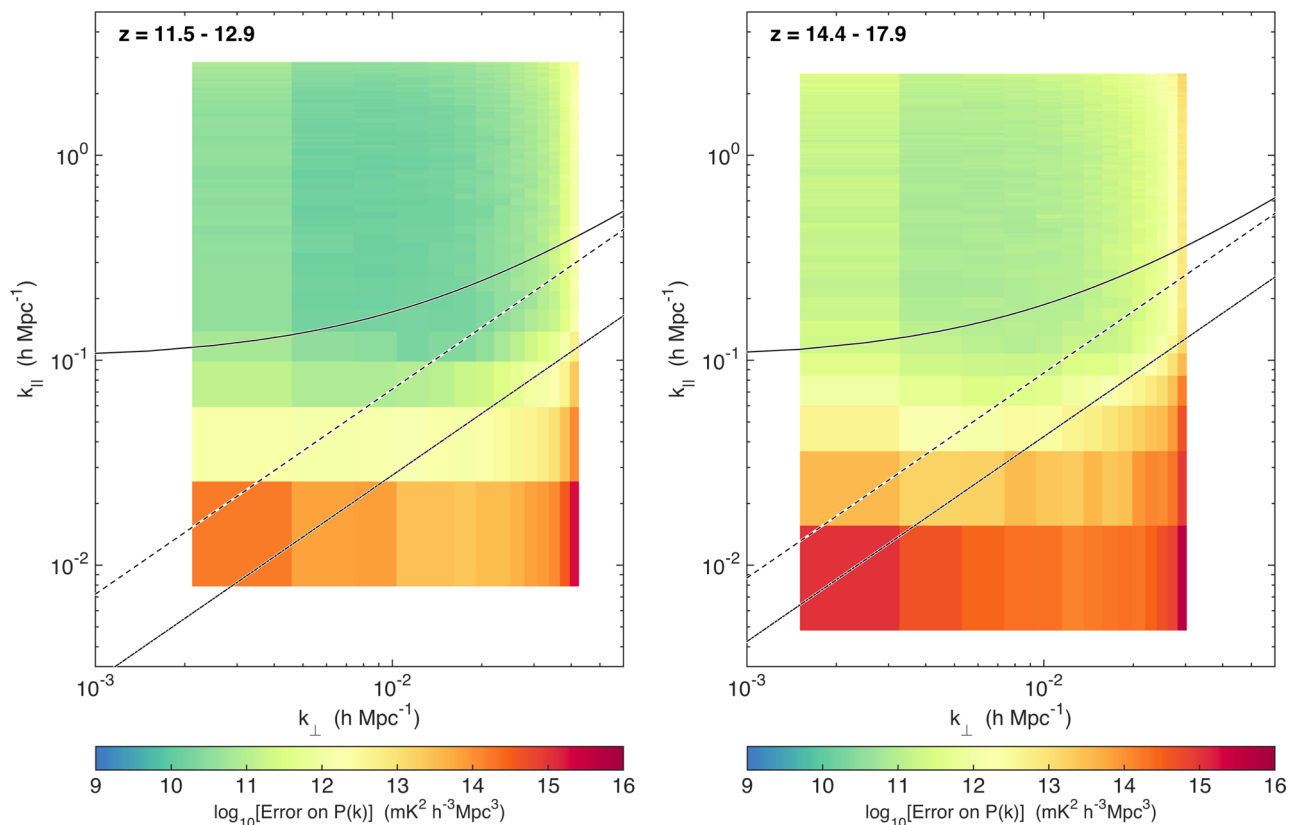


Figure 18. The errors on \hat{p} arising from residual foregrounds and thermal noise are determined by looking at even/odd difference cubes and foreground-subtracted residual cubes using the method of D15. We show the error bars on our cylindrical power spectrum here, seeing that errors arising from foregrounds are contained within the wedge. These foreground errors are maximized at the smallest and largest k_{\perp} arising from large power in diffuse emission and increasing thermal noise from a dropoff in baseline density, respectively.

and digital artefacts are known to contaminate autocorrelations and likely exist below our current noise level.

4.3 Power spectra comparison between nights of varying ionospheric activity

An open question is whether or not the ionosphere will significantly hamper measurements of the power spectrum. The fact that the severity of ionospheric effects increase with λ^2 makes the question especially pertinent at low frequency. Changes in foreground emission induced by ionospheric effects can enter the power spectrum in two ways: through calibration and through the foreground residuals themselves. We check whether either of these potential error sources have an observable effect on our 1d power spectrum in Fig. 22 by comparing power spectra derived from 1.4 h of Band 1 data on September 5, over which ionospheric activity was comparatively mild to the same number of hours of Band 1 data on September 6 where differential refraction was approximately twice as severe.

We find that the power spectra, which are estimated from data outside of the wedge, are consistent with each other. This result confirms the intuitive idea that since ionospheric errors in the foreground model are spectrally smooth (evolving as $\sim\lambda^2$), they should be contained within the wedge. We also extended our 1d power spectrum estimation into the wedge to see whether the foreground detections appeared to be significantly different and find that they are not. This suggests that the random errors induced by the ionosphere average down with time. It is important to keep in mind that the spatial scales being probed in our analysis are relatively

large, on the order of $\gtrsim 2.5^{\circ}$, while ionospheric refraction at these frequencies effects sub-arcminute scales. Hence the contamination that we might expect from ionospheric refraction should be small. Amplitude scintillation effects are prominent on short baselines (V15) and likely dominate any contamination, however their spectral coherence still constrains them to be predominantly within the wedge (V16).

4.4 First upper limits on the 21 cm power spectrum during the pre-reionization epoch

We limit our 1d power spectra to redshift widths of $\Delta z \sim 1.5$ to minimize effects from cosmic evolution. A redshift interval of $\Delta z \sim 0.5$ is the range most cited in the literature over which the statistics of the brightness temperature field are expected to be stationary (Mao et al. 2008). However, at higher redshift, the frequency range corresponding to $\Delta z \approx 0.5$ decreases as $(1+z)^{-2}$ with $\Delta z = 0.5$ corresponding to a bandwidth of only 2.45 MHz by $z = 16$. Reducing our bandwidth to such a small interval leads to poor k resolution which we prefer to maintain for assessing systematics. Since we are far from a detection, we opt for a larger redshift interval than we would otherwise use if we were actually observing the cosmological signal. In Fig. 23 we show 1d power spectra derived from our three hours of observing. Vertical error bars give 2σ uncertainties and the horizontal error bars give the width of our window functions. The amplitudes of our power spectrum values are consistent with thermal noise except for the regions of k -space below $k_{\parallel} \lesssim 0.5 h \text{ Mpc}^{-1}$. At $k \approx 1 h \text{ Mpc}^{-1}$, where our measurements are

