Publications of the Astronomical Society of Australia (2019), **36**, e026, 17 pages
doi: 10.1017/pasa.2019.21

CAMBRIDGE UNIVERSITY PRESS

Research Paper

The FHD/εppsilon Epoch of Reionisation power spectrum pipeline

N. Barry[1,2], A. P. Beardsley[3], R. Byrne[4], B. Hazelton[4,5], M. F. Morales[4,6], J. C. Pober[7] and I. Sullivan[8]

[1]School of Physics, The University of Melbourne, Parkville, VIC 3010, Australia. [2]ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), [3]School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287, USA, [4]Department of Physics, University of Washington, Seattle, WA 98195, USA, [5]eScience Institute, University of Washington, Seattle, WA 98195, USA, [6]Dark Universe Science Center, University of Washington, Seattle, 98195, USA, [7]Department of Physics, Brown University, Providence, RI 02906, USA and [8]Department of Astronomy, University of Washington, Seattle, WA 98195, USA

Abstract

Epoch of Reionisation (EoR) data analysis requires unprecedented levels of accuracy in radio interferometer pipelines. We have developed an imaging power spectrum analysis to meet these requirements and generate robust 21 cm EoR measurements. In this work, we build a signal path framework to mathematically describe each step in the analysis, from data reduction in the Fast Holographic Deconvolution (FHD) package to power spectrum generation in the εppsilon package. In particular, we focus on the distinguishing characteristics of FHD/εppsilon: highly accurate spectral calibration, extensive data verification products, and end-to-end error propagation. We present our key data analysis products in detail to facilitate understanding of the prominent systematics in image-based power spectrum analyses. As a verification to our analysis, we also highlight a full-pipeline analysis simulation to demonstrate signal preservation and lack of signal loss. This careful treatment ensures that the FHD/εppsilon power spectrum pipeline can reduce radio interferometric data to produce credible 21 cm EoR measurements.

Keywords: cosmology: dark ages, reionisation, first stars – instrumentation: interferometers – methods: data analysis – techniques: interferometric

(Received 05 January 2019; Revised 02 May 2019; Accepted 15 May 2019)

1. Introduction

Structure measurements of the Epoch of Reionisation (EoR) have the potential to revolutionise our understanding of the early universe. Many radio interferometers are pursuing these detections, including the MWA (Tingay et al. 2013; Bowman et al. 2013; Wayth et al. 2018), PAPER (Parsons et al. 2010), LOFAR (Yatawatta et al. 2013; van Haarlem et al. 2013), HERA (Pober et al. 2014; DeBoer et al. 2017), and the SKA (Mellema et al. 2013; Koopmans et al. 2014). The sheer amount of data to be processed even for the most basic of interferometric analyses requires sophisticated pipelines.

Furthermore, these analyses must be very accurate; the EoR signal is many orders of magnitude fainter than the foregrounds. Systematics from inaccurate calibration, spatial/frequency transform artefacts, and other sources of spectral contamination will preclude an EoR measurement. Understanding and correcting for these systematics have been the main focus of improvements for MWA Phase I EoR analyses, including the Fast Holographic Deconvolution (FHD)/εppsilon pipeline, since their description in Jacobs et al. (2016).

FHD/εppsilon power spectrum analyses have been prevalent in the literature, including data reduction for MWA Phase I (Beardsley et al. 2016), MWA Phase II (Li et al. 2018), and for PAPER (Kerrigan et al. 2018), with HERA analyses planned. It is also flexible enough to be used for pure simulation, particularly in investigating calibration effects (Barry et al. 2016; Byrne et al. 2019). Combined with its focus on spectrally accurate calibration, as well as end-to-end data-matched model simulations and error propagation, FHD/εppsilon is well suited for EoR measurements and studies.

In this work, we describe the FHD/εppsilon power spectrum pipeline in full with a consistent mathematical framework, highlighting recent improvements and key features. We trace the signal path from its origin through the main components of our data reduction analysis. Our final power spectrum products, including uncertainty estimates, are detailed extensively. We perform many diagnostics with various types of power spectra, which provide confirmations for our improvements to the pipeline.

As further verification of the FHD/εppsilon power spectrum pipeline, we present proof of signal preservation. Confidence in EoR upper limits relies on our ability to avoid absorption of the signal itself. A full end-to-end analysis simulation within our pipeline proves that we do not suffer from signal loss, thereby adding credibility to our EoR upper limits.

First, we detail the signal path framework in Section 2 to provide a mathematical foundation. In Section 3, we build an analytical description of FHD, focusing on the recent improvements in calibration. Section 4 outlines the data products from FHD and our choice of integration methodology. Propagating errors and creating power spectra in εppsilon are described in Section 5. Finally, we illustrate all of our power spectra data products in Section 6, including proof of signal preservation.

Author for correspondence: N. Barry, E-mail: nichole.barry@unimelb.edu.au

Cite this article: Barry N, Beardsley AP, Byrne R, Hazelton B, Morales MF, Pober JC and Sullivan I. The FHD/εppsilon Epoch of Reionisation power spectrum pipeline. *Publications of the Astronomical Society of Australia* **36**, e026, 1–17. https://doi.org/10.1017/pasa.2019.21

2. The signal path framework

The sky signal is modified during the journey from when it was emitted to when it was recorded. In order to uncover the EoR, the true sky must be separated from all intervening effects, including those introduced via the instrument. We describe all known signal modifications to build a consistent, mathematical framework.

There are three main types of signal modification: those that occur before, during, and after interaction with the antenna elements, as shown in Figure 1. We adhere to notation from Hamaker, Bregman, & Sault (1996) whenever possible in our brief catalogue of interactions.

B: Before antenna

- Faraday rotation from interaction with the ionosphere, **F**.
- Source position offsets **O** due to variation in ionospheric thickness.
- Unmodelled signals caused by radio frequency interference (RFI), **U**.

S: At antenna

- Parallactic rotation **P** between the rotating basis of the sky and the basis of the antenna elements.
- Antenna correlations from cross-talk between antennas, **X**.
- Antenna element response, **C**, usually referred to as the beam.[a]
- Errors in the expected nominal configuration and beam model, **D**.

E: After antenna

- Electronic gain amplitude and phase **R** from a typical response of each antenna.
- Gain amplitude changes from temperature effects **T** on amplifiers.
- Gain amplitude and phase oscillations **K** due to cable reflections, both at the end of the cables and at locations where the cable is kinked.
- Frequency correlations **A** caused by aliasing in polyphase filter banks or other channelisers.

Each contribution can be modelled as a matrix which depends on $[f, t, P, AB]$: frequency, time, instrumental polarisation, and antenna cross-correlations. The expected contributions along the signal path are thus $\mathbf{B} = \mathbf{UOF}$, $\mathbf{S} = \mathbf{DCXP}$, and $\mathbf{E} = \mathbf{AKTR}$. The visibility measurement equation takes the form

$$\mathbf{v}^{\text{meas}} \sim \mathbf{E}\,\mathbf{S}\,\mathbf{B}\,\mathbf{I}^{\text{true}} + \text{noise}, \tag{1}$$

where $\mathbf{v}^{\text{meas}}$ are the measured visibilities, $\mathbf{I}^{\text{true}}$ is the true sky, **B** are contributions that occur between emission and the ground, **S** are contributions from the antenna configuration, and **E** are contributions from the electronic response. We