



Improving the Epoch of Reionization Power Spectrum Results from Murchison Widefield Array Season 1 Observations

N. Barry^{1,2}, M. Wilensky³, C. M. Trott^{2,4}, B. Pindor^{1,2}, A. P. Beardsley⁵, B. J. Hazelton^{3,6}, I. S. Sullivan⁷, M. F. Morales³, J. C. Pober⁸, J. Line^{2,4}, B. Greig^{1,2}, R. Byrne³, A. Lanman⁸, W. Li⁸, C. H. Jordan^{2,4}, R. C. Joseph^{2,4}, B. McKinley^{2,4}, M. Rahimi^{1,2}, S. Yoshiura⁹, J. D. Bowman⁵, B. M. Gaensler¹⁰, J. N. Hewitt¹¹, D. C. Jacobs⁵, D. A. Mitchell^{2,12}, N. Udaya Shankar¹³, S. K. Sethi¹³, R. Subrahmanyan^{2,13}, S. J. Tingay^{2,4}, R. L. Webster^{1,2}, and J. S. B. Wyithe^{1,2}

¹The University of Melbourne, School of Physics, Parkville, VIC 3010, Australia; nichole.barry@unimelb.edu.au

²ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

³University of Washington, Department of Physics, Seattle, WA 98195, USA

⁴International Centre for Radio Astronomy Research, Curtin University, Perth, WA 6845, Australia

⁵Arizona State University, School of Earth and Space Exploration, Tempe, AZ 85287, USA

⁶University of Washington, eScience Institute, Seattle, WA 98195, USA

⁷University of Washington, Department of Astronomy, Seattle, WA 98195, USA

⁸Brown University, Department of Physics, Providence, RI 02912, USA

⁹Faculty of Science, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan

¹⁰Dunlap Institute for Astronomy and Astrophysics, University of Toronto, ON M5S 3H4, Canada

¹¹MIT Kavli Institute for Astrophysics and Space Research, Cambridge, MA 02139, USA

¹²CSIRO Astronomy and Space Science (CASS), PO Box 76, Epping, NSW 1710, Australia

¹³Raman Research Institute, Bangalore 560080, India

Received 2019 March 14; revised 2019 August 6; accepted 2019 August 29; published 2019 October 7

Abstract

Measurements of 21 cm Epoch of Reionization (EoR) structure are subject to systematics originating from both the analysis and the observation conditions. Using 2013 data from the Murchison Widefield Array (MWA), we show the importance of mitigating both sources of contamination. A direct comparison between results from Beardsley et al. and our updated analysis demonstrates new precision techniques, lowering analysis systematics by a factor of 2.8 in power. We then further lower systematics by excising observations contaminated by ultra-faint RFI, reducing by an additional factor of 3.8 in power for the zenith pointing. With this enhanced analysis precision and newly developed RFI mitigation, we calculate a noise-dominated upper limit on the EoR structure of $\Delta^2 \leq 3.9 \times 10^3 \text{ mK}^2$ at $k = 0.20 \text{ h Mpc}^{-1}$ and $z = 7$ using 21 hr of data, improving previous MWA limits by almost an order of magnitude.

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

1. Introduction

The evolution of structure in the early universe holds significant value to our understanding of cosmology and astrophysics. However, it remains unobserved. Of particular interest is the Epoch of Reionization (EoR), the time period where near-uniform neutral hydrogen coalesced into stars and galaxies. Observations of extragalactic sources (e.g., Robertson et al. 2010) and the cosmic microwave background (CMB; e.g., Planck Collaboration et al. 2016) have put some constraints on the neutral hydrogen fraction, though not much is known beyond an approximate timing of when the EoR occurred.

A direct detection of the structure of neutral hydrogen, and how it evolves in time, is possible using the 21 cm hyperfine transition. The measured frequency corresponds to a line-of-sight distance due to the narrow emission width, and thus the structure of the EoR can be measured as a function of time. Many collaborations are seeking an EoR 21 cm detection, including the Giant Metrewave Radio Telescope (Paciga et al. 2011), LOw Frequency Array (LOFAR; Yatawatta et al. 2013; van Haarlem et al. 2013), Precision Array for Probing the Epoch of Reionization (PAPER; Parsons et al. 2010), Hydrogen Epoch of Reionization Array (Pober et al. 2014; DeBoer et al. 2017), and the Murchison Widefield Array (MWA; Bowman et al. 2013; Tingay et al. 2013; Wayth et al. 2018).

A recent sky-averaged detection of the early EoR using the Experiment to Detect the Global EoR Signature (EDGES; Bowman et al. 2018) revealed an absorption profile twice as bright as predicted, thereby suggesting new physics. While this detection does not probe structure, it sets a precedent for the potential impact of direct measurements of the EoR. So far, there have been no detections of the EoR structure due to the challenges associated with analysis and spectral accuracy.

Astrophysical foregrounds are many orders of magnitude brighter than the EoR signal. Therefore, our ability to measure EoR structure depends on how well we can separate the foregrounds from the signal in analysis. Power spectrum space naturally separates spectrally smooth foregrounds from the predicted spectral variation of the EoR signal while quickly averaging down the thermal noise.

Recent discoveries of analysis signal loss have changed the state of the field and reemphasized the importance of analysis pipeline efficacy and signal loss simulations (Cheng et al. 2018). As a result, the best published upper limits on EoR structure have remained close to $\sim 10^4 \text{ mK}^2$ (Paciga et al. 2013; Dillon et al. 2015; Beardsley et al. 2016; Patil et al. 2017), with the most recent competitive MWA limits as $\Delta^2 \leq 2.7 \times 10^4 \text{ mK}^2$ at $k = 0.27 \text{ h Mpc}^{-1}$ and $z = 7.1$ published by Beardsley et al. (2016).

In this work, we first improve on the best MWA limits through pure analysis techniques alone. We begin with the identical 32 hr data set and the same code packages as Beardsley et al. (2016), with the only difference between the analyses being our new precision techniques. This highlights the importance of our analysis pipeline, proving that the dominant systematic preventing lower limits was the spectral accuracy of our techniques. Then, we further improve upon this limit through additional systematic mitigation using the latest RFI-detection methodologies, reducing the data set to 21 hr.

We achieve a new EoR upper limit at $k = 0.20 h \text{ Mpc}^{-1}$ and $z = 7$ of $\Delta^2 \leq 3.9 \times 10^5 \text{ mK}^2$. This limit is systematic dominated, and would not benefit significantly from further integration. However, the zenith-pointing subset gives a similar limit, which is noise dominated, indicating pointing or beam-related errors as a dominant systematic in our analysis. We do not expect to be able to detect the EoR with less than a few hundred hours of integration (Beardsley et al. 2013), but a smaller integration is still capable of proving the viability of our analysis and foreground mitigation techniques.

Descriptions of the MWA instrument and data sets used in this work are given in Section 2. Then, we briefly summarize the major components of the Fast Holographic Deconvolution (FHD)/Error Propagated Power Spectrum with Interleaved Observed Noise (ϵ pssilon) data analysis pipeline in Section 3, focusing on recent improvements. In Section 4, we compare our updated analysis to Beardsley et al. (2016) by reducing the same data set. We further improve this limit in Section 5 by using the latest RFI-mitigation techniques to select a subset of the data, and we provide validation of this new limit in Section 6. Finally, we discuss potential future improvements in Section 7.

2. Observations and the MWA

To understand the most recent improvements to the EoR upper limit, we must first briefly describe the main instrumental aspects of the MWA. We also summarize the data preprocessing, focusing on our nominal RFI flagging, as well as detail the data sets used in this work.

2.1. The MWA

The MWA is a radio interferometer located on a designated radio-quiet site in Western Australia. It has a dense cluster of elements at the core, which allows for extra sensitivity on EoR modes, and it has a pseudo-random scatter of elements at longer baselines for imaging. Due to its design, there is an opportunity to utilize MWA-made catalogs for instrumental calibration and foreground subtraction in EoR analysis. Upgrades to the MWA have separated the scientific cases into two independent layouts (Wayth et al. 2018), however, all data reduced in this paper uses the original Phase I layout (Tingay et al. 2013).

The sky voltage is observed with 128 elements of 16 dual-polarization dipoles in a 4×4 layout over a ground screen. Analog delays are added to the signal path to point each dipole toward the target, and then each set of 16 dipoles is beamformed to form one element. The signal then travels to the receiver, where it is digitized in a first-stage coarse frequency channelizer. This introduces aliasing every 1.28 MHz which must be removed via flagging in the analysis pipeline (Prabu et al. 2015). Each 1.28 MHz channel is

multiplied by a quantized gain to achieve a flatter response and to avoid bit quantization errors. A bandwidth of 30.72 MHz is then selected and transported to the correlator, where a second-stage channelizer creates 10 kHz channels. The signals are then correlated and averaged to a specified time and frequency set, specifically 0.5 s and 40 kHz for a 112 s interval per observation in this work.

We observe a few targeted fields of the sky for EoR science. This is achieved through a “drift and shift” strategy, where the instrument is pointed toward the target field every 30 minutes using analog delays (Trott 2014). In between shifts, the pointing remains constant and each new pointing results in a different beam shape and sampling of the sky. For EoR science, we only reduce data from pointings near zenith to minimize the effect of beam modeling errors. We use two pointings before zenith to two pointings after zenith in this work, targeting the “EoR0” field at R.A. = 0.00^{h} , decl. = -27° . This field is naturally low in sky temperature due to minimal foreground emission, and thus a great candidate for EoR observing.

2.2. Preprocessing

We run a pre-pipeline package to generate input data files that are RFI-flagged and frequency/time averaged. This simultaneously reduces data size and mitigates adverse RFI effects. We use COTTER, which subsequently calls the flagging package AOFLAGGER¹⁴ (Offringa et al. 2015).

We input the 0.5 s and 40 kHz sampled data into AOFLAGGER, which performs RFI flagging at this highest resolution. Line-like RFI features are found via a summed frequency and time threshold method after the sky contribution is estimated in the visibilities (Offringa et al. 2010). Then, a time and frequency morphological technique is used to catch broadband and variable features (Offringa et al. 2012).

We perform further flagging based off of known MWA instrumental effects. In particular, we flag the first two seconds and the last four seconds of an observation to mitigate beamformer lag and potential for dropped integration blocks in some 2013 data, respectively. For frequency, we flag the first two and last two frequency channels per coarse band to avoid channelizer aliasing, as well as the center DC channel per coarse band to avoid adverse bit rounding errors.

Averaging is then performed to output 80 kHz frequency and 2 s time resolution in a standard UVFITS format.

2.3. Data Sets

A 32 hr data set of observations was used to calculate EoR limits for the MWA in Beardsley et al. (2016), and we will use the same data set to verify our pipeline improvements in Section 4. This data set measures approximately 167–197 MHz from 2013 August until November. The target field is EoR0, and only pointings near zenith are included.

These 1029 observations were originally chosen for their clean statistics; they passed multiple mitigation tests, including pointing-based cuts, window power ratio cuts, polarization difference cuts, and residual image rms cuts (see Beardsley et al. 2016 for quality control techniques). They represent the best data at the time for MWA EoR analysis.

In addition to our pipeline-verification data sets, we apply improved quality assurance techniques to remove additional

¹⁴ <https://sourceforge.net/p/aoflagger/wiki/Home/>

contaminated data from the original 1029 observations to generate our best EoR upper limits in Section 5. All data with signatures of digital TV, amounting to 311 observations, are completely removed. A further 40 observations are removed by cutting data with RFI occupancy fraction greater than 40%. This creates a subset of about 21 hr, or 678 observations, from the previous data set. A description of our improved RFI-detection techniques is provided in Section 5.1.

As another part of our pipeline verification, we compare a small data set analyzed with FHD/ ϵ psilon to an analysis from RTS/CHIPS (Mitchell et al. 2008; Trott et al. 2016) in Section 6.1. This small data set only consists of the zenith pointing of the 678-observation data set, amounting to approximately 5 hr.

3. The FHD/ ϵ psilon Pipeline

Accuracy in analysis is a crucial aspect to EoR power spectrum measurements. Every element of our data analysis, including calibration, foreground subtraction, observation integration, and power spectrum estimation is highly bespoke to the accuracy requirements necessary to detect the EoR.