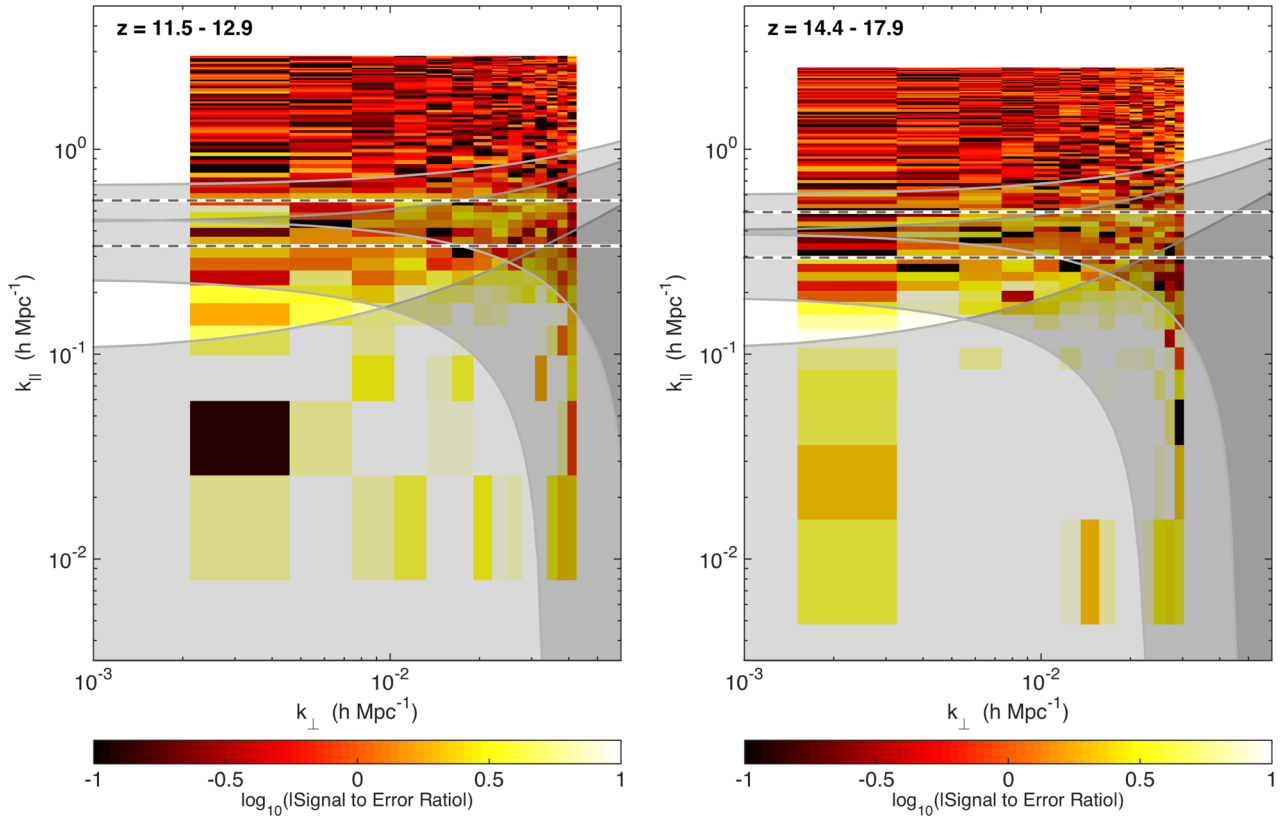


Figure 19. The foreground contamination within the wedge along with residual detections due to miscalibrated fine frequency features in the bandpass are especially clear in plots of the ratio between power and the error bars estimated by the empirical covariance method of D15. We overplot the wedge with a $0.1 h \text{ Mpc}^{-1}$ buffer along with the wedge translated to cable reflection delays of our 90 and 150 m receiver to beamformer cables to highlight the effect of this systematic.

well described by thermal noise, our upper limits are on the order of 100 times higher than the results presented in D15 in which a similar three hour upper limit was established at $\approx 180 \text{ MHz}$. This factor of ≈ 100 is consistent with what we expect from equation (8). The sky temperature increases with decreasing frequency as $f^{-2.6}$ leading to a factor of ≈ 30 from T_{sys}^2 while λ^4/A_e^2 introduces an additional factor of 4–10.

The detections at small k are many orders of magnitude larger than the expected cosmological signal from a 21cmFAST simulation (Mesinger, Furlanetto & Cen 2011) (blue solid lines) so they cannot possibly originate from the redshifted H I emission. Instead, these detections are most likely the miscalibrated reflection structure observed in our 2D power spectra. We shade out regions of the k axis in which we expect contamination given the reflections discussed above and find that they correspond to the same modes where detections are observed. These systematic detections occupy the regions of Fourier space where our interferometer has the greatest sensitivity to the cosmic signal. Since we do not expect the systematics dominated regions to integrate down, a detection with the MWA in its current state using the techniques presented in this work would have to take place at $k \gtrsim 0.5 h \text{ Mpc}^{-1}$, requiring over 10^5 h of integration – a rather infeasible time-scale. Thus, in order to probe the pre-reionization epoch, improvements in calibration and/or changes in the hardware of the MWA will have to be implemented. We note that at lower redshifts, the primary beam is smaller and the k -modes occupied by reflections are farther away from the sensitivity sweet spot, so it is less likely that this problem will prevent the MWA from detecting the EoR power spectrum.

Our best upper limits fall within the region of Fourier space with systematic errors and, while we do not expect them to integrate down with more observing time, we can infer that $\Delta^2(k)$ is less than $2.5 \times 10^7 \text{ mK}^2$ at $k = 0.18 h \text{ Mpc}^{-1}$ and $z = 12.2$, $8.3 \times 10^7 \text{ mK}^2$ at $k = 0.21 h \text{ Mpc}^{-1}$ and $z = 15.35$, and $2.7 \times 10^8 \text{ mK}^2$ at $k = 0.22 h \text{ Mpc}^{-1}$ and $z = 17.05$, all at 95 per cent confidence.

4.5 The outlook for EoR measurements on the MWA

A pertinent question arising from our analysis is how much the observed reflections impact or limit observations with the MWA of the Epoch of Reionization power spectrum at higher frequencies. The answer depends significantly on the calibration and reduction approach and as of now, several different efforts using alternative calibration and reduction schemes are being undertaken (Jacobs et al. 2016). The analyses in D15 and Beardsley et al. (in preparation) are calibrated in a way similar to this work, employing limited calibration parameters to avoid detrimental modelling errors. After an integration time of ≈ 3 hours, D15 also observe the cable reflections above the thermal noise level, however the smallest k mode occupied by the shortest 90 m cable lies at $k \approx 0.4 h \text{ Mpc}^{-1}$ while the delay width is narrower due to the smaller primary beam, causing the wedge to occupy fewer k_{\parallel} modes. As a result, there are regions of k -space below the shortest reflection that are still consistent with noise. Such miscalibrated structure should be highly detectable after ≈ 10 – 30 h of integration but the results of such an analysis are still forthcoming (Beardsley et al., in preparation). If the reflections can be corrected to the $\sim 10^{-3}$ level as was done in

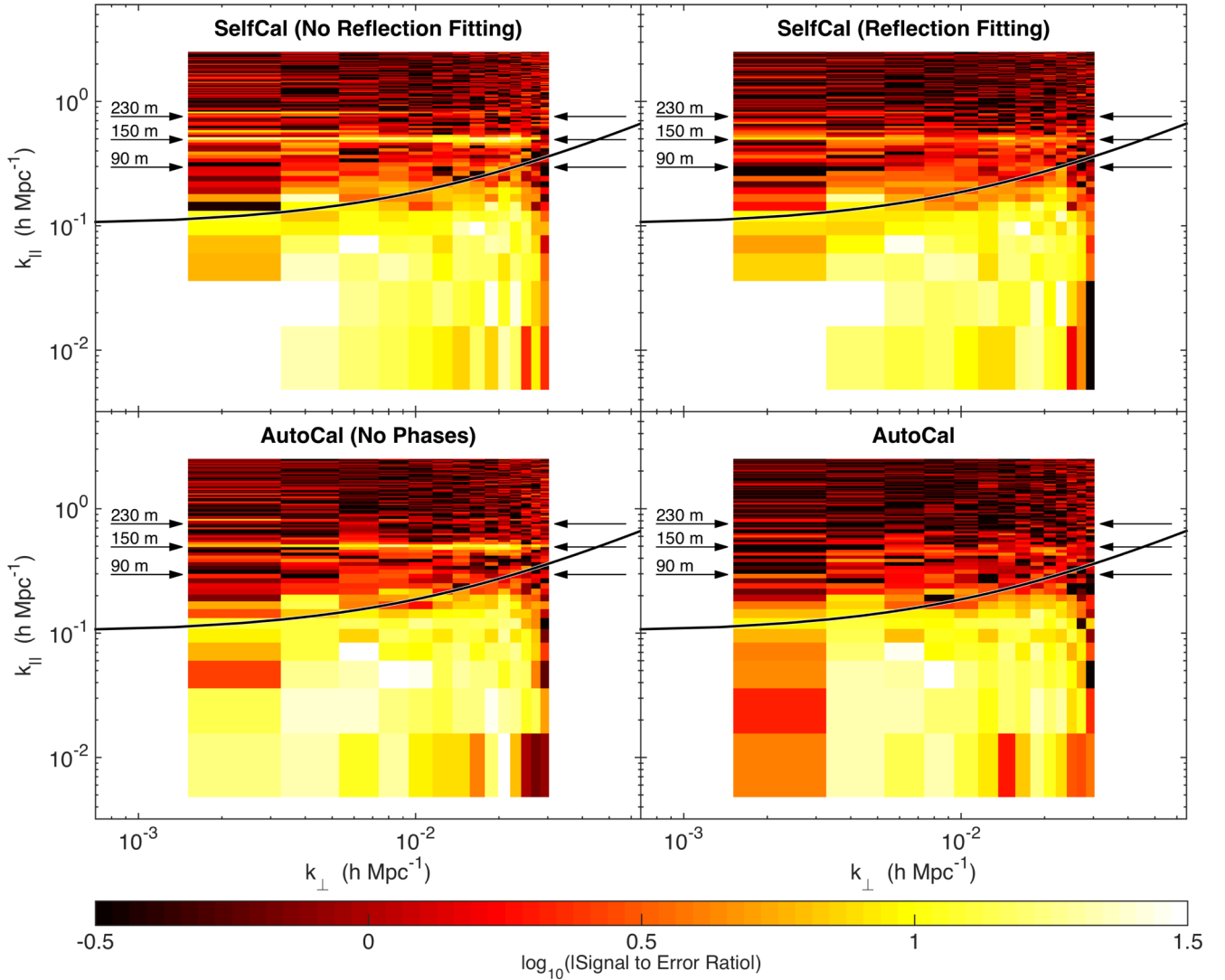


Figure 20. Here we show the ratio between our 2D power spectrum and the error bars estimated by the empirical covariance method of D15. On the top left, we show our data calibrated using our initial calibration (see Section 2.2) with no attempt made to correct for standing wave structure in the MWA bandpass. Bright, band-like structures are clearly visible at the delays associated with reflections. On the top right, we show a first attempt to correct for cable reflections by fitting a sinusoidal model to rather noisy calibration solutions that had been integrated over a night of observing (1.5 h each night). While the bands appear weaker, they are still quite visible above the noise. In the bottom right panel, we show the same plot with calibration solutions using scaled autocorrelations described in Section 3.4. In the lower left panel we show a power spectrum with calibration solutions using autocorrelations for the amplitudes but without any attempt to correct reflections in the phase solutions. Pronounced reflection features are visible in this power spectrum, indicating that any mismodelled reflection structure in the phases will contaminate our measurement.

this analysis, the region below the first reflection should remain free of contamination from the beam-former to receiver reflections.

Additional calibration pipelines, which include far greater degrees of freedom, such as the Real Time System (RTS) (Mitchell et al. 2008; Ord et al. 2010) and the reduction pipeline discussed in Offringa et al. (2016) include direction-dependent calibration, ionospheric phase fitting, and greater frequency resolution, are also being applied to MWA data sets. A recent upper limit at 180 MHz derived from RTS calibrated data and the CHIPS power spectrum estimator did not show evidence of the cable reflections being present (Trott et al. 2016). It is likely that the enhanced degrees of freedom allowed by RTS calibration did a better job at removing the structure from reflections but it is difficult to tell given that the error bars due to thermal noise at the comoving scales relevant to

these reflections are an order of magnitude larger than those in D15 because of the shorter integration time. Ultimately, the increase in the number of fitting parameters may enhance the removal of instrumental chromaticity in the EoR window, however simulations by Barry et al. (2016) show that small errors in ones calibration model will introduce power into the window in excess of the 21 cm signal unless the intrinsic bandpass is smooth enough to be modelled by a small number of parameters or a source model exists with an accuracy significantly beyond what is currently available. Ultimately, the increase in fitting parameters may enhance the removal of instrumental chromaticity in the EoR window. Whether they can be introduced without adding power into the EoR window in excess of the signal, due to small errors in source modelling is still an open question that is currently being investigated.

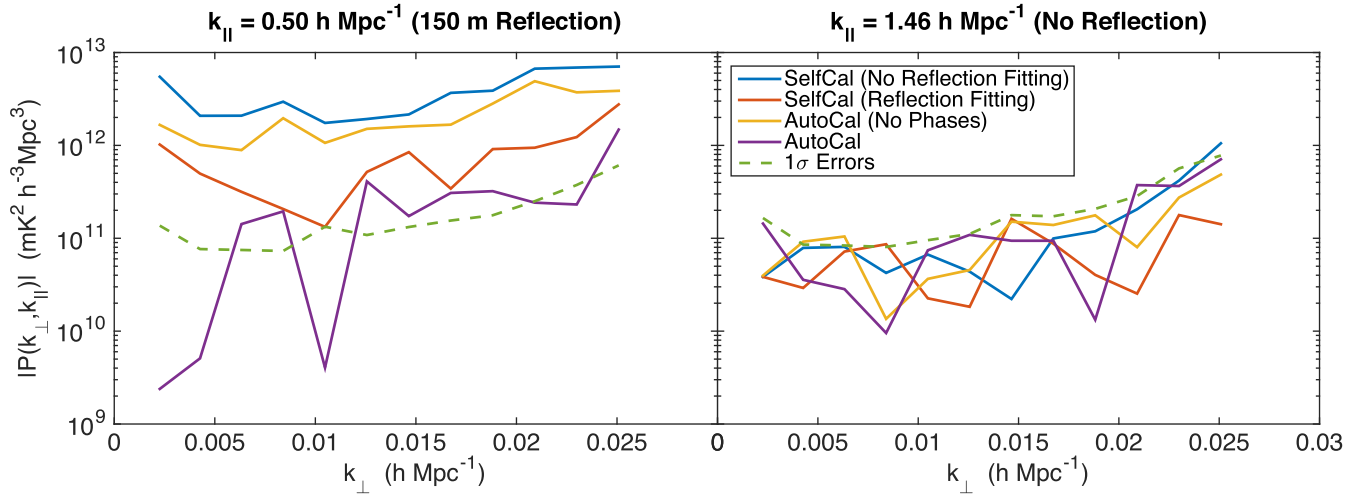


Figure 21. The level of power at a fixed k_{\parallel} corresponding to the delay of reflections from our 150 m cable (left), and comparing it to a value of k_{\parallel} unaffected by cable reflections (right). The blue line shows the power spectrum level for calibration in which the bandpass is modelled as a polynomial with no attempt to correct fine frequency scale reflections. We see that power is on the order of ~ 50 times the thermal noise level (green-dashed line). Attempting to fit the reflections to calibration solutions integrated over each night gives us an improvement in the power level by roughly an order of magnitude (orange solid line). Using calibration solutions derived from autocorrelations brings down the reflection power by another factor of a few (purple solid line) but is still unable to bring the majority of measurements below the $\sim 1\sigma$ level. While we think that the autocorrelations accurately capture the fine frequency structure of the gains, we are still forced to model this fine frequency structure and predict it in the phases. Residual power is likely due to inaccuracies in this modelling. The right-hand panel shows all data below the estimated noise level. This is due to the fact that in Dillon et al. (2014) it is shown that the method for calculating error bars layed-out in Liu & Tegmark (2011); Dillon et al. (2013) slightly overestimates the noise.

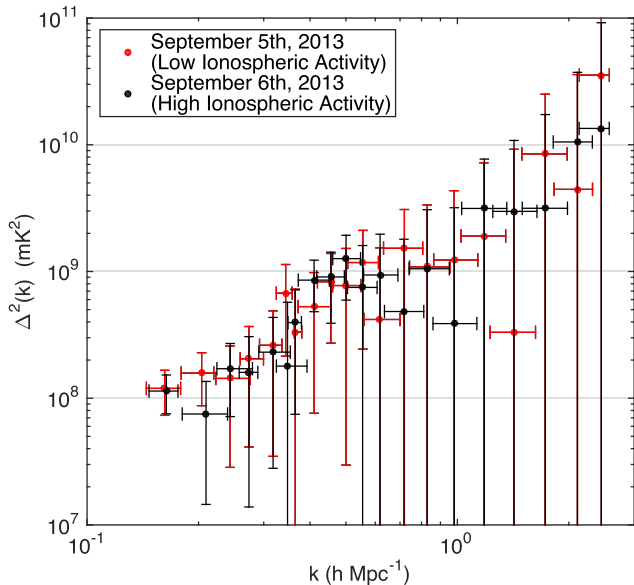


Figure 22. The Band 1, 1 d power spectra from our two nights of observing: 2013 September 5 (black) and 2013 September 6 (red). We saw in Fig. 7, that the magnitude of refractions on September 6 were on average twice as severe. The two power spectra nearly indistinguishable (within error bars) despite the significant differences in conditions, indicating that ionospheric systematics do not have a significant effect after three hours of integration, even at these low frequencies.