There are two separate, open-source packages in our pipeline: FHD¹⁵ and ϵ psilon.¹⁶ FHD is an implementation of an efficient deconvolution algorithm (Sullivan et al. 2012), though now we use its precision analysis capabilities to calibrate and image each observation without deconvolution (Barry et al. 2019). ϵ psilon calculates the resulting power spectrum and noise estimates from an integration of many observations. This general setup shares similar aspects with other imaging pipelines (Paciga et al. 2011; Patil et al. 2014, 2017; Shaw et al. 2014, 2015; Dillon et al. 2015; Ewall-Wice et al. 2016; Jacobs et al. 2016; Trott et al. 2016). However, a key focus of FHD/ ϵ psilon is end-to-end error propagation for error estimation. This is relatively uncommon among power spectrum estimators, but it is an integral feature in both ϵ psilon and CHIPS (Trott et al. 2016).

In this section, we summarize the main computational tasks and outputs of the FHD/ ϵ psilon pipeline, highlighting the significant accuracy improvements we have made since Beardsley et al. (2016) and Jacobs et al. (2016). Specifically, our pipeline is an image-based analysis that generates a reconstructed power spectrum (Morales et al. 2019). The resulting errors we aim to reduce are thus specific to this style of analysis.

Figure 1 provides context for how our analysis packages fit into the data flow. For a full description of the tasks and outputs of the FHD/ ϵ psilon pipeline, please see Barry et al. (2019).

3.1. FHD

In brief, FHD calculates calibrated images from measured visibility data. Various transformations and assumptions must take place to achieve these results, therefore the narrative of our data reduction is that of accuracy and precision. We summarize the main steps in the analysis, focusing on recent improvements.

First, we estimate the dipole response using an image-space beam from simulated Jones matrices for each frequency. Then, we build the response of the beamformed element and

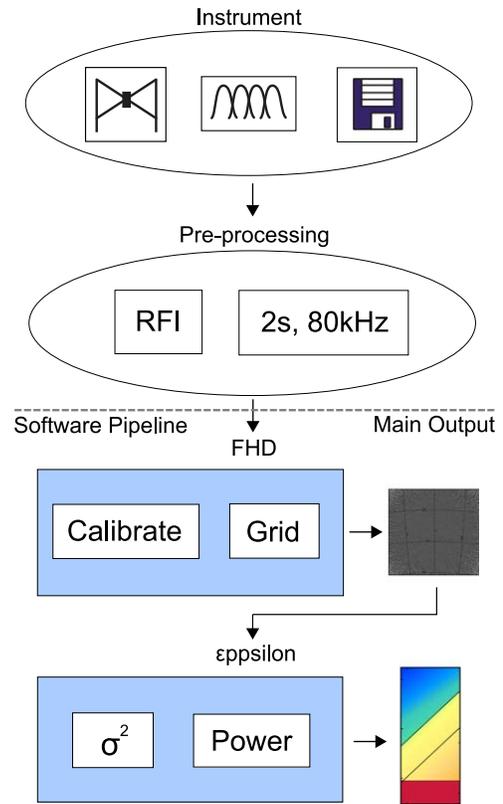


Figure 1. Data flow from measurement to power spectrum. There are four main blocks: (1) measurement, electronics, and raw data output from the instrument (Section 2.1), (2) RFI detection and averaging during preprocessing (Section 2.2), (3) calibration, gridding, and image creation during the FHD analysis (Section 3.1), and (4) error propagation, power spectrum estimation, and binning during the ϵ psilon analysis (Section 3.2).

transform it into the $\{u, v, f\}$ -domain, or the space of the visibility measurements. However, there is an accuracy limitation to simulations which model the Jones matrices at low dB. To account for this, we build a mask in uv -space to cut at this low dB level, and we renormalize the beam such that the edge smoothly goes to zero. This recent improvement ensures a smoother beam in Fourier space, and reduces systematics in sensitive EoR modes.

This beam kernel, or uv -beam response, is as instrumentally accurate and smooth as possible, and thus will be used in calibration. However, we now also use a modified gridding kernel in power spectrum estimation, analogous to the concept of a Tapered Gridded Estimator (Choudhuri et al. 2014, 2016). We apply the square of a Blackman-Harris window to the beam image. This acts as an image-space weighting to the observation when normalizations are properly taken into account. However, this creates correlations between image pixels; we investigate the importance of this effect in the power spectrum in Section 6.3. In this work, we refer to the kernel matched to the instrument beam as the “instrumental gridding kernel” and the kernel incorporating the image domain window function as the “modified gridding kernel.”

Next, we build model visibilities for calibration. We generate a model uv -plane at half-wavelength resolution by performing a discrete Fourier transform (DFT) of the flux and position of over 10,000 sources to a specified set of grid points in uv -space. At every baseline location in uv -space, we multiply the model uv -plane by our instrumental gridding kernel and integrate the

¹⁵ <https://github.com/EoRImaging/FHD>

¹⁶ <https://github.com/EoRImaging/epsilon>

result. This generates model visibilities, or an estimate of what the instrument measured.

We are reaching levels of precision that are affected by finite sum integration errors, especially in the process of estimating model visibilities (Kerrigan et al. 2018). Since the modified gridding kernel is a convolution in uv -space, it smooths out these errors as a function of frequency. This has important consequences in power spectrum space, and thus we use the modified gridding kernel when making power spectra and the instrumental gridding kernel when calibrating. This results in two sets of model visibilities: one for calibration that is instrumentally accurate, and one for power spectra that is frequency-smooth.

The introduction of the GaLactic and Extragalactic All-sky Murchison Widefield Array (GLEAM) survey in recent years has greatly increased the number of point sources in our sky model (Hurley-Walker et al. 2017). As a result, this has improved the accuracy of our sky-based calibration and the dynamic range in the power spectrum. Since our absolute flux scale has changed with GLEAM, our absolute power spectrum normalization has changed as well. Future works that scientifically compare EoR upper limits will need to match absolute flux scales and incorporate flux scale uncertainties.

We then calibrate the measured visibilities using the model visibilities in a sky-based calibration. An alternating direction implicit method is used to minimize a χ -squared equation to determine the gains from each element, which are assumed to account for the difference between the data and model (Mitchell et al. 2008; Salvini & Wijnholds 2014). This necessitates several assumptions, including independence in time, elements, polarization, and frequency.

These assumptions have consequences in the power spectrum. In Beardsley et al. (2016), we employed a method to smooth the bandpass gains as a function of frequency to reduce spectral structure. In our current implementation, we use the auto-visibility to calculate the resulting amplitude per element. The auto-visibility do not measure any structure on the sky, and thus are not subject to the incomplete sky errors as a function of frequency in Barry et al. (2016). However, the noise statistics and the overall amplitude will not be the same as the cross-visibility. We continue to use the fit cross-phases with a cable reflection in calibration, as the auto-visibility do not contain any phase information. For a full characterization of our calibration, please see Barry et al. (2019).

Errors in the bit statistics can affect both the cross- and auto-visibility if there is bit over or undersaturation (Barry 2018). We have found that in 2013 data, the upper part of the band (~ 187.5 – 197 MHz) was adversely affected by the bit statistics. An improperly tuned channel gain resulted in bit oversaturation, leading to an uncorrectable bias in the data. We now completely flag this part of the band in 2013 data, which limits the usable redshifts in power spectrum measurements.

Once we have calibrated, we grid the data and model visibilities onto the uv -plane. Each visibility value per frequency is multiplied by our modified gridding kernel and pixelated onto a regular uv -grid. In addition, we also grid beam kernels of integrated value 1 and squared beam kernels of integrated value 1 to separate uv -planes. These two additional uv -planes are the sampling map and variance map, respectively, and will be used in ϵ ppsilong for power spectrum estimation and error propagation. In order to generate noise estimates in ϵ ppsilong, we split each uv -plane into two interleaved time steps.

Since we did not incorporate w -projection effects (Cornwell et al. 2008), our gridded uv -planes are not in a basis that is coherent across observations. Therefore, we choose to Fourier transform the sampling map, the variance map, and the unweighted data uv -planes separately into image space. We interpolate to HEALPix projection,¹⁷ which is a fixed basis and therefore integrable across observations.

When we use image space to integrate, we are subject to aliasing effects due to a limited extent in the image. Our modified gridding kernel acts as a window filter, which greatly reduces image aliasing effects. In order to retain similar effective area to our previous 560 deg^2 , we now image 8090 deg^2 .

3.2. ϵ ppsilong

Our power spectrum pipeline, ϵ ppsilong, calculates the data power, model power, residual power, observed noise, expected noise, and uncertainty estimates from integrated images. We determine these various data and noise products to build reliable upper limits.

First, ϵ ppsilong transforms each of the integrated images back into the $\{u, v, f\}$ -domain. This includes the separate interleaved time samples for the data, the sampling map, and the variance map for each polarization. Each data product is then weighted with the sampling map, and similarly the variance map is weighted with the square of the sampling map.

At this point, we choose to perform weighted sums and differences of the data to build both a propagated noise spectrum and a power spectrum simultaneously. Using our interleaved time samples, we calculate maximum-likelihood estimates of the mean and the noise by using the sampling-map-weighted variances as the new weights. This weighting scheme is also used to calculate maximum-likelihood uncertainty estimates. These uncertainty estimates are compared to T_{sys} calculations propagated directly from the time-interleaved visibility differences, and thus we have consistent uncertainty estimates for every quantity within the full FHD/ ϵ ppsilong pipeline.

To continue onto power spectra, we must transform into the $\{k_x, k_y, k_z\}$ -domain. While u and v are easily converted into k_x and k_y via cosmological parameters, the transform from f into k_z is nuanced. Our frequency sampling is irregular: baselines move as a function of frequency and we have flagged frequencies due to RFI and instrumental systematics. Therefore, we perform a Lomb–Scargle periodogram (Lomb 1976; Scargle 1982) to find an orthogonal basis. The phase is not preserved in the Lomb–Scargle, thus restricting its use to power spectrum calculations and destroying information useful to the bispectrum (Bharadwaj & Pandey 2005), which may prove to be a vital statistic in the future (Majumdar et al. 2018; Trott et al. 2019; Watkinson et al. 2019).

Immediately before we perform the frequency transform, we first remove the mean of the amplitude. This reduces contamination from the bright, intrinsic foregrounds coupling into higher k_z -modes during the transform. We add this term back in as the DC component in the $\{k_x, k_y, k_z\}$ -domain to preserve power. This new average-removal method improves the distribution of power in 2D and 1D power spectrum compared to previous analyses.

We now construct our cross-power estimation from the maximum-likelihood mean and noise of the 3D $\{k_x, k_y, k_z\}$ -cube.

¹⁷ HEALPix: the Hierarchical Equal Area isoLatitude Pixelization of a sphere (Górski et al. 2005).

The power of the mean minus the power of the noise, divided by four, gives the same power estimation as the cross power between the interleaved time samples (Barry et al. 2019).

Since we choose to form the cross power after transformations, we are able to easily propagate our noise throughout ϵ ppsiilon. We assume negligible cross correlation between pixels, hence the propagation is simple sum/difference error propagation. We compare this propagated expected noise to the observed noise calculated from even-odd differences in Section 6.3 to test this assumption. This is particularly important since increasing the integrated image area increases the uv -resolution.

In order to calculate upper limits and related products from the 3D $\{k_x, k_y, k_z\}$ -cube, we calculate maximum-likelihood weighted averages of the bins. To form 2D power spectra as a function of k -modes perpendicular to the line of sight, k_{\perp} , and k -modes parallel to the line of sight, k_{\parallel} , we average in cylindrical regions. To form 1D power spectra as a function of $|k|$, we average in spherical regions.

The final outputs of our FHD/ ϵ ppsiilon pipeline are 2D and 1D power spectra products of the calibrated data, model data, residual data, expected noise, observed noise, and error bars. These various products will be used to determine our upper limits in Section 4 and Section 5 and to provide evidence of signal preservation and proper error propagation in Section 6.

4. Direct Limit Comparison

In Beardsley et al. (2016), the FHD/ ϵ ppsiilon pipeline was used to reduce 32 hr of MWA data for an EoR upper limit. The data reduction was dominated by systematics, and thus larger integrations would not have benefited the analysis significantly. At the time, it was unknown whether the limiting systematics were related to the data or to the analysis.