5 CONCLUSIONS AND FUTURE EXPERIMENTAL CONSIDERATIONS

In this paper, we have presented low frequency radio observations with the MWA at unprecedentedly high redshifts between 11.6 and 17.9. Our goals in conducting these observations were to place upper limits on the 21 cm power spectrum during the Epoch of

X-ray heating and to assess the levels of systematics which are expected to be generally worse than at EoR frequencies. These systematics include ionospheric effects, RFI (due to the FM band) and increased thermal noise. We need to control these systematics if we are to learn the detailed properties of the sources that heated the IGM; be they the first generation of stellar mass black holes, the hot interstellar medium left over from the first supernovae explosions in the Universe, or dark matter annihilation.

With regards to RFI, we have found after three hours of integration that existing algorithms are sufficient to flag RFI below the FM band. Within the FM band, we have found that only a handful of channels are contaminated continuously and that after discarding them our power spectra do not show any evidence of RFI contamination. This bodes well for future planned 21 cm experiments at the MRO such as the SKA-low which is expected to make high signal to noise detections of the power spectrum (Koopmans et al. 2015). However, we are still many orders of magnitude above the level of a detection and reducing the thermal noise through longer integrations may reveal lower level RFI.

Over two nights of observing, we encounter different ionospheric conditions, observed quantitatively using the differential refraction metric described in Cohen & Röttgering (2009). We establish that ionospheric fluctuations are the source of observed position shifts by comparing the level of refraction in our two observing bands and find that they exhibit the expected λ^2 evolution. Diffractive scales on the second night of ≈ 5 km are a factor of 2 shorter than the first night, indicative of more severe ionospheric activity. When we compare the 1 d power spectra derived from an equal amount of data on each night, we find that they are very similar to each other, lending support to the idea that since ionospheric effects on calibration and foreground residuals are spectrally smooth, they should not contaminate the EoR window.

While the majority of foreground power is contained within the wedge, we find high-significance foreground detections within the EoR window out to a $k_{\parallel} \lesssim 0.5 h \text{ Mpc}^{-1}$. These contaminated

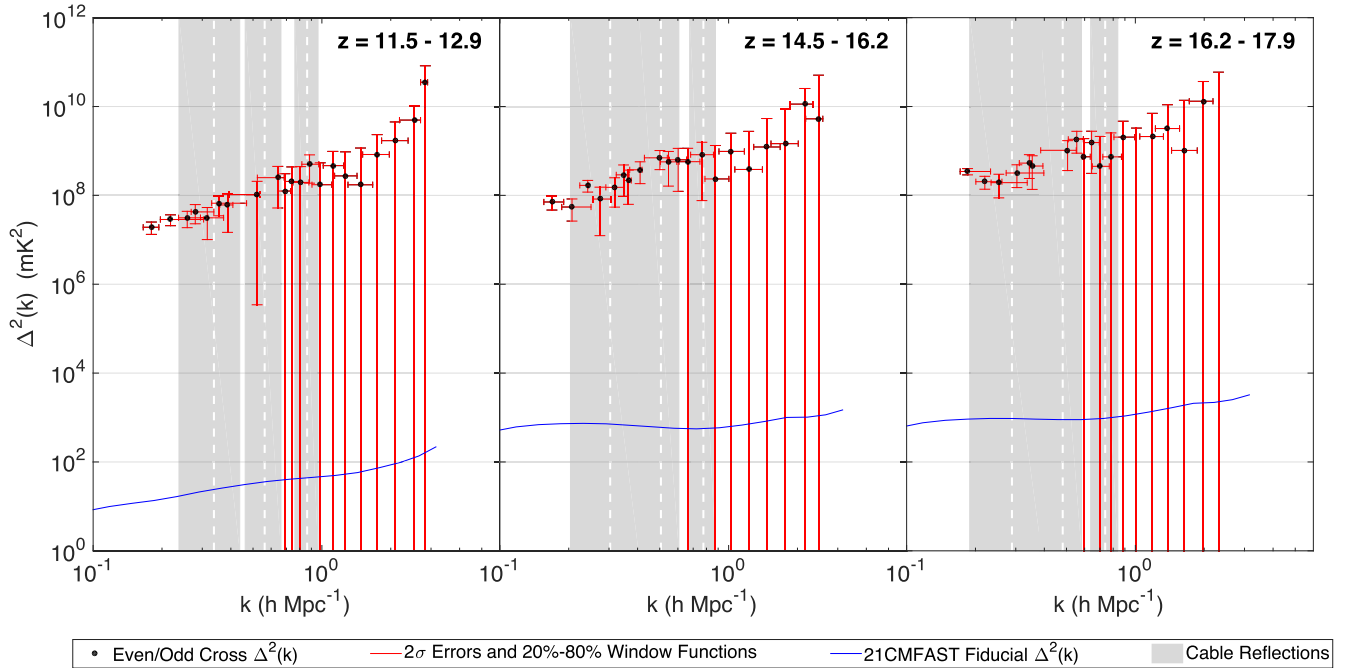


Figure 23. Dimensionless 1d power spectra derived by Integrating spherical shells excluding the foreground contaminated wedge region with a $0.1 h \text{ Mpc}^{-1}$ buffer. Black dots indicate the mean estimated from the weighted average in each bin. Vertical error bars denote the 2σ uncertainties while horizontal error bars indicate the width of window functions. We also shade regions of k -space that we expect to have some level of foreground contamination due to uncalibrated cable reflection structure. Grey shaded regions clearly correspond to regions in which our power spectrum measurements are not consistent with thermal noise. We note that where our upper limits do agree with thermal noise, the power spectrum is on the order of ~ 100 times larger than the upper limits set with the MWA at $\approx 180 \text{ MHz}$ (D15). This factor is reasonable given that the sky noise (noise power spectrum) scales with $\sim f^{-2.6}$ ($f^{-5.2}$) and the primary beam solid angle increases as $\sim f^2$.

regions are consistent with miscalibrated cable reflections. We are able to obtain an order of magnitude improvement on removing the worst of these features using fits to autocorrelations, however they still limit our sensitivity at the 2σ – 5σ level. In addition, since autocorrelations are generally contaminated by RFI and digital artefacts, it is likely that in reducing the dominant obstacle in our data, we have introduced additional features that are below the noise level of this analysis. Since the reflections occupy the regions of k -space where we would otherwise expect the greatest cosmological sensitivity, our best upper limits are a factor of a few larger than the limits we would obtain if we were thermal noise limited. Cable reflections are especially pernicious at higher redshifts because the increasing primary beam width adds foreground power to delays ever closer to the horizon. While supra horizon emission off of the wedge moves up in k_{\parallel} , the modes occupied by cable reflections move down, increasing in width. The EoR window is crushed between the shortest reflection mode and the top of the wedge.