Many improvements have been made to the FHD/ ϵ ppsiilon pipeline since then, as outlined in Section 3. By reducing the same data set, we can determine how significant these improvements are toward lowering the MWA EoR limit. This will also help identify the nature of the dominating systematics in Beardsley et al. (2016).

Our updated analysis completely flags the upper part of the band due to an improperly tuned channel gain. To make a direct comparison, we rebin the data in Beardsley et al. (2016) and in our updated analysis to include the range 168.555–187.275 MHz. We avoid frequency regions where the bit statistics indicate truncation or saturation, and are thus different between the auto-visibilitys and the cross-visibilitys. This includes the first coarse band (167.115–168.235 MHz) and the upper part of the band (187.595–197.675 MHz). In addition, we avoid one extra channel each near these ranges (168.395 and 187.435 MHz) since they have flagged contributions, which can have consequences in power spectrum space (Offringa et al. 2019). With the Blackman-Harris window applied, we have an effective bandwidth of 9.4 MHz at approximately redshift 7.

The calibration catalog, and thus the absolute flux scale, has changed between the two analyses. The previous catalog, KGS¹⁸ (Carroll et al. 2016), was matched to the flux density of the MWA Commissioning Survey (MWACS; Hurley-Walker et al. 2014). Their absolute flux scale is significantly higher due to

residual flux and primary beam errors compared to our current catalog, GLEAM (Hurley-Walker et al. 2017). In order to have a fair comparison, we need to apply a scale factor to the Beardsley et al. (2016) analysis. We compare the mean calibration amplitudes for the frequency range of interest, and find that the previous analysis was 28% brighter in power for the E–W polarization and 23% brighter in power for the N–S polarization. We scale down both the Beardsley et al. (2016) upper limits and thermal noise by this correction factor for our comparison.

In Beardsley et al. (2016), the integrated HEALPix image was limited to a square $\sim 20^\circ$ across, both to select the clean center of the antenna beam and to reduce computational costs. This hard image cut was found to cause aliasing in the uv -plane, and was replaced by the modified gridding kernel discussed in Section 3.1. This modified kernel performs the same task of selecting data from the center of the antenna beam, without the hard edge, but requires using an integrated HEALPix map that is over $10\times$ larger in area. This change in integration scheme results in different levels of contamination in the thermal noise, simulated in Section 6.3. Therefore, our noise is about a factor of 3 lower than Beardsley et al. (2016) for the same data set.

4.1. 2D Power Spectrum Comparison

The 2D power spectrum is a useful diagnostic due to the characteristic contamination regions, and thus we can draw useful conclusions via their comparison. We create 2D power spectra for three data products: the calibrated data, the subtraction model, and the residual data after foreground removal.

We apply the frequency mask to both data sets, and then perform cylindrical averaging in k_x and k_y to generate k_{\perp} , or k -modes perpendicular to the line of sight. In addition, the k_z -component is relabeled k_{\parallel} , or k -modes parallel to the line of sight.

Power contamination in the $\{k_{\perp}, k_{\parallel}\}$ -space occurs in distinctive regions. Foreground contamination is present at low k_{\parallel} since it does not vary quickly as a function of frequency. Instrument chromaticity also couples the foregrounds into higher k_{\parallel} -modes along a constant slope called the “foreground wedge” (Datta et al. 2010; Morales et al. 2012; Parsons et al. 2012; Trott et al. 2012; Vedantham et al. 2012; Hazelton et al. 2013; Pober et al. 2013; Thyagarajan et al. 2013; Liu et al. 2014). Above this region is the “EoR window,” and this is where we expect to be able to make our measurements. Horizontal contamination lines constant in k_{\parallel} are caused from regular flagging of channelizer aliasing (described in Section 2). For a full description of the 2D power spectrum from ϵ ppsiilon, please see Barry et al. (2019).

Figure 2 shows the diagnostic 2D power spectra of the calibrated data, the model, and the residual for the N–S polarization for both the Beardsley et al. (2016) data reduction and our updated data analysis. There are several key differences.

1. The power in the foreground wedge is less contained in the 2016 analysis; spectrally smooth foregrounds should contaminate the lowest k_{\parallel} -mode by orders of magnitude more than other modes. This is an indication that foreground power was coupled to higher k_{\parallel} -modes more

¹⁸ KATALOGSS, the KDD (Knowledge Discovery in Databases) Astrometry, Trueness, and Apparent Luminosity of Galaxies in Snapshot Surveys, abbreviated as KGS.

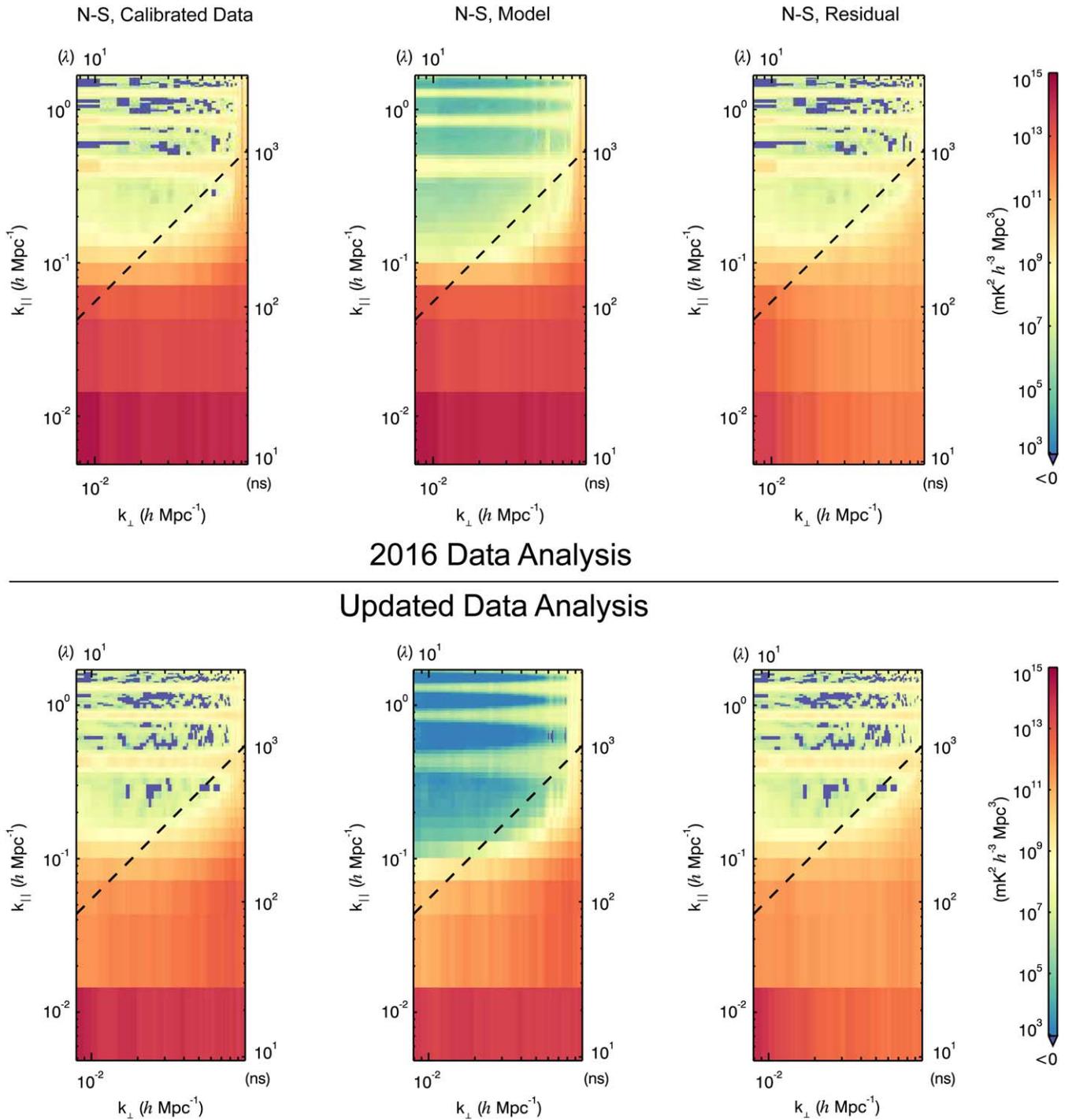


Figure 2. 2D power spectra comparison between the Beardsley et al. (2016) analysis (top row) and our updated analysis (bottom row) with the same observation data set and binning scheme. The calibrated data (left column), the subtraction model (middle column), and the residual (right column) 2D power spectra are shown for the N–S polarization. Our updated precision techniques and improved calibration reduce foreground coupling into the EoR window (above the dashed line).

than expected. We mitigate this in our updated analysis with improved calibration and the average-removal technique in power spectrum estimation.

2. The model power in the updated analysis is much lower in the EoR window. This is a strong indication that our analysis is more precise, and couples less foreground power into the sensitive measurement modes than before. Many improved techniques described in Section 3 contribute to this, including the modified gridding kernel.

3. Contamination in the EoR window for both the calibrated data and residual is reduced in the updated analysis, especially between the horizontal flagging harmonics. This is most likely due to higher precision in the model and updated calibration techniques using auto-visibility.

The 2D comparison in Figure 2 demonstrates significant improvement in sensitive measurement regions and in the expected distribution of foreground power. Considerable

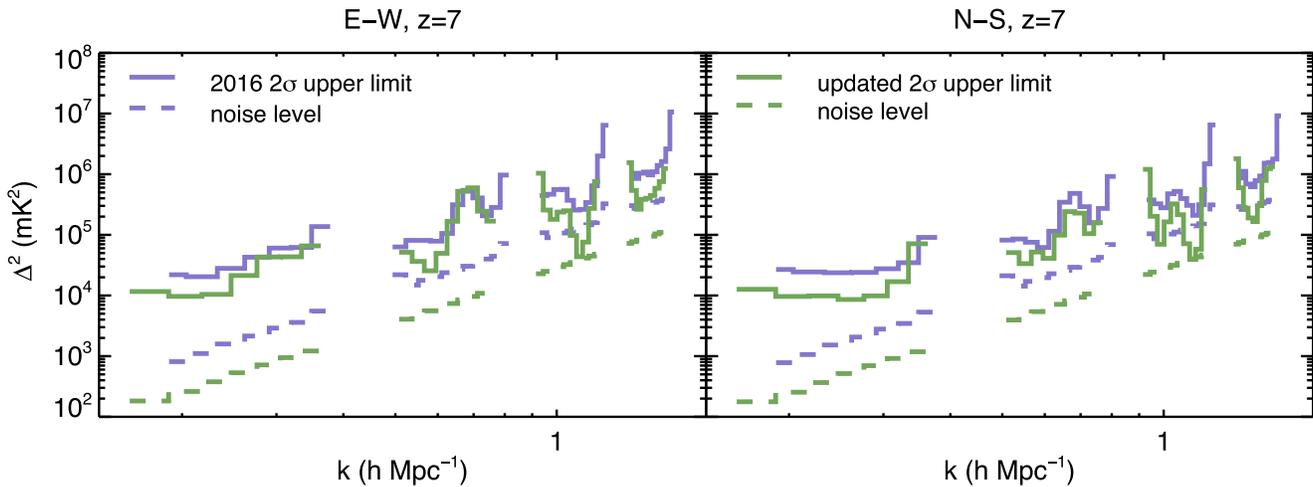


Figure 3. 1D EoR upper limit comparison between Beardsley et al. (2016) (purple) and our updated analysis (green) for the E–W and N–S polarizations at a band centered on redshift 7 for the same 1029-observation data set. The dashed lines are the thermal noise levels of each analysis. Our updated analysis has less power contamination on most k -modes.

advances have been made for the calibrated data and the model, resulting in an improved residual 2D power spectrum.

4.2. 1D Upper Limit Comparison

In order to show quantitative improvement, we calculate the 1D EoR upper limit between the two analyses. We use a similar mask and binning scheme to that presented in Beardsley et al. (2016) to have a direct comparison of the contamination levels in sensitive regions.

First, we exclude wavelengths in k_{\perp} , which are (1) greater than 70λ due to low uv -coverage and (2) less than 10λ due to increased contamination from coarse band harmonics. Then, we apply a k_{\parallel} -mask to all bins lower than $0.15 h \text{ Mpc}^{-1}$ to avoid a systematic floor from foreground leakage. Finally, we remove the foreground wedge with a small buffer; we exclude along the horizon line with an increased slope of 14% (Dillon et al. 2015). We also exclude k_{\perp} -bins associated with the coarse band harmonics, however, our analysis cannot exactly match Beardsley et al. (2016) due to a difference in resolution.

Figure 3 shows the comparison of the 1D EoR upper limits between the Beardsley et al. (2016) analysis and our updated analysis for the E–W and N–S polarizations at redshift 7. The limits are described as $\Delta^2(k) = k^3 P_{21}(k) / (2\pi^2)$ in units of mK^2 . Again, the only difference between the two approaches is the analysis; the raw data and the k -space binning scheme have remained the same. The updated analysis from this work has a lower EoR upper limit for most k -modes.

In particular, the most sensitive measurement modes at small k have lower EoR upper limits. The best mode in the N–S polarization from the rebinning of the Beardsley et al. (2016) analysis is $\Delta^2 \leq 2.37 \times 10^4 \text{ mK}^2$ at $k = 0.25 h \text{ Mpc}^{-1}$, while the best mode in the updated analysis is $\Delta^2 \leq 8.59 \times 10^3 \text{ mK}^2$. We have improved by a factor of 2.8. Likewise, for the best modes in the E–W polarization, we have improved by a factor of 2.1.

This is an indication that a major systematic in the Beardsley et al. (2016) analysis is related to the precision of the data reduction. There is a consistent improvement for most k -modes, suggesting that a contamination floor in the EoR window caused by foreground-coupled power has been mitigated by our new techniques. Figure 2 supports this conclusion; major

power reduction in the model 2D power spectra, especially in the EoR window, indicates a significant improvement in analysis precision.

By analyzing the same data set in Beardsley et al. (2016) with our updated pipeline, we directly compared how our new precision techniques affected the 2D and the 1D power spectrum. Now, we have a consistent narrative that the FHD/ ϵ ppsi pipeline has improved significantly, and that analysis systematics have been greatly reduced.

5. Updated Limit

Our improvements to the FHD/ ϵ ppsi pipeline (Section 3) lowered the previous MWA EoR upper limits significantly (Section 4). However, we can further lower these limits using new techniques to remove faint-RFI-contaminated observations. To calculate the best possible limits from 2013 data, we will choose a subset of the Beardsley et al. (2016) data set along with a new binning scheme. Using these 678 observations, we report an updated MWA EoR upper limit.

5.1. Data Selection

We use AOFLAGGER for primary RFI flagging in the FHD/ ϵ ppsi pipeline (Section 2.2). However, we are able to identify a substantial number of observations with leftover ultra-faint RFI contamination using SSINS¹⁹ (Wilensky et al. 2019).

SSINS operates by time-differencing visibilities to subtract out the slowly varying sky, and then averaging the amplitudes of these visibility differences over the set of baselines in the array. This leaves a single dynamic spectrum per polarization in which flagging can be performed. This is called the sky-subtracted incoherent noise spectrum (SSINS) of the observation. At the expense of more finely grained baseline-to-baseline information, this gives a dramatic sensitivity boost that allows identification of RFI well below the thermal noise levels of a single baseline. In addition, a match-shape filter is implemented within the SSINS framework to further boost sensitivity to known RFI contaminants such as digital television (DTV).

¹⁹ <https://github.com/mwilensky768/SSINS>

While SSINS can be used to flag time/frequency regions, we opted instead only to use it to catalog particular types and levels of contamination. This information was then used to cut out entire 112 s observations with residual faint RFI, rather than try to recover the observations. Two main RFI cuts were made: (1) any observations containing DTV signals identified by the SSINS match filter were removed, and (2) any observations where over 40% of the SSINS samples²⁰ were identified as contaminated by the SSINS flagging procedures were removed. These procedures identified 311 observations with DTV and a further 40 observations with high levels of RFI occupancy.

Figure 4 shows the residual 2D power spectra for the remaining 678 observations. We analyze the frequency band range 168.555–187.275 MHz (approximately redshift 7) to avoid known instrumental effects (Section 4). The general features of the 2D power spectrum from Figure 2 are still present: foregrounds are most prevalent in the lowest k_{\parallel} -mode and couple into the foreground wedge, channelizer flagging harmonics cause horizontal contamination lines, and there is lower power in the EoR window.

However, there are some key differences when comparing to the full 1029-observation integration in Figure 2. The EoR window power is generally lower between the channelizer flagging harmonics, especially in the N–S polarization. A notable exception is the power at the lower left-hand corner of the EoR window; it remains fairly contaminated in the N–S polarization. The systematic that causes this contamination is unknown.

There is also noteworthy differences between the polarizations in Figure 4. The E–W polarization tends to have higher power than the N–S polarization in the EoR window, which will have significant consequences in 1D EoR upper limits. The cause for this is not yet known.

By understanding the regions affected by contamination, we can make logical cuts in $\{k_{\perp}, k_{\parallel}\}$ -space. This will allow us to calculate 1D EoR upper limits that are free of known contaminated k -modes.

5.2. EoR Upper Limit

We can now perform informed $\{k_{\perp}, k_{\parallel}\}$ -cuts on the data to generate EoR upper limits. We avoid contaminated regions in Figure 4 by making the following selections:

- Spatial modes outside of $18\text{--}80 \lambda$ are masked. This avoids low k_{\perp} -modes that are contaminated in the EoR window, and it removes any potential effects from poor uv -coverage at high k_{\perp} -modes.
- We restrict line-of-sight modes to be greater than or equal to $0.15 h \text{ Mpc}^{-1}$. While the foreground wedge is mostly contained within the horizon line, there is some leakage at small k_{\perp} into the EoR window, which is avoided with this mask.
- We add a small buffer to the horizon slope to avoid sub-horizon leakage, which is especially prevalent in the E–W polarization. Our slope is 15% larger than the horizon, similar to the slopes used in Beardsley et al. (2016) and Dillon et al. (2015).

Contours in Figure 4 highlight the 2D region we use to generate 1D power spectra with these conditions.

²⁰ Given instrumental effects in the MWA, we did not include coarse band edges in this classification.

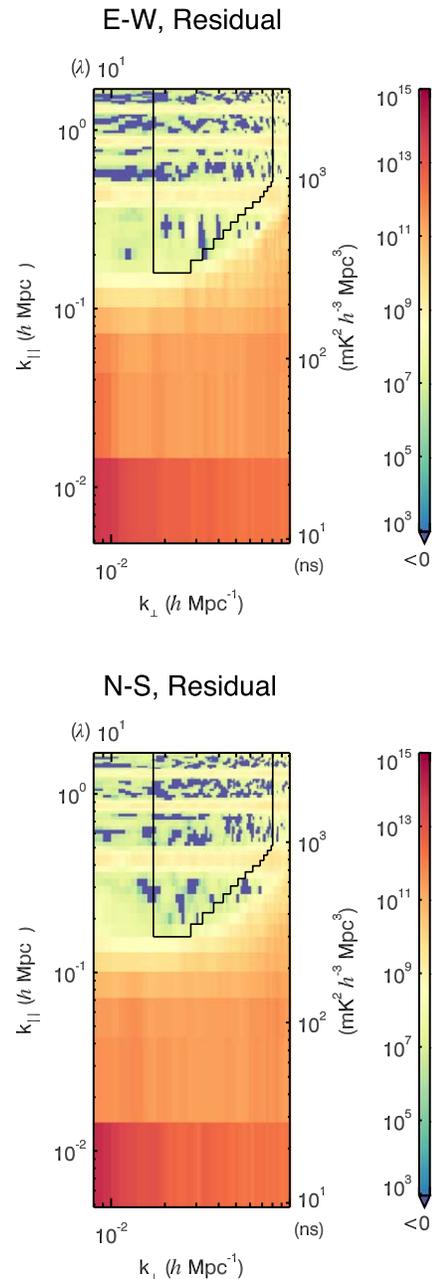


Figure 4. Residual 2D power spectra for the E–W and N–S polarizations for 678 observations selected with SSINS. Contours show the region where we make $\{k_{\parallel}, k_{\perp}\}$ selections to avoid known regions of contamination for 1D power spectra.

By choosing regions to integrate, we are potentially introducing a selection bias. This is unavoidable in a foreground-avoidance analysis; we must excise regions that we know are dominated by foregrounds in order to produce meaningful limits. To reduce the potential for selection bias, we only produce masks and cuts via foreground information from the 2D power spectrum. This lowers the degrees of freedom in our choice, and thus the potential bias.

Using these binning selections, we compare EoR upper limits from the full 1029-observation data set and the 678-observation subset selected with SSINS in the top panel of Figure 5. There is little to no change in the E–W polarization and a small improvement in the N–S polarization. However,

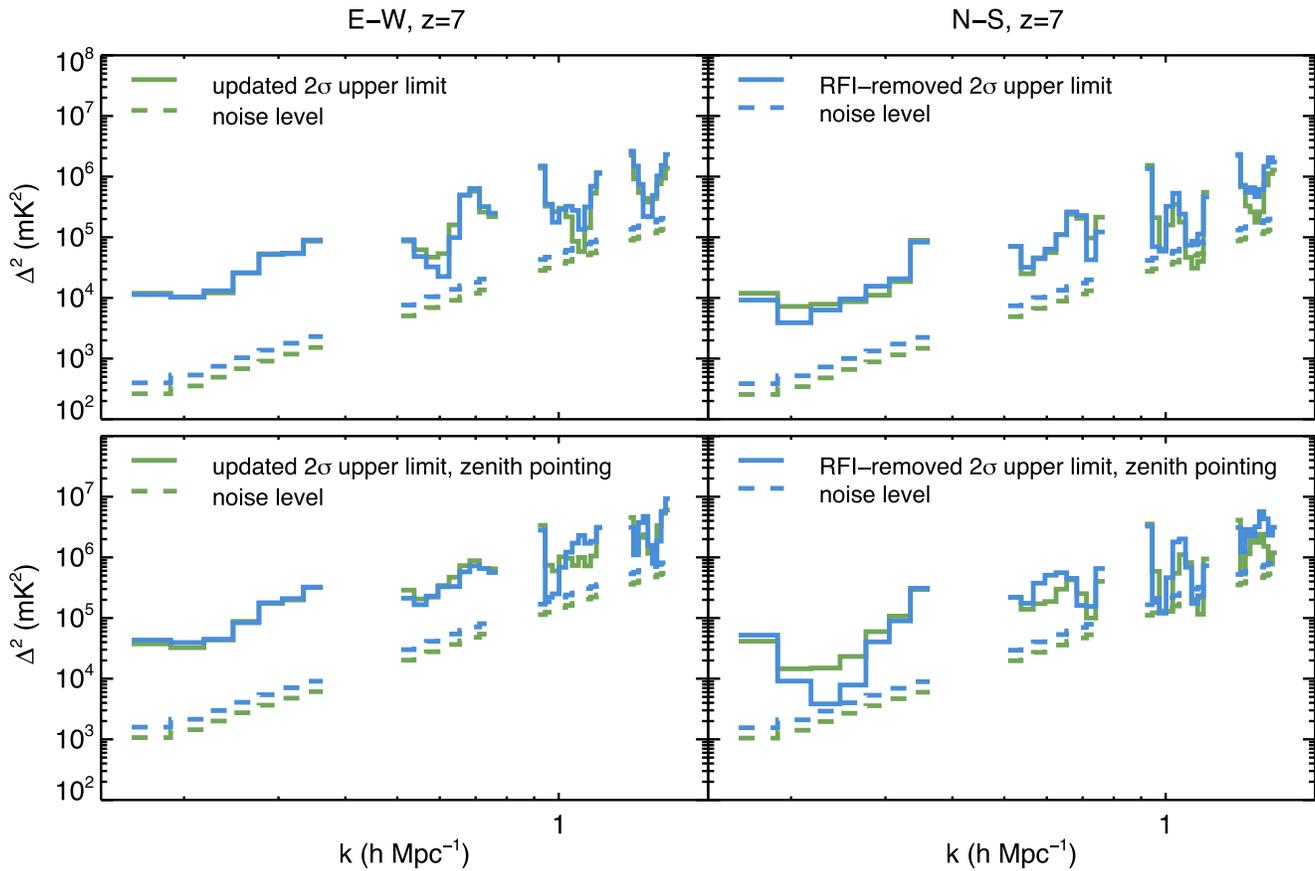


Figure 5. 1D EoR upper limit comparison between our updated analysis for the 1029-observation data set (green) and our updated analysis for the RFI-removed 678-observation data set (blue) for the E–W and N–S polarizations at a band centered on redshift 7. There is marginal improvement over the full integration (top panel), however, there is significant improvement over the zenith-pointing subset (bottom panel). The dashed lines are the thermal noise levels of each analysis.