While our observations on the MWA will not integrate down below $\approx 10^8 \text{ mK}^2$ at $k \lesssim 0.5 h \text{ Mpc}^{-1}$ and is limited by the intrinsic spectral structure of the instrument, the systematics encountered in this analysis do not prevent 21 cm observations at high redshift in general. A robust source catalog, that includes emission all the way down to the horizon along with precise models of the primary beam will lead to less foreground power bleeding from the edge of the wedge, (Thyagarajan et al. 2015a,b; Pober et al. 2016) and potentially open up a foreground free region under the first cable reflection. Resolving the question of cosmological signal loss and mixing of foreground spectral structure from large to short baselines may enable us to calibrate with more free parameters, better capturing the spectral structure of the bandpass. More robust calibration

of these features may also be obtainable with a redundant array (Wieringa 1992; Liu et al. 2010; Zheng et al. 2014). The 128-tile MWA has very little redundancy by design, however an additional 128 tile expansion is expected to introduce two highly redundant, hex-packed, subarrays (Tingay, private communication). The final plan for HERA, which is currently under construction, is dominated by 331 hexagonally packed dishes. Its layout is designed to take advantage of redundant calibration as well (Pober et al. 2014). Finally, calibration using injected signals (Patra et al. 2015) can also be employed to make high precision measurements of the bandpass.

The most sure way of eliminating reflection features is to remove them in hardware either by ensuring better impedance matching on the cable connections, changing the cable lengths to move reflections out of the window, or early digitization. The current HERA design employs cables no longer than 35 m in length, translating to $k_{\parallel} = 0.09 h \text{ Mpc}^{-1}$ at $z = 16$ and ensures that reflections within the dish are below an acceptable level (Ewall-Wice et al. 2016b; Patra et al., in preparation; Thyagarajan et al. 2016), while the planned MWA phase III upgrade and the SKA are considering digitization at the beamformers (Tingay, private communication), eliminating reflections altogether.

While measurements of the 21 cm line at EoR frequencies can teach us about the nature of UV photon sources and constrain cool thermal histories, a significant number of scenarios predict saturation of heating’s contribution to brightness temperature fluctuations during reionization. In order to learn of the detailed properties of the sources that heated the IGM and to exploit the full potential of the 21 cm line as a cosmological and astrophysical probe, we will invariably want to extend our search to as low a frequency as possible. In this work we have obtained a first look at the

systematics facing us in this high redshift realm and have found that most of them are navigable. As of now, our primary limitation lies in the design of our instrument and calibration, both of which can be dramatically improved on relatively short time-scales. Ultimately, we expect a combination of improvements in instrumental design including shorter/no cables to keep reflections inside of the wedge and redundant baseline layouts allowing for more robust calibration to allow for much deeper integrations in the near future.

ACKNOWLEDGEMENTS

We would like to thank the referees for their helpful comments. We extend our gratitude to Jeff Zhang and Adrian Liu for useful discussions. This work was supported by NSF Grants AST-0457585, AST-0821321, AST-1105835, AST-1410719, AST-1410484, AST-1411622, and AST-1440343, by the MIT School of Science, by the Marble Astrophysics Fund, and by generous donations from Jonathan Rothberg and an anonymous donor. AEW acknowledges support from the National Science Foundation Graduate Research Fellowship under Grant No. 1122374. AM acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 638809 – AIDA).

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 MWA comes from the U.S. National Science Foundation (grants AST-0457585, PHY-0835713, CAREER-0847753, and AST-0908884), the Australian Research Council (LIEF grants LE0775621 and LE0882938), the U.S. Air Force Office of Scientific Research (grant FA9550-0510247), and the Centre for All-sky Astrophysics (an Australian Research Council Centre of Excellence funded by grant CE110001020). Support is also provided by the Smithsonian Astrophysical Observatory, the Raman Research Institute, the Australian National University, and the Victoria University of Wellington (via grant MED-E1799 from the New Zealand Ministry of Economic Development and an IBM Shared University Research Grant). The Australian Federal government provides additional support via the Commonwealth Scientific and Industrial Research Organisation (CSIRO), National Collaborative Research Infrastructure Strategy, Education Investment Fund, and the Australia India Strategic Research Fund, and Astronomy Australia Limited, under contract to Curtin University. We acknowledge the iVEC Petabyte Data Store, the Initiative in Innovative Computing and the CUDA Center for Excellence sponsored by NVIDIA at Harvard University, and the International Centre for Radio Astronomy Research (ICRAR), a Joint Venture of Curtin University and The University of Western Australia, funded by the Western Australian State government.

REFERENCES

- Ali Z. S. et al., 2015, *ApJ*, 809, 61
 Arora B. S. et al., 2015, *PASA*, 32, 29
 Asad K. M. B. et al., 2015, *MNRAS*, 451, 3709
 Baek S., Semelin B., Di Matteo P., Revaz Y., Combes F., 2010, *A&A*, 523, A4
 Barkana R., Loeb A., 2001, *Phys. Rep.*, 349, 125
 Barkana R., Loeb A., 2005, *ApJ*, 626, 1
 Barry N., Hazelton B., Sullivan I., Morales M. F., Pober J. C., 2016, preprint ([arXiv:1603.00607](https://arxiv.org/abs/1603.00607))
 Bartels J., Heck N. H., Johnston H. F., 1939, *J. Geophys. Res.*, 44, 411
 Bernardi G. et al., 2013, *ApJ*, 771, 105
 Bowman J. D., Rogers A. E. E., 2010, *Nature*, 468, 796
 Bowman J. D. et al., 2013, *PASA*, 30, 31
 Burns J. O. et al., 2012, *Adv. Space Res.*, 49, 433
 Carilli C. L., Gnedin N. Y., Owen F., 2002, *ApJ*, 577, 22
 Christian P., Loeb A., 2013, *J. Cosmol. Astropart. Phys.*, 9, 14
 Ciardi B. et al., 2013, *MNRAS*, 428, 1755
 Cohen A. S., Röttgering H. J. A., 2009, *AJ*, 138, 439
 Cotton W. D., 2012, *J. Astron. Instrum.*, 1, 1250001
 da Rosa A. V., Waldman H., Bendito J., Garriotto O. K., 1973, *J. Atmos. Terr. Phys.*, 35, 1429
 Datta A., Bowman J. D., Carilli C. L., 2010, *ApJ*, 724, 526
 Di Matteo T., Perna R., Abel T., Rees M. J., 2002, *ApJ*, 564, 576
 Dillon J. S., Liu A., Tegmark M., 2013, *Phys. Rev. D*, 87, 043005
 Dillon J. S. et al., 2014, *Phys. Rev. D*, 89, 023002
 Dillon J. S. et al., 2015a, *Phys. Rev. D*, 91, 023002
 Dillon J. S. et al., 2015b, *Phys. Rev. D*, 91, 123011 (D15)
 Ewall-Wice A., Dillon J. S., Mesinger A., Hewitt J., 2014, *MNRAS*, 441, 2476
 Ewall-Wice A., Hewitt J., Mesinger A., Dillon J. S., Liu A., Pober J., 2016a, *MNRAS*, 458, 2710
 Ewall-Wice A. et al., 2016b, preprint ([arXiv:1602.06277](https://arxiv.org/abs/1602.06277))
 Fixsen D. J. et al., 2011, *ApJ*, 734, 5
 Fomalont E. B., Ebner K. A., van Breugel W. J. M., Ekers R. D., 1989, *ApJ*, 346, L17
 Furlanetto S. R., 2006a, *MNRAS*, 370, 1867
 Furlanetto S. R., 2006b, *MNRAS*, 371, 867
 Furlanetto S. R., Loeb A., 2002, *ApJ*, 579, 1
 Furlanetto S. R., Oh S. P., Briggs F. H., 2006, *Phys. Rep.*, 433, 181
 Geil P. M., Gaensler B. M., Wyithe J. S. B., 2011, *MNRAS*, 418, 516
 Greenhill L. J., Bernardi G., 2012, preprint ([arXiv:1201.1700](https://arxiv.org/abs/1201.1700))
 Greig B., Mesinger A., 2015, *MNRAS*, 449, 4246
 Griffiths D. J., 2013, *Introduction to Electrodynamics*, 4th edn. Pearson, Boston, MA
 Hales S. E. G., Baldwin J. E., Warner P. J., 1988, *MNRAS*, 234, 919
 Hancock P. J., Murphy T., Gaensler B. M., Hopkins A., Curran J. R., 2012, *MNRAS*, 422, 1812
 Haslam C. G. T., Salter C. J., Stoffel H., Wilson W. E., 1982, *A&AS*, 47, 1
 Hazelton B. J., Morales M. F., Sullivan I. S., 2013, *ApJ*, 770, 156
 Helmboldt J. F., Intema H. T., 2012, *Radio Sci.*, 47, 0
 Helmboldt J. F., Lane W. M., Cotton W. D., 2012, *Radio Sci.*, 47, 5008
 Jacobs D. C. et al., 2011, *ApJ*, 734, L34
 Jacobs D. C. et al., 2013, *ApJ*, 776, 108
 Jacobs D. C. et al., 2015, *ApJ*, 801, 51
 Jacobs D. et al., 2016, *ApJ*, preprint ([arXiv:1605.06978](https://arxiv.org/abs/1605.06978))
 Jelić V., Zaroubi S., Labropoulos P., Bernardi G., de Bruyn A. G., Koopmans L. V. E., 2010, *MNRAS*, 409, 1647
 Jelić V. et al., 2014, *A&A*, 568, A101
 Jester S., Falcke H., 2009, *New Astron. Rev.*, 53, 1
 Kassim N. E., Perley R. A., Erickson W. C., Dwarakanath K. S., 1993, *AJ*, 106, 2218
 Kassim N. E. et al., 2007, *ApJS*, 172, 686
 Koopmans L. V. E. et al., 2015, preprint ([arXiv:1505.07568](https://arxiv.org/abs/1505.07568))
 Lane W. M., Cotton W. D., van Velzen S., Clarke T. E., Kassim N. E., Helmboldt J. F., Lazio T. J. W., Cohen A. S., 2014, *MNRAS*, 440, 327
 Large M. I., Mills B. Y., Little A. G., Crawford D. F., Sutton J. M., 1981, *MNRAS*, 194, 693
 Liu A., Tegmark M., 2011, *Phys. Rev. D*, 83, 103006
 Liu A., Tegmark M., Morrison S., Lutomirski A., Zaldarriaga M., 2010, *MNRAS*, 408, 1029
 Liu A., Parsons A. R., Trott C. M., 2014a, *Phys. Rev. D*, 90, 023018
 Liu A., Parsons A. R., Trott C. M., 2014b, *Phys. Rev. D*, 90, 023019
 Loi S. T. et al., 2015a, *Geophys. Res. Lett.*, 42, 3707
 Loi S. T. et al., 2015b, *Radio Sci.*, 50, 574
 Lonsdale C. J. et al., 2009, *IEEE Proc.*, 97, 1497
 Mack K. J., Wyithe J. S. B., 2012, *MNRAS*, 425, 2988
 McMullin J. P., Waters B., Schiebel D., Young W., Golap K., 2007, in Shaw R. A., Hill F., Bell D. J., eds. *ASP Conf. Ser. Vol. 376, Astronomical Data*