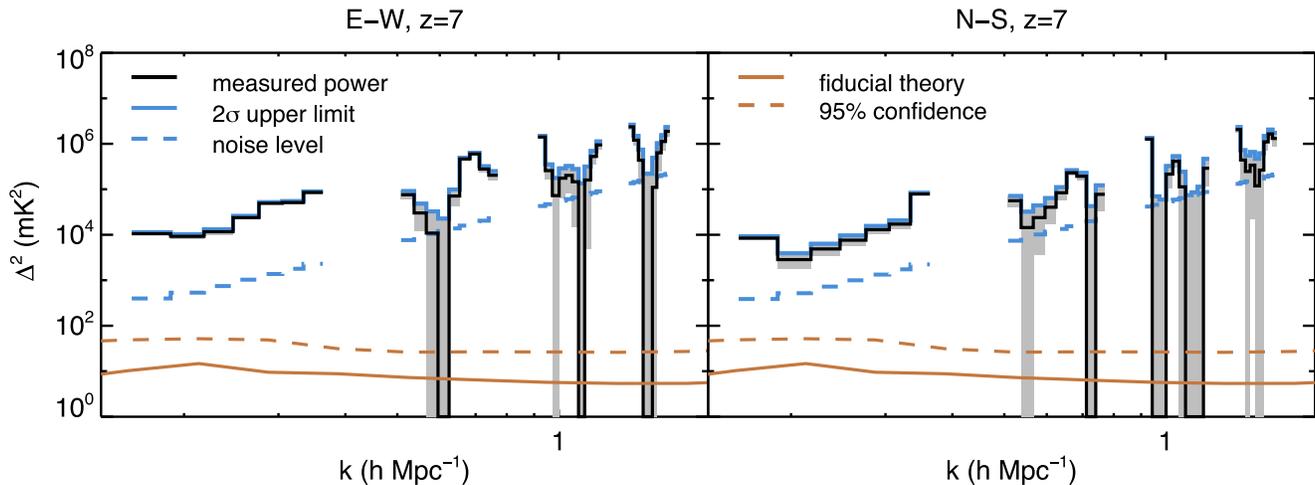


Figure 6. 1D measured power spectra (black), the 2σ error bars (gray), the 2σ EoR upper limits (solid blue), and the 1σ thermal noise levels (dashed blue) for the E–W and N–S polarizations using 678 observations selected with SSINS. We also present an example fiducial EoR theory power spectrum (solid brown) along with the theoretical 2σ upper limits on the 21 cm power spectrum amplitude (brown dashed) obtained using existing observational constraints (see Appendix A for further details). These constitute our best EoR upper limits in this work. We are noise dominated for many k -modes, including our lowest EoR upper limit.

the improvement via data selection from SSINS is highly dependent on pointing.

The bottom panel of Figure 5 shows the *zenith pointing* of both the 1029-observation data set and the 678-observation subset. In this case, removing additional RFI using SSINS improves the upper limit by a factor of 3.8 in the N–S

polarization. Other pointings do not benefit from further RFI flagging despite similar RFI-detection levels, which indicates dominating systematics due to beam errors.

We present the 1D measured power spectra, the 2σ error bars, the 2σ EoR upper limits, and the 1σ thermal noise levels in Figure 6, along with an example EoR fiducial theory and its

associated 2σ upper limits (see Appendix A for further details).

The measured power is the cross-power spectrum between the interleaved time samples (Section 3.2), and can fluctuate to be negative if below the thermal noise. The 1σ noise level is the estimated noise from the integration using the fully propagated uncertainty estimates. The corresponding 2σ error bars are shown for each power estimate, and can reach zero if the signal is consistent with a non-detection. The 2σ EoR upper limit is calculated from the measured power and the variances, where a prior of being greater than the thermal noise is enforced. We omit bins that are severely affected by the channelizer-aliasing flagging.

Our aim with this new integration is to be dominated by noise, not systematics. Since this is not enough data to theoretically detect the EoR, any detection is most likely that of a systematic in our data or analysis. Therefore, having a signal consistent with zero is ideal. We have many power estimates consistent with a non-detection, the majority of which are at higher k -modes. However, our lowest limits are systematic dominated. We can produce a similar, *noise-dominated* limit ($\Delta^2 \leq 3.8 \times 10^3 \text{ mK}^2$ at $k = 0.23 \text{ h Mpc}^{-1}$) using just the zenith pointing. Given the observing approach of the MWA, it is not practical to use just zenith-pointing measurements of a designated field to detect the EoR in a timely fashion, therefore future efforts will need to reduce this systematic.

There are features in Figure 6 that indicate other sources of contamination via systematics. First, the flagging from the channelizer aliasing causes foreground coupling on nearby k -modes. This indicates that a systematic floor might be present on *all* k -modes at a lower level than this analysis can achieve, and thus will need to be addressed in the future. We know that only five modes in Figure 6 *could* theoretically be below the 95% confidence limit of the fiducial EoR by analyzing the power in the model, most likely due to this systematic. In addition, known cable reflections contaminate certain modes: 0.4 h Mpc^{-1} (90 m), 0.7 h Mpc^{-1} (150 m), 1.0 h Mpc^{-1} (230 m), 1.4 h Mpc^{-1} (320 m), 1.7 h Mpc^{-1} (400 m), and 2.3 h Mpc^{-1} (524 m). These modes cannot be used for EoR detection (Barry et al. 2016), and may contaminate surrounding modes at a low level (Ewall-Wice et al. 2016).

The features present in the EoR upper limit, along with the features we cut via our mask and binning scheme, can help us determine future areas of improvement in our analysis and RFI-mitigation techniques (Section 7).

6. Data Analysis Validation

Now that we have a new EoR power spectrum upper limit, we present various procedures to prove its validity. Specifically, we compare results with another pipeline, perform a signal loss simulation, and present error propagation results from ϵ ppsiilon.

6.1. Pipeline Comparison

We cross-validate our FHD/ ϵ ppsiilon pipeline results with a different set of packages: the Real Time System (RTS; Mitchell et al. 2008; Ord et al. 2010) and CHIPS (Cosmological HI Power Spectrum, Trott et al. 2016) pipeline. We have extensively used this verification approach in both Jacobs et al. (2016) and Beardsley et al. (2016) to indicate proper treatment of normalization, polarization, and signal preservation. By

comparing power spectra from the same data set, we can demonstrate the continued robustness of our analysis pipelines.

The RTS calibrates raw data and produces visibilities, therefore it plays a similar role to FHD. However, the methodology is quite different. Specifically, the RTS can perform direction-dependent calibration via a peeling method. Approximately 1000 sources are used for a preliminary direction-independent calibration, and then five bright sources are used to estimate the specific calibration in the local region. Separate calibration values are calculated per 1.28 MHz (per coarse band). This constitutes the largest philosophical difference, but a variety of processes are distinct, including other calibration parameters, beam calculations, and techniques to enforce spectral smoothness. Given the importance of these features in the power spectrum space (Section 3.1), comparisons with the RTS are relevant for validation.

The RTS-calibrated visibilities are then processed by CHIPS, which starts by gridding the visibilities onto the $\{u, v, w\}$ -plane. By choosing discrete w -projection planes, CHIPS avoids the need to use image space for integrating observations. Therefore, many of the required aliasing mitigation techniques discussed in Section 3.1 are not necessary with this package. CHIPS also uses advanced systematic mitigation techniques, including an inverse-covariance weighting scheme during power estimation (Kay 1993). The noise calculation used for this comparison is a propagated noise from the even-odd difference, similar to the approach in ϵ ppsiilon.

We compare the zenith-pointing subset for our cross-validation in Figure 7. To remain consistent, we apply the same binning scheme in Section 5.2 to each data reduction. This foreground/systematic avoidance scheme was chosen with the FHD/ ϵ ppsiilon analysis in mind; they may not be the best cuts for RTS/CHIPS. Nevertheless, the EoR upper limits are roughly consistent with each analysis, and follow a general trend within the same order of magnitude for all k .

Since CHIPS uses an inverse-covariance weighting, the RTS/CHIPS analysis recovers more k -modes in the channelizer-aliasing harmonics. This technique also allows the RTS/CHIPS analysis to probe further within the foreground wedge, and thus their lowest limit is at their lowest k -mode. In contrast, the FHD/ ϵ ppsiilon analysis has lower systematics between the flagging harmonics.

In general, the two analyses produce similar EoR upper limits from the same data set. While 5 hr is not enough to create a competitive limit, we show continued consistency between two unique pipelines.

6.2. Pipeline Simulation

In addition to our analysis cross-validation with RTS/CHIPS, we simulate our pipeline end-to-end to test for signal loss. This demonstrates self-consistency and signal preservation throughout the analysis.

In order to validate FHD specifically, we run an in situ simulation. FHD is naturally an instrument simulator; we produce model visibilities that represent the response of the instrument to tens of thousands of point sources on the sky. These model visibilities, plus a theoretical EoR response, can be input into FHD as simulation data.

We run two pipeline-verification tests: (1) theoretical EoR signal through the entire pipeline with no calibration or subtraction and (2) theoretical EoR in a point-source sky where only a subset of the brightest point sources are subtracted with

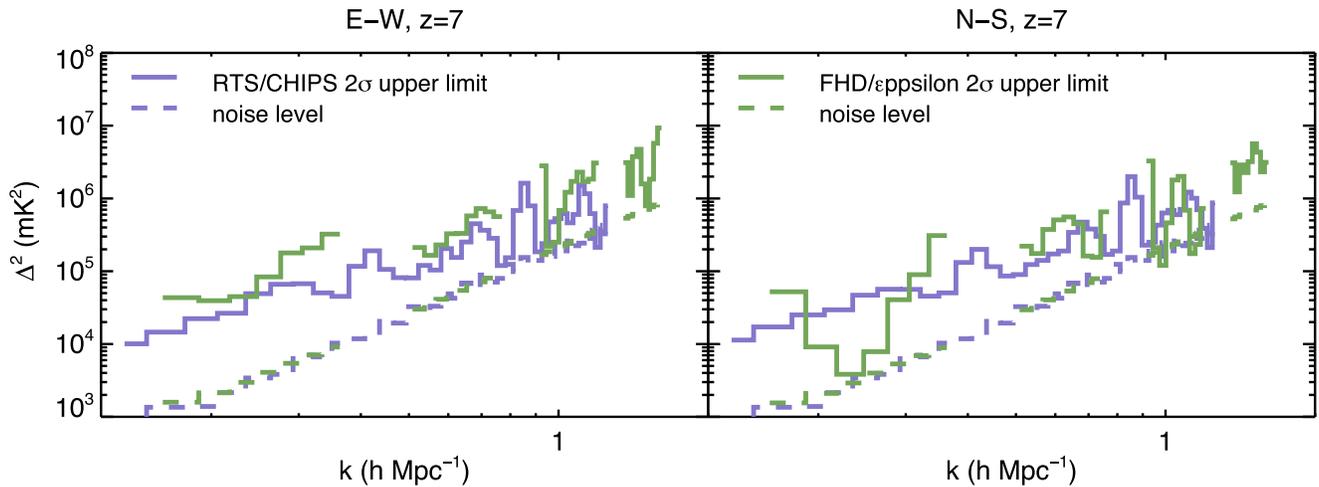


Figure 7. Cross-validation analysis of the EoR upper limits and associated noise levels on the zenith-pointing subset from the FHD/ ϵ pipeline (green) and the RTS/CHIPS pipeline (purple). RTS/CHIPS recovers more k -modes in known systematic-dominated regions via advanced techniques, while FHD/ ϵ pipeline produces lower systematics at some low k -modes. In general, their consistency with each other demonstrates robustness in our analysis techniques.

no calibration effects. These two tests investigate whether we recover the input EoR signal and if we can theoretically detect the EoR with realistic foregrounds disregarding calibration effects, respectively.