- Analysis Software and Systems XVI. Astron. Soc. Pac., San Francisco, p. 127
- Mao Y., Tegmark M., McQuinn M., Zaldarriaga M., Zahn O., 2008, *Phys. Rev. D*, 78, 023529
- Mesinger A., Furlanetto S., Cen R., 2011, *MNRAS*, 411, 955
- Mesinger A., Ferrara A., Spiegel D. S., 2013, *MNRAS*, 431, 621
- Mesinger A., Ewall-Wice A., Hewitt J., 2014, *MNRAS*, 439, 3262
- Mineo S., Gilfanov M., Sunyaev R., 2012, *MNRAS*, 419, 2095
- Mirabel I. F., Dijkstra M., Laurent P., Loeb A., Pritchard J. R., 2011, *A&A*, 528, A149
- Mitchell D. A., Greenhill L. J., Wayth R. B., Sault R. J., Lonsdale C. J., Cappallo R. J., Morales M. F., Ord S. M., 2008, *IEEE J. Sel. Top. Signal Process.*, 2, 707
- Moore D. F., Aguirre J. E., Parsons A. R., Jacobs D. C., Pober J. C., 2013, *ApJ*, 769, 154
- Moore D. et al., 2015, preprint ([arXiv:1502.05072](https://arxiv.org/abs/1502.05072))
- Morales M. F., Hewitt J., 2004, *ApJ*, 615, 7
- Morales M. F., Wyithe J. S. B., 2010, *ARA&A*, 48, 127
- Morales M. F., Hazelton B., Sullivan I., Beardsley A., 2012, *ApJ*, 752, 137
- Neben A. R. et al., 2015, *Radio Sci.*, 50, 614
- Offringa A. R., van de Gronde J. J., Roerdink J. B. T. M., 2012, *A&A*, 539, A95
- Offringa A. R. et al., 2014, *MNRAS*, 444, 606
- Offringa A. R. et al., 2015, *PASA*, 32, 8
- Offringa A. R. et al., 2016, *MNRAS*, 458, 1057
- Ord S. M. et al., 2010, *PASP*, 122, 1353
- Ord S. M. et al., 2015, *PASA*, 32, 6
- Paciga G. et al., 2013, *MNRAS*, 433, 639
- Pacucci F., Mesinger A., Mineo S., Ferrara A., 2014, *MNRAS*, 443, 678
- Parsons A. R. et al., 2010, *AJ*, 139, 1468
- Parsons A. R., Pober J. C., Aguirre J. E., Carilli C. L., Jacobs D. C., Moore D. F., 2012, *ApJ*, 756, 165
- Parsons A. R. et al., 2014, *ApJ*, 788, 106
- Parsons A. R., Liu A., Ali Z. S., Cheng C., 2016, *ApJ*, 820, 51
- Patra N., Subrahmanyan R., Raghunathan A., Udaya Shankar N., 2013, *Exp. Astron.*, 36, 319
- Patra N., Bray J., Ekers R., Roberts P., 2015, preprint ([arXiv:1502.05862](https://arxiv.org/abs/1502.05862))
- Perley R. A., Roser H.-J., Meisenheimer K., 1997, *A&A*, 328, 12
- Pober J. C. et al., 2013, *ApJ*, 768, L36
- Pober J. C. et al., 2014, *ApJ*, 782, 66
- Pober J. C. et al., 2015, *ApJ*, 809, 62
- Pober J. C. et al., 2016, *ApJ*, 819, 8
- Prabu T. et al., 2015, *Exp. Astron.*, 39, 73
- Pritchard J. R., Furlanetto S. R., 2006, *MNRAS*, 367, 1057
- Pritchard J. R., Furlanetto S. R., 2007, *MNRAS*, 376, 1680
- Rideout W., Coster A., 2006, *GPS Solut.*, 10, 219
- Rogers A. E. E., Bowman J. D., 2008, *AJ*, 136, 641
- Rufenach C. L., 1972, *J. Geophys. Res.*, 77, 4761
- Rybicki G. B., Lightman A. P., 1979, *Radiative Processes in Astrophysics*. Wiley-Interscience, New York
- Santos M. G., Amblard A., Pritchard J., Trac H., Cen R., Cooray A., 2008, *ApJ*, 689, 1
- Singleton D., 1974, *J. Atmos. Terr. Phys.*, 36, 113
- Sitwell M., Mesinger A., Ma Y.-Z., Sigurdson K., 2014, *MNRAS*, 438, 2664
- Slee O. B., 1995, *Aust. J. Phys.*, 48, 143
- Sokolowski M. et al., 2015, *PASA*, 32, 4
- Sullivan I. S. et al., 2012, *ApJ*, 759, 17
- Tegmark M., 1997a, *Phys. Rev. D*, 55, 5895
- Tegmark M., 1997b, *ApJ*, 480, L87
- Thompson A. R., Moran J. M., Swenson G. W., 1986, *Interferometry and Synthesis in Radio Astronomy*. Wiley-Interscience, New York
- Thyagarajan N. et al., 2013, *ApJ*, 776, 6
- Thyagarajan N. et al., 2015a, *ApJ*, 804, 14
- Thyagarajan N. et al., 2015b, *ApJ*, 807, L28
- Thyagarajan N. et al., 2016, preprint ([arXiv:1603.08958](https://arxiv.org/abs/1603.08958))
- Tingay S. J. et al., 2013a, *PASA*, 30, 7
- Tingay S. J. et al., 2013b, *J. Phys. Conf. Ser.*, 440, 012033
- Titheridge J. E., 1973, *J. Atmos. Terr. Phys.*, 35, 981
- Tremblay S. E. et al., 2015, *PASA*, 32, 5
- Trott C. M., Tingay S. J., 2015, *ApJ*, 814, 27
- Trott C., Wayth R., 2016, *PASA*, 33, e019
- Trott C. M., Wayth R. B., Tingay S. J., 2012, *ApJ*, 757, 101
- Trott C. M. et al., 2016, *ApJ*, 818, 139
- Valdés M., Evoli C., Mesinger A., Ferrara A., Yoshida N., 2013, *MNRAS*, 429, 1705
- van Haarlem M. P. et al., 2013, *A&A*, 556, A2
- Vedantham H. K., Koopmans L. V. E., 2015, *MNRAS*, 453, 925 (V15)
- Vedantham H. K., Koopmans L. V. E., 2016, *MNRAS*, 458, 3099 (V16)
- Vedantham H., Udaya Shankar N., Subrahmanyan R., 2012, *ApJ*, 745, 176
- Voytek T. C., Natarajan A., Jáuregui García J. M., Peterson J. B., López-Cruz O., 2014, *ApJ*, 782, L9
- Wayth R. B. et al., 2015, *PASA*, 32, 25
- Wieringa M. H., 1992, *Exp. Astron.*, 2, 203
- Yeh K. C., Flaherty B. J., 1966, *J. Geophys. Res.*, 71, 4557
- Zheng H. et al., 2014, *MNRAS*, 445, 1084