Figure 8 shows the measured power from these two simulations in 1D power spectrum space. We include the foreground wedge and the full bandwidth in this binning scheme. Our calculated power from the input EoR visibilities indicates that we do not suffer from signal loss or normalization errors; we recover the input power across all k -modes. When there are residual foregrounds in the simulation, we still recover the input EoR simulation in the EoR window for a large range of k -modes. Large k -modes (above $\sim 1.0 h \text{ Mpc}^{-1}$) are coupled to foreground contamination via our gridding kernel resolution (Beardsley et al. 2016).

Therefore, we have verified the ability of FHD to detect the EoR in ideal conditions without calibration effects for a wide range of k -modes. We report the results of these types of simulations whenever we reanalyze data with updated software, therefore similar simulations were performed in Barry et al. (2016, 2019) on the FHD/ ϵ pipeline.

6.3. Uncertainty Estimation Verification

We also conduct verification experiments within ϵ pipeline to test the validity of our noise assumptions. Given that our best EoR upper limit is noise dominated, it is crucial to investigate our uncertainty measurements.

First, we can determine the level of contamination in the noise due to choosing image space as our integration basis. While noiseless pipeline simulations have demonstrated that our measured power estimation recovers the EoR in Section 6.2, the noise can still be adversely affected by image-space aliasing.

We can calculate the noise contamination level by comparing the analysis of one observation using various integration schemes. By only analyzing one observation, it is not necessary to go to image space because integration is not required. Therefore, we have the capability to compare to noise that is not affected by the transform to image space.

In this work, we have two varieties of HEALPix integration schemes. Beardsley et al. (2016) used a square $\sim 20^\circ$ region,

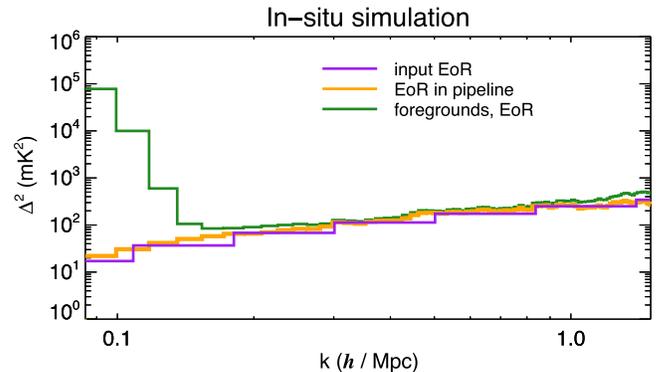


Figure 8. Measured 1D power for two in situ simulations on the FHD/ ϵ pipeline. We input simulated EoR visibilities (purple) into the pipeline and recover the expected power (orange). If we add foregrounds and only subtract a subset (green), we still recover the underlying EoR signal for most all k -modes.

while our updated analysis uses a region that includes more than $10\times$ the area and includes a modified gridding kernel. These two tests, along with the noise level calculated without an image-space transform, are shown in Figure 9.

The excess contamination from the image-space transform is flat across all k -modes. The integration scheme from Beardsley et al. (2016) is high by about a factor of 6, whereas our updated integration scheme is high by about a factor of 2. While this is an improvement, we need to mitigate this contamination in the future. We *do not* apply a correction factor in this work, therefore our EoR upper limit is higher than it could theoretically be for our analysis.

We also analytically calculate the error propagation throughout ϵ pipeline, thereby avoiding the ambiguity of bootstrapped errors. However, in order to make this manageable, we assume there are no cross-correlations between pixels in $\{u, v, f\}$ -space even though there are various stages in the analysis that could cause correlations. Increasing the integrated image size creates smaller uv -pixels, which are more likely to be correlated. Applying the modified gridding kernel and the frequency window will also correlate pixels.

To investigate the assumption of independent pixels in our analytic calculation, we also measure the standard deviation of the noise power created from the even-odd difference. These

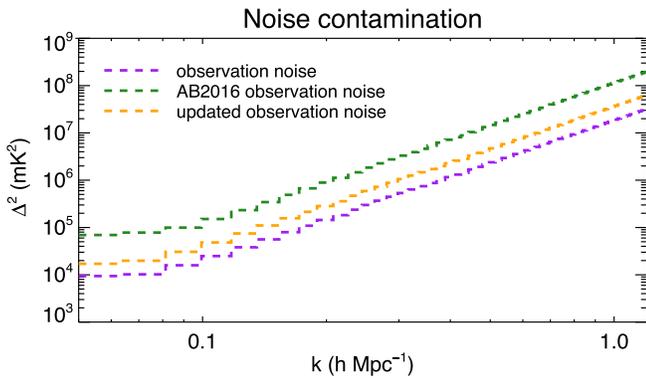


Figure 9. Calculated thermal noise levels for one observation using various integration schemes. Avoiding image space results in the lowest noise level (purple). Using a small image to match Beardsley et al. (2016) results in a factor of 6 contamination (green), and using our updated method results in a factor of 2 contamination (orange). Mitigating this effect is left for future work.

uncertainties are noisy; they will be subject to random noise variations per pixel. This gives us two different noise estimates: (1) an observed noise calculated from the power, and (2) an expected noise calculated from error propagation of the input cubes. The observed noise and expected noise should be the same magnitude if our error propagation and correlation assumptions are well-founded.

Figure 10 shows the noise and error 2D power spectra calculated from the 678-observation data set in Section 5.2. The observed noise (top right) and the expected noise (top left) are very similar, and their ratio (bottom right) is very close to 1. There is some deviation from 1 in poor uv -coverage regions, however these will not contribute to the 1D power spectrum given our binning schemes. The error bars (bottom left) are related to the expected noise.

By comparing the observed noise to the expected noise, we have investigated whether cross-correlations caused by our analysis techniques have a significant effect. A ratio close to unity indicates no excessive cross-correlations.

7. Discussion

Our new EoR upper limit with the open-source FHD/ ϵ ppsi pipeline is almost an order of magnitude better than previous data reductions, and there have been many contributing factors to this improvement. These developments can be classified under four main modifications: (1) change in flux density scale, (2) change in analysis that reduced the measured power spectrum value, (3) change in analysis that reduced the contamination in the noise, and (4) change in RFI mitigation.

Adopting a new catalog for better calibration and subtraction accuracy lowered our power spectrum normalization by approximately 1.3. Improving our analysis through the modified gridding kernel and other various techniques reduced contamination in both the measured power and the noise for a combined reduction of 2.8 in power. Finally, excising observations contaminated with faint RFI lowered our limit by a factor of 3.8, but only for the zenith pointing. This has highlighted the critical aspects of precision data analysis on the EoR, and indicates areas that we can continue to develop.

Future data reductions with FHD/ ϵ ppsi can further improve the EoR upper limit by enforcing spectral smoothness in systematic errors. This encompasses a large breadth of error types, including those from the theoretical beam kernel, sky

calibration, and HEALPix interpolation. However, some frequency-dependent errors are inherent to the analysis, like the discreteness of the uv -plane during the estimation of the model visibilities and the gridding process (Kerrigan et al. 2018).

This has culminated in a new general approach in our analysis. We sacrifice modes *not* within the EoR window in order to keep modes in sensitive measurement regions as spectrally precise as possible. For example, the tapered gridded estimator that we apply to the beam kernel when we calculate power spectra is not an accurate representation of the sky. Our subtraction model visibilities will be dominated from sources in the center of the beam, and sources near the horizon will be significantly down-weighted. In order to recover modes in the foreground wedge, sidelobe sources must be included and be precise (Pober et al. 2016). However, the modified gridding kernel reduces the spectral dependence of the discrete-based errors in the EoR window. Thus, we knowingly do not recover modes within the foreground wedge in order to reduce errors in the EoR window.

In the future, we plan to investigate other methods to reduce spectral errors which affect the sensitive measurement modes, specifically those based on the instrument (e.g., the beam and calibration). However, this is difficult if the actual response is spectrally complicated. It is easier for the errors we make to be spectrally smooth if the response is *also* spectrally smooth. This will advise upgrades to the MWA and the design of the Square Kilometre Array (SKA).

Comparisons with other data pipelines have also demonstrated potential ways to remove systematics from more k -modes. The RTS/CHIPS pipeline uses inverse-covariance weighting and other methods to help remove systematics caused from flagging channelizer aliasing. We are dominated on some k -modes from this effect in the FHD/ ϵ ppsi pipeline, with potential for it to affect all k -modes at a low power level. Incorporating these advanced techniques into power spectrum estimation will be an important aspect of lowering the EoR upper limit in our data reductions.

One of the major improvements to our EoR upper limit has come from using the package SSINS to remove RFI-contaminated observations from our integrations. We are beginning to enter an analysis regime where low-level contamination can dominate the power in the EoR window. In the future, we plan to use the SSINS package to its fullest potential in order to recover observations with faint DTV, while still maintaining a high degree of RFI-clean observations. Nevertheless, we have only gained significant improvements from further RFI flagging in the zenith pointing of the N-S polarization. This indicates that there is significant beam errors in non-zenith pointings and some residual systematic in the E-W polarization. Mitigating these contaminations will be investigated in the future.

To conclude our work, we put our results in the context of the wider EoR community. Previous publications from the MWA report $\Delta^2 \leq 2.7 \times 10^4 \text{ mK}^2$ at $k = 0.27 \text{ h Mpc}^{-1}$ and $z = 7.1$ (Beardsley et al. 2016). The PAPER Collaboration, who previously had the lowest EoR upper limits (Ali et al. 2015), has recently reanalyzed these calculations due to signal loss (Cheng et al. 2018; Kolopanis et al. 2019). The next lowest current limits come from the LOFAR Collaboration, with $\Delta^2 \leq 6.3 \times 10^3 \text{ mK}^2$ at $k = 0.053 \text{ h Mpc}^{-1}$ and $z = 10.1$ (Patil et al. 2017).