APPENDIX A: THE EFFECT OF CABLE REFLECTIONS ON TILE GAINS

Throughout this work, we reference several expressions dealing with standing waves on cables that arise from mismatched impedances at their connections. In this section we derive these expressions for the reader's convenience. Discussions of this problem can be found in most elementary electricity and magnetism texts.

An voltage signal, $A(x, t)$ incident on the end of a transmission line with impedance Z_0 and length L that is terminated by some resistance R_L will be partially reflected $B(x, t)$ and transmitted $C(x, t)$. The amplitudes of the reflected and transmitted components can be found by enforcing continuity in the voltage across the connection and are given by

$$B(L, t) = \frac{Z_0 - R_L}{Z_0 + R_L} A(L, t) \equiv \tilde{R}A(L, t) \quad (\text{A1})$$

$$C(L, t) = \frac{2Z_0}{Z_0 + R_L} A(L, T) \equiv \tilde{T}A(L, t) \quad (\text{A2})$$

The impedance of a length L coaxial line is given by

$$Z_0 = R_0 + i \left(2\pi f \ell_0 L - \frac{1}{2\pi f c_0 L} \right) \quad (\text{A3})$$

where c_0 is the capacitance per unit length and ℓ_0 is the inductance per unit length. A ubiquitous undergraduate electricity and magnetism exercise involves finding these quantities for a coaxial cable filled with a dielectric of permittivity ϵ and permeability μ (Griffiths 2013), yielding

$$c_0 = \frac{2\pi\epsilon}{\ln \frac{d_o}{d_i}} \quad (\text{A4})$$

and

$$\ell_0 = \frac{\mu}{2\pi} \ln \frac{d_o}{d_i}. \quad (\text{A5})$$

Here d_i is the radius of the inner wire of the coaxial cable and d_o is the radius of the outer shell. It is clear from equation (A3) that the reflection coefficients are dependent on frequency in a way that is influenced by the cable geometry and dielectric properties.

Now we consider the coaxial cable terminated on both ends with reflection coefficients \tilde{R}_0 and \tilde{R}_L . A monochromatic voltage signal with frequency f entering the cable at $x = 0$ with amplitude $s(f)$ will travel to the end of the cable ($x = L$) where part of it will be

transmitted and the other part reflected. The complex amplitude of the transmitted component is $\tilde{T}_L(f)s(f)e^{\pi i \tau f}$ while the reflected component has complex amplitude $\tilde{R}_L s(f)e^{\pi i \tau f}$, where τ is the time it takes for the signal to propagate down the length of the cable and back. The reflected component will travel back down to $x = 0$ and be re-reflected and transmitted with an amplitude of $\tilde{T}_L(f)\tilde{R}_0\tilde{R}_L s(f)e^{3\pi i \tau f}$. We may compute the total output at $x = L$ as a series of transmitted waves where the n th summand has gone through n partial reflections,

$$s_{eff}(f) = \tilde{T}_0 e^{\pi i \tau f} \sum_{n=0}^{\infty} \left(\tilde{R}_0 \tilde{R}_L e^{2\pi i \tau f} \right)^n \quad (\text{A6})$$

$$= s_{eff}(f) \tilde{T}_0 e^{\pi i \tau} \frac{1}{1 - \tilde{R}_0 \tilde{R}_L e^{2\pi i \tau f}}. \quad (\text{A7})$$

The term $\tilde{T}_0 e^{\pi i \tau f}$ has a phase and amplitude that evolves gradually with frequency so we may treat it as part of a smooth complex gain $g(f)$ which will include the contributions from all other steps in the signal path. The gain of the tile in the presence of reflections becomes

$$g(f) \rightarrow g(f)' = g(f) \frac{1}{1 - r e^{i(2\pi\tau + \phi)}} \quad (\text{A8})$$

where $r e^{i\phi} = \tilde{R}_0 \tilde{R}_L$, both terms potentially evolving with frequency.

APPENDIX B: THE POWER SPECTRUM OF IONOSPHERIC PHASE FLUCTUATIONS FROM MEASUREMENTS OF DIFFERENTIAL REFRACTION

In this section, we derive the relationship between the structure function of source offsets and the underlying power spectrum of ionospheric phase fluctuations. We will adopt the common assumption that the TEC above the MWA, and hence the phases added to transiting electromagnetic waves are described approximately by a Gaussian random field (Rufenach 1972; Singleton 1974) whose power spectrum we denote as $P(\mathbf{k})$. In Section 3.2, we measure the differential refraction of source positions which we may express in terms of the gradients of the phase screen.

$$\begin{aligned} D(\theta) &= 2 \left(\frac{c}{2\pi f} \right)^2 \langle \nabla\phi(\mathbf{r}_0) \cdot \nabla\phi(\mathbf{r}_0) - \nabla\phi(\mathbf{r}_0) \cdot \nabla\phi^*(\mathbf{r}_0 + \mathbf{r}) \rangle \\ &= 2 \left(\frac{c}{2\pi f} \right)^2 [\rho_{\nabla}(0) - \rho_{\nabla}(\mathbf{r})], \end{aligned} \quad (\text{B1})$$

where $\rho_{\nabla}(\mathbf{r})$ is the correlation function of the ionospheric gradients. We can write $\rho_{\nabla}(\mathbf{r})$ in terms of the power spectrum by expanding $\nabla\phi(\mathbf{r})$ in terms of its Fourier components

$$\nabla\phi(\mathbf{r}) = \frac{i}{(2\pi)^2} \int d^2\mathbf{k} \tilde{\phi}(\mathbf{k}) \mathbf{k} e^{i\mathbf{k}\cdot\mathbf{r}} \quad (\text{B2})$$

Hence,

$$\rho_{\nabla}(\mathbf{r}) = \frac{1}{(2\pi)^2} \int d^2\mathbf{k} k^2 e^{-i\mathbf{k}\cdot\mathbf{r}} P(\mathbf{k}), \quad (\text{B3})$$

where we have used the definition of the power spectrum,

$$\langle \tilde{\phi}(\mathbf{k}) \tilde{\phi}(\mathbf{k}') \rangle = (2\pi)^2 P(\mathbf{k}) \delta_D^{(2)}(\mathbf{k} - \mathbf{k}'). \quad (\text{B4})$$

If we assume isotropy of the field, we have

$$\rho_{\nabla}(r) = \frac{1}{2\pi} \int dk k^3 P(k) J_0(kr). \quad (\text{B5})$$

Thus, by measuring the structure function of source offsets, we effectively measure the power spectrum of the ionospheric fluctuations.