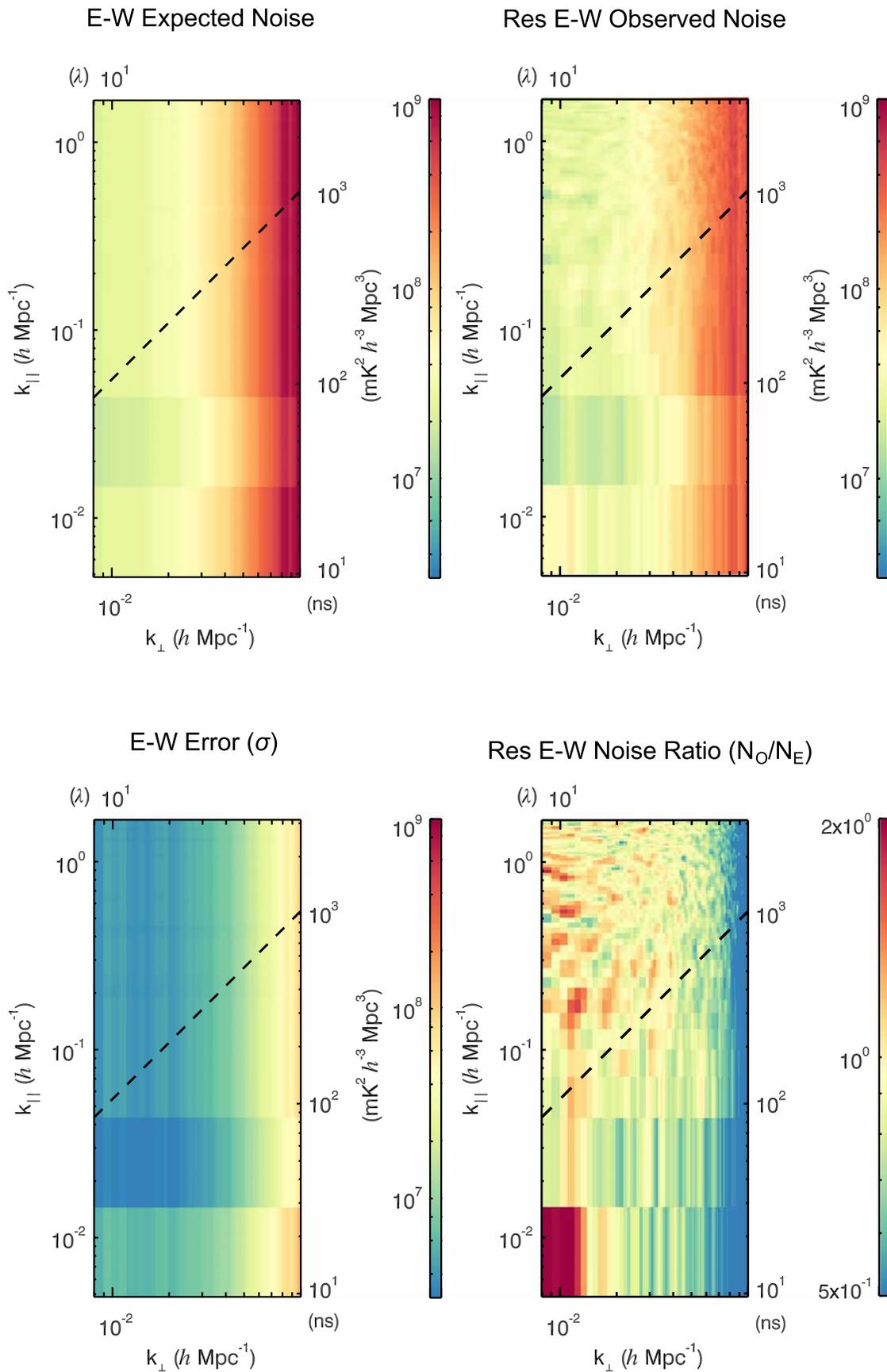


Figure 10. Observed noise from the resulting power (top right), the expected noise from the analytic uncertainty estimate (top left), the resulting analytic error bars (bottom left), and the ratio between the expected and observed noise (bottom right). We validate our analytic error propagation and assumptions in ϵ_{psion} via the noise ratio, which is very close to 1.

Our best limit from this work is $\Delta^2 \leq 3.9 \times 10^3 \text{ mK}^2$ at $k = 0.20 \text{ h Mpc}^{-1}$ and $z = 7$ in the N-S polarization on 21 hr. This is currently the lowest upper limit on EoR structure in the literature. We note that all detections in this work are those of systematics, and longer integrations are needed to reduce thermal noise to the level of the EoR.

We report all k -modes and associated limit calculations from this work in Appendix B for ease of comparison. In addition, both the data²¹ and the software²² is publicly available. The full list of observations, software versions, and settings files necessary to recreate this analysis are available upon request.

By incorporating pipeline improvements to reduce analysis systematics and by removing RFI contamination to reduce observational systematics, we are now in the regime where we are noise dominated in our lowest limit. This puts emphasis on both fronts: lowering the EoR upper limit will require development in analysis precision and observational contamination mitigation. We have proven that both are crucial for detecting the structure of the EoR in the power spectrum.

This research was supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. This work was supported by resources awarded under Astronomy Australia Limited’s merit allocation scheme on the OzSTAR national facility at Swinburne University of Technology. OzSTAR is funded by Swinburne University of Technology and the National Collaborative Research Infrastructure Strategy (NCRIS). A.P.B. is supported by an NSF Astronomy and Astrophysics Postdoctoral Fellowship under #1701440. C.M.T. is supported by an ARC Future Fellowship under grant FT180100196. This work has been funded by National Science Foundation grants AST-1410484, AST-1506024, AST-1613040, AST-1613855, AST-1643011, and AST-1835421, and Amazon/SKA Astrocompute led by D. Jacobs at Arizona State. The International Centre for Radio Astronomy Research (ICRAR) is a Joint Venture of Curtin University and The University of Western Australia, funded by the Western Australian State government. This scientific work makes use of the Murchison Radio-astronomy Observatory, operated by CSIRO. We acknowledge the Wajarri Yamatji people as the traditional owners of the Observatory site. Support for the operation of the MWA is provided by the Australian Government (NCRIS), under a contract to Curtin University administered by Astronomy Australia Limited. We acknowledge the Pawsey Supercomputing Centre which is supported by the Western Australian and Australian Governments.

Facility: MWA.

Appendix A

Theoretical Upper Limits on the 21 cm Power Spectrum

To construct the fiducial theory 21 cm power spectrum used in Figure 6 (brown-solid curve) we use 21CMFAST (Mesinger et al. 2011), adopting the latest astrophysical model parameterization from Park et al. (2019) and assume that we are in the saturated limit (the 21 cm spin temperature is considerably larger than the CMB temperature).

²¹ Data available via MWA All-Sky Virtual Observatory portal (<http://www.mwatelescope.org/data>) and a public Amazon Web Services mirror.

²² Software is open source and freely available as part of the EoRImaging GitHub repository (<https://github.com/EoRImaging>).

Rather than using just a single fiducial model, we can also explore the allowed variation in the amplitude of the theoretical 21 cm power spectrum owing to the uncertainties in the underlying reionization astrophysics. We achieve this by calculating theoretical 2σ upper limits on the 21 cm power spectrum model (brown-dashed curve in Figure 6) following a similar approach to that of Pober et al. (2016). That is, we use 21CMC (Greig & Mesinger 2015), a Monte Carlo Markov chain (MCMC) sampler of 3D reionization simulations, and use existing observational constraints on the reionization epoch to constrain the theoretical astrophysical models. Specifically, we only consider the six astrophysical parameters governing the ionizing sources from Park et al. (2019) and assume the spin temperature is saturated. These six astrophysical parameters include both a mass-dependent escape fraction and fraction of gas in stars, a minimum turnover mass for haloes hosting star-forming galaxies, and a star formation timescale. Additionally, we ignore recombinations, and thus include a mean photon horizon, R_{mfp} parameter.

For our observational constraints, we use the limits on the intergalactic medium neutral fraction at the tail end of reionization (at $z = 5.9$; $x_{\text{HI}} < 0.06 + 0.05 (1\sigma)$) from the dark pixel statistics from quasar spectroscopy (McGreer et al. 2015), the latest estimate for the electron scattering optical depth from Planck ($\tau = 0.054 \pm 0.007$, Planck Collaboration et al. 2018) and the ultraviolet galaxy luminosity functions at $z = 6, 7, 8$, and 10 (Bouwens et al. 2015, 2017; Oesch et al. 2018). Post-processing the output 21 cm power spectrum data from each sampled astrophysical model from the MCMC, we can construct marginalized probability distribution functions (PDFs) for the power spectrum amplitude as a function of Fourier mode, k . Our theoretical limit on the 21 cm power spectrum is then obtained from sampling these PDFs.

Appendix B

All Calculated EoR Upper Limits

We report all EoR upper limits from this work in Table 1 to aid the community in creating comparisons. This data set consists of 678 observations, whose selection from the full 1029-observation data set is discussed in Section 5.1. The various masks in our binning scheme are presented in Section 5.2, and we choose to analyze the frequency range $\sim 168.5\text{--}187.3 \text{ MHz}$ to avoid known instrumental effects.

The E-W values (left) and N-S values (right) are reported in Table 1. We include the k (h Mpc^{-1}), the upper limit Δ_U^2 (mK^2), the lower uncertainty bound Δ_L^2 (mK^2), the measured power Δ^2 (mK^2), and the 1σ thermal noise (mK^2). If the k -mode is consistent with a non-detection (e.g., the measured power or the calculated lower uncertainty bound is negative) we report a lower uncertainty bound of zero.

ORCID iDs

N. Barry  <https://orcid.org/0000-0003-2064-6979>
 C. M. Trott  <https://orcid.org/0000-0001-6324-1766>
 B. Pindor  <https://orcid.org/0000-0003-3240-2437>
 A. P. Beardsley  <https://orcid.org/0000-0001-9428-8233>
 B. J. Hazelton  <https://orcid.org/0000-0001-7532-645X>
 I. S. Sullivan  <https://orcid.org/0000-0001-8708-251X>
 M. F. Morales  <https://orcid.org/0000-0001-7694-4030>
 J. C. Pober  <https://orcid.org/0000-0002-3492-0433>
 J. Line  <https://orcid.org/0000-0002-9130-5920>
 B. Greig  <https://orcid.org/0000-0002-4085-2094>

Table 1
The Calculated EoR Upper Limits and Related Values for Each Polarization

E-W					N-S				
k	$\Delta_{\bar{U}}^2$	$\Delta_{\bar{L}}^2$	Δ^2	Thermal	k	$\Delta_{\bar{U}}^2$	$\Delta_{\bar{L}}^2$	Δ^2	Thermal
0.174	1.13×10^4	9.75×10^3	1.05×10^4	3.98×10^2	0.174	9.22×10^3	7.67×10^3	8.45×10^3	3.87×10^2
0.203	1.03×10^4	8.14×10^3	9.21×10^3	5.35×10^2	0.203	3.89×10^3	1.81×10^3	2.85×10^3	5.21×10^2
0.232	1.31×10^4	1.02×10^4	1.16×10^4	7.46×10^2	0.232	6.32×10^3	3.42×10^3	4.87×10^3	7.25×10^2
0.261	2.58×10^4	2.17×10^4	2.38×10^4	1.03×10^3	0.261	9.61×10^3	5.61×10^3	7.61×10^3	9.99×10^2
0.290	5.19×10^4	4.64×10^4	4.92×10^4	1.37×10^3	0.290	1.56×10^4	1.03×10^4	1.29×10^4	1.33×10^3
0.319	5.47×10^4	4.75×10^4	5.11×10^4	1.79×10^3	0.319	2.07×10^4	1.37×10^4	1.72×10^4	1.74×10^3
0.349	8.99×10^4	8.07×10^4	8.53×10^4	2.30×10^3	0.349	8.32×10^4	7.42×10^4	7.87×10^4	2.23×10^3
0.523	9.11×10^4	6.06×10^4	7.58×10^4	7.63×10^3	0.523	7.09×10^4	4.13×10^4	5.61×10^4	7.40×10^3
0.552	4.81×10^4	1.22×10^4	3.02×10^4	8.97×10^3	0.552	3.20×10^4	0	1.44×10^4	8.71×10^3
0.581	3.25×10^4	0	1.09×10^4	1.05×10^4	0.581	4.43×10^4	3.60×10^3	2.39×10^4	1.02×10^4
0.610	2.26×10^4	0	-8.02×10^3	1.21×10^4	0.610	6.38×10^4	1.69×10^4	4.04×10^4	1.17×10^4
0.639	9.88×10^4	4.35×10^4	7.12×10^4	1.38×10^4	0.639	1.11×10^5	5.70×10^4	8.38×10^4	1.34×10^4
0.668	4.94×10^5	4.31×10^5	4.63×10^5	1.59×10^4	0.668	2.62×10^5	2.00×10^5	2.31×10^5	1.54×10^4
0.697	6.36×10^5	5.64×10^5	6.00×10^5	1.81×10^4	0.697	2.28×10^5	1.58×10^5	1.93×10^5	1.75×10^4
0.726	3.20×10^5	2.38×10^5	2.79×10^5	2.04×10^4	0.726	4.25×10^4	0	-3.93×10^3	1.98×10^4
0.755	2.49×10^5	1.57×10^5	2.03×10^5	2.30×10^4	0.755	1.22×10^5	3.29×10^4	7.75×10^4	2.23×10^4
0.929	1.50×10^6	1.33×10^6	1.41×10^6	4.29×10^4	0.929	1.35×10^6	1.19×10^6	1.27×10^6	4.16×10^4
0.958	3.52×10^5	1.64×10^5	2.58×10^5	4.70×10^4	0.958	7.01×10^4	0	-6.04×10^4	4.56×10^4
0.987	1.77×10^5	0	7.25×10^4	5.14×10^4	0.987	5.91×10^4	0	-1.15×10^5	4.99×10^4
1.017	2.88×10^5	6.33×10^4	1.75×10^5	5.61×10^4	1.017	3.24×10^5	1.06×10^5	2.15×10^5	5.44×10^4
1.046	3.24×10^5	8.01×10^4	2.02×10^5	6.09×10^4	1.046	5.32×10^5	2.96×10^5	4.14×10^5	5.91×10^4
1.075	2.78×10^5	1.48×10^4	1.46×10^5	6.57×10^4	1.075	2.42×10^5	0	1.13×10^5	6.38×10^4
1.104	1.35×10^5	8.92×10^4	-4.46×10^4	7.17×10^4	1.104	7.41×10^4	0	-1.92×10^5	6.96×10^4
1.133	3.16×10^5	4.87×10^3	1.60×10^5	7.76×10^4	1.133	8.57×10^4	0	-1.86×10^5	7.53×10^4
1.162	6.94×10^5	3.59×10^5	5.27×10^5	8.37×10^4	1.162	1.13×10^5	0	-1.36×10^5	8.12×10^4
1.191	1.13×10^6	7.65×10^5	9.46×10^5	9.02×10^4	1.191	4.65×10^5	1.15×10^5	2.90×10^5	8.75×10^4
1.365	2.63×10^6	2.08×10^6	2.36×10^6	1.36×10^5	1.365	2.32×10^6	1.80×10^6	2.06×10^6	1.32×10^5
1.394	1.50×10^6	9.16×10^5	1.21×10^6	1.45×10^5	1.394	7.22×10^5	1.60×10^5	4.41×10^5	1.40×10^5
1.423	7.45×10^5	1.29×10^5	4.37×10^5	1.54×10^5	1.423	5.48×10^5	0	2.46×10^5	1.49×10^5
1.452	2.20×10^5	0	-2.95×10^5	1.63×10^5	1.452	6.58×10^5	2.24×10^4	3.40×10^5	1.59×10^5
1.481	2.18×10^5	0	-3.40×10^5	1.70×10^5	1.481	4.67×10^5	0	1.19×10^5	1.65×10^5
1.510	4.90×10^5	0	1.10×10^5	1.78×10^5	1.510	6.16×10^5	0	2.64×10^5	1.73×10^5
1.539	1.02×10^6	2.46×10^5	6.35×10^5	1.94×10^5	1.539	1.48×10^6	7.24×10^5	1.10×10^6	1.89×10^5
1.568	1.54×10^6	7.12×10^5	1.12×10^6	2.06×10^5	1.568	2.05×10^6	1.25×10^6	1.65×10^6	2.00×10^5
1.597	2.32×10^6	1.45×10^6	1.88×10^6	2.18×10^5	1.597	1.73×10^6	8.90×10^5	1.31×10^6	2.11×10^5

Note. We include the k ($h \text{ Mpc}^{-1}$), the upper limit $\Delta_{\bar{U}}^2$ (mK^2), the lower uncertainty bound $\Delta_{\bar{L}}^2$ (mK^2), the measured power Δ^2 (mK^2), and the 1σ thermal noise (mK^2).

R. Byrne <https://orcid.org/0000-0003-4980-2736>
A. Lanman <https://orcid.org/0000-0003-2116-3573>
C. H. Jordan <https://orcid.org/0000-0002-1220-2940>
R. C. Joseph <https://orcid.org/0000-0003-3457-4670>
B. McKinley <https://orcid.org/0000-0002-9006-1450>
S. Yoshiura <https://orcid.org/0000-0003-0581-5973>
J. D. Bowman <https://orcid.org/0000-0002-8475-2036>
B. M. Gaensler <https://orcid.org/0000-0002-3382-9558>
J. N. Hewitt <https://orcid.org/0000-0002-4117-570X>
D. C. Jacobs <https://orcid.org/0000-0002-0917-2269>
D. A. Mitchell <https://orcid.org/0000-0002-1828-1969>
N. Udaya Shankar <https://orcid.org/0000-0001-8974-3369>
R. Subrahmanyan <https://orcid.org/0000-0001-9913-900X>
S. J. Tingay <https://orcid.org/0000-0002-8195-7562>
R. L. Webster <https://orcid.org/0000-0002-5325-2709>
J. S. B. Wyithe <https://orcid.org/0000-0001-7956-9758>

References

Ali, Z. S., Parsons, A. R., Zheng, H., et al. 2015, *ApJ*, 809, 61
Barry, N. 2018, PhD thesis, University of Washington
Barry, N., Beardsley, A. P., Byrne, R., et al. 2019, *PASA*, 36, e026
Barry, N., Hazelton, B., Sullivan, I., Morales, M. F., & Pober, J. C. 2016, *MNRAS*, 461, 3135
Beardsley, A. P., Hazelton, B. J., Morales, M. F., et al. 2013, *MNRAS: Letters*, 429, L5
Beardsley, A. P., Hazelton, B. J., Sullivan, I. S., et al. 2016, *ApJ*, 833, 102
Bharadwaj, S., & Pandey, S. K. 2005, *MNRAS*, 358, 968
Bouwens, R. J., Oesch, P. A., Illingworth, G. D., Ellis, R. S., & Stefanon, M. 2017, *ApJ*, 843, 129
Bouwens, R. J., et al. 2015, *ApJ*, 803, 34
Bowman, J. D., Cairns, I., Kaplan, D. L., et al. 2013, *PASA*, 30, e031
Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., & Mahesh, N. 2018, *Natur*, 555, 67
Carroll, P. A., Line, J., Morales, M. F., et al. 2016, *MNRAS*, 461, 4151
Cheng, C., Parsons, A. R., Kolopanis, M., et al. 2018, *ApJ*, 868, 26
Choudhuri, S., Bharadwaj, S., Ghosh, A., & Ali, S. S. 2014, *MNRAS*, 445, 4351
Choudhuri, S., Bharadwaj, S., Roy, N., Ghosh, A., & Ali, S. S. 2016, *MNRAS*, 459, 151
Cornwell, T. J., Golap, K., & Bhatnagar, S. 2008, *ISTSP*, 2, 647
Datta, A., Bowman, J. D., & Carilli, C. L. 2010, *ApJ*, 724, 526
DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017, *PASP*, 129, 045001
Dillon, J. S., Neben, A. R., Hewitt, J. N., et al. 2015, *PhRvD*, 91, 123011
Ewall-Wice, A., Dillon, J. S., Hewitt, J. N., et al. 2016, *MNRAS*, 460, 4320
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, *ApJ*, 622, 759
Greig, B., & Mesinger, A. 2015, *MNRAS*, 449, 4246
Hazelton, B. J., Morales, M. F., & Sullivan, I. S. 2013, *ApJ*, 770, 156
Hurley-Walker, N., Callingham, J. R., Hancock, P. J., et al. 2017, *MNRAS*, 464, 1146
Hurley-Walker, N., Morgan, J., Wayth, R. B., et al. 2014, *PASA*, 31, e045
Jacobs, D. C., Hazelton, B. J., Trott, C. M., et al. 2016, *ApJ*, 825, 114

- Kay, S. M. 1993, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Upper Saddle River, NJ: Prentice-Hall)
- Kerrigan, J. R., Pober, J. C., Ali, Z. S., et al. 2018, *ApJ*, **864**, 131
- Kolopanis, M., Jacobs, D. C., Cheng, C., et al. 2019, *ApJ*, in press (arXiv:1909.02085)
- Liu, A., Parsons, A. R., & Trott, C. M. 2014, *PhRvD*, **90**, 023018
- Lomb, N. R. 1976, *Ap&SS*, **39**, 447
- Majumdar, S., Pritchard, J. R., Mondal, R., et al. 2018, *MNRAS*, **476**, 4007
- McGreer, I. D., Mesinger, A., & D’Odorico, V. 2015, *MNRAS*, **447**, 499
- Mesinger, A., Furlanetto, S., & Cen, R. 2011, *MNRAS*, **411**, 955
- Mitchell, D. A., Greenhill, L. J., Wayth, R. B., et al. 2008, *ISTSP*, **2**, 707
- Morales, M. F., Beardsley, A., Pober, J., et al. 2019, *MNRAS*, **483**, 2207
- Morales, M. F., Hazelton, B., Sullivan, I., & Beardsley, A. 2012, *ApJ*, **752**, 137
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Labbé, I., & Stefanon, M. 2018, *ApJ*, **855**, 105
- Offringa, A. R., de Bruyn, A. G., Biehl, M., et al. 2010, *MNRAS*, **405**, 155
- Offringa, A. R., Gronde, J. J. v. d., & Roerdink, J. B. T. M. 2012, *A&A*, **539**, A95
- Offringa, A. R., Mertens, F., & Koopmans, L. V. E. 2019, *MNRAS*, **484**, 2866
- Offringa, A. R., Wayth, R. B., Hurley-Walker, N., et al. 2015, *PASA*, **32**, e008
- Ord, S. M., Mitchell, D. A., Wayth, R. B., et al. 2010, *PASP*, **122**, 1353
- Paciga, G., Albert, J. G., Bandura, K., et al. 2013, *MNRAS*, **433**, 639
- Paciga, G., Chang, T.-C., Gupta, Y., et al. 2011, *MNRAS*, **413**, 1174
- Park, J., Mesinger, A., Greig, B., & Gillet, N. 2019, *MNRAS*, **484**, 933
- Parsons, A. R., Backer, D. C., Foster, G. S., et al. 2010, *AJ*, **139**, 1468
- Parsons, A. R., Pober, J. C., Aguirre, J. E., et al. 2012, *ApJ*, **756**, 165
- Patil, A. H., Yatawatta, S., Koopmans, L. V. E., et al. 2017, *ApJ*, **838**, 65
- Patil, A. H., Zaroubi, S., Chapman, E., et al. 2014, *MNRAS*, **443**, 1113
- Planck Collaboration, Adam, R., Aghanim, N., et al. 2016, *A&A*, **596**, A108
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018, arXiv:1807.06209
- Pober, J. C., Greig, B., & Mesinger, A. 2016, *MNRAS*, **463**, L56
- Pober, J. C., Hazelton, B. J., Beardsley, A. P., et al. 2016, *ApJ*, **819**, 8
- Pober, J. C., Liu, A., Dillon, J. S., et al. 2014, *ApJ*, **782**, 66
- Pober, J. C., Parsons, A. R., Aguirre, J. E., et al. 2013, *ApJL*, **768**, L36
- Prabu, T., Srivani, K. S., Roshi, D. A., et al. 2015, *ExA*, **39**, 73
- Robertson, B. E., Ellis, R. S., Dunlop, J. S., McLure, R. J., & Stark, D. P. 2010, *Natur*, **468**, 49
- Salvini, S., & Wijnholds, S. J. 2014, *A&A*, **571**, A97
- Scargle, J. D. 1982, *ApJ*, **263**, 835
- Shaw, J. R., Sigurdson, K., Pen, U.-L., Stebbins, A., & Sitwell, M. 2014, *ApJ*, **781**, 57
- Shaw, J. R., Sigurdson, K., Sitwell, M., Stebbins, A., & Pen, U.-L. 2015, *PhRvD*, **91**, 083514
- Sullivan, I. S., Morales, M. F., Hazelton, B. J., et al. 2012, *ApJ*, **759**, 17
- Thyagarajan, N., Shankar, N. U., Subrahmanyan, R., et al. 2013, *ApJ*, **776**, 6
- Tingay, S. J., Goeke, R., Bowman, J. D., et al. 2013, *PASA*, **30**, e007
- Trott, C. M. 2014, *PASA*, **31**, e026
- Trott, C. M., Pindor, B., Procopio, P., et al. 2016, *ApJ*, **818**, 139
- Trott, C. M., Watkinson, C. A., Jordan, C. H., et al. 2019, *PASA*, **36**, e023
- Trott, C. M., Wayth, R. B., & Tingay, S. J. 2012, *ApJ*, **757**, 101
- van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, *A&A*, **556**, A2
- Vedantham, H., Shankar, N. U., & Subrahmanyan, R. 2012, *ApJ*, **745**, 176
- Watkinson, C. A., Giri, S. K., Ross, H. E., et al. 2019, *MNRAS*, **482**, 2653
- Wayth, R. B., Tingay, S. J., Trott, C. M., et al. 2018, *PASA*, **35**, 3033
- Wilensky, M. J., Morales, M. F., Hazelton, B. J., et al. 2019, *PASP*, in press (arXiv:1906.01093)
- Yatawatta, S., de Bruyn, A. G., Brentjens, M. A., et al. 2013, *A&A*, **550**, A136