APPENDIX C: THE AMPLITUDE OF SCINTILLATION NOISE IN MWA VISIBILITIES

In this section, we estimate the amplitude of scintillation noise present in each of the 2 s time steps that we interleave to estimate the system temperature. The time between the interleaved steps used to compute our system temperature is smaller than the coherence time given in V15. However, computing the amplitude of scintillation noise, assuming that it is entirely decorrelated between our 2 s time steps allows us to place an upper limit on what systematic bias in T_{sys} that might arise. We estimate the level of scintillation noise a baseline with length \mathbf{b} arising from a source population with a number density per solid angle and intrinsic flux bin given by,

$$\frac{d^2 N(S_i, f)}{dS_i d\Omega} = C S_i^{-\alpha} f^{-\beta}, \quad (\text{C1})$$

using equation (2.7) in V16

$$\sigma_{\text{scint}}^2[V(\mathbf{b})] = 4 S_{\text{eff}}^2 \int d^2\mathbf{k} P(\mathbf{k}) \sin^2(\pi\lambda h \mathbf{k}^2 - \pi\mathbf{b} \cdot \mathbf{q})$$

$$S_{\text{eff}}^2 \approx \frac{C B_{\text{eff}}(f) f^{-\beta}}{3 - \alpha} S_{\text{max}}^{3-\alpha}. \quad (\text{C2})$$

Here $P(\mathbf{k})$ is the power spectrum of ionospheric phase fluctuations and S_{max} is the maximal apparent source flux for which ionospheric effects have not been calibrated out. B_{eff} is the effective primary beam of the instrument and can be computed from the equation

$$B_{eff}(f) = \int d\Omega B^{\alpha-1}(f, \ell) \quad (\text{C3})$$

where $B(f, \ell)$ is the antenna primary beam. This equation is derived assuming a small field of view $\lesssim 10^\circ$. However, V16 find that it is accurate to within ≈ 10 per cent for substantially larger fields such as the MWA's.

Since we do not attempt to calibrate out the fluctuations on 2 s intervals, this source flux can be obtained by setting the number of sources in the field of view of the instrument with fluxes equal to S_{max} to one (V15),

$$S_{\text{max}}(f) = \left(\frac{\alpha - 1}{C f^{-\beta} B_{\text{eff}}(f)} \right)^{1/(1-\alpha)} \quad (\text{C4})$$

For our source population, we use fits by Di Matteo et al. (2002) to the source counts observed in the 6C survey (Hales, Baldwin & Warner 1988) at 151 MHz,

$$\frac{d^2 N}{dS_i d\Omega}(S_i, f_0 = 150 \text{ MHz}) = k \left(\frac{S_i}{0.880 \text{ Jy}} \right)^{-\gamma} \quad (\text{C5})$$

where $k = 4000 \text{ sr}^{-1} \text{ Jy}^{-1}$ and $\gamma = 2.5$. Assuming that all of the sources have a spectral index close to the observed mean of $\delta = 0.8$, we determine the frequency dependence of the source counts by setting the number of sources with fluxes above flux S_i at $f_0 = 150 \text{ MHz}$ equal to the number of sources at f with fluxes greater than $S'_i = S_i (f/f_0)^{-\delta}$. Doing this, we obtain

$$\frac{d^2 N(S_i, f)}{dS_i d\Omega} = k \left(\frac{f}{f_0} \right)^{\delta(1-\gamma)} S_i^{-\gamma} \quad (\text{C6})$$

which is similar to the expression in Trott & Tingay (2015) except for an order-unity difference in the frequency power law which was neglected in that work since $f/f_0 \approx 1$ and here, where $f/f_0 \approx 1/2$ accounts for an ≈ 25 per cent enhancement in the source counts.

We substitute $\beta = \delta(\gamma - 1) = 1.2$, $C = kf_0^{-\delta(1-\gamma)} = 1.6 \times 10^6 \text{Jy}^{-1} \text{sr}^{-1} \text{MHz}^{-2.5}$, $\alpha = \gamma = 2.5$, and an effective beam area of C1. Using the short dipole model of the MWA beam, we compute a $B_{\text{eff}}(83 \text{ MHz})$ of 0.33 sr. From these numbers, we obtain $S_{\text{eff}} = 212.3 \text{ Jy}$.

The final ingredient is $P(\mathbf{k})$ which we compute from our fits of our differential refraction measurements described in appendix B and using the functional form in equation (19). Applying equation (C2), we obtain values for $\sigma_{\text{scint}}[V(\mathbf{b})]$ between ≈ 4 and 6 Jy on the 30 min intervals on September 5 and ≈ 1 –2 Jy on the 30 min intervals on September 6 at 83 MHz.

We estimate the noise on a single antenna for each 2 s interleaved time interval is given by (Morales & Hewitt 2004)

$$\sigma_v = \frac{k_b T_{\text{sys}}}{A_e \sqrt{2 df \tau}} \quad (\text{C7})$$

where df is the channel width, A_e is the effective area of the tile, and $T_{\text{sys}} \approx T_{\text{sky}} = 60(\lambda/\text{meter})^{-2.6} \text{ K}$ (Rogers & Bowman 2008; Fixsen et al. 2011) which dominates the MWA's system temperature at lower frequency. Using $df = 80 \text{ kHz}$ and $\tau = 2 \text{ s}$, we obtain $\sigma_v \approx 315 \text{ K}$. Hence, on a single 2 s integration for each of our visibilities, scintillation noise contributes at the level of $\lesssim 2$ per cent relative to the system noise during the most severe times and $\lesssim 0.3$ per cent during the calmest intervals.

¹ Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

² MIT Kavli Institute for Astrophysics and Space Research, Cambridge, MA 02139, USA

³ Department of Astronomy and Berkeley Center for Cosmological Physics, University of California, Berkeley, Berkeley, CA 94720, USA

⁴ Harvard–Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA

⁵ Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy

⁶ Netherlands Institute for Radio Astronomy (ASTRON), PO Box 2, NL-7990 AA Dwingeloo, The Netherlands

⁷ ARC Centre of Excellence for All-sky Astrophysics (CAASTRO), Redfern 2016, Australia

⁸ Department of Physics, University of Washington, Seattle, WA 98195, USA

⁹ Arizona State University, School of Earth and Space Exploration, Tempe, AZ 85287, USA

¹⁰ Square Kilometre Array South Africa (SKA SA), Park Road, Pinelands 7405, South Africa

¹¹ Department of Physics and Electronics, Rhodes University, Grahamstown 6140, South Africa

¹² Australian National University, Research School of Astronomy and Astrophysics, Canberra, ACT 2611, Australia

¹³ MIT Haystack Observatory, Westford, MA 01886, USA

¹⁴ International Centre for Radio Astronomy Research, Curtin University, Perth, WA 6845, Australia

¹⁵ Sydney Institute for Astronomy, School of Physics, The University of Sydney, NSW 2006, Australia

¹⁶ Dunlap Institute for Astronomy and Astrophysics, University of Toronto, ON M5S 3H4, Canada

¹⁷ eScience Institute, University of Washington, Seattle, WA 98195, USA

¹⁸ Victoria University of Wellington, School of Chemical and Physical Sciences, Wellington 6140, New Zealand

¹⁹ University of Wisconsin–Milwaukee, Department of Physics, Milwaukee, WI 53201, USA

²⁰ Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, MI 48109, USA

²¹ School of Physics, The University of Melbourne, Parkville, VIC 3010, Australia

²² CSIRO Astronomy and Space Science (CASS), PO Box 76, Epping, NSW 1710, Australia

²³ National Centre for Radio Astrophysics, Tata Institute for Fundamental Research, Pune 411007, India

²⁴ Raman Research Institute, Bangalore 560080, India

²⁵ Department of Physics, Brown University, Providence, RI 02912, USA

²⁶ National Radio Astronomy Observatory, Charlottesville and Greenbank, Charlottesville 22903, USA

²⁷ International Centre for Radio Astronomy Research, University of Western Australia, Crawley, WA 6009, Australia

This paper has been typeset from a \LaTeX file prepared by the author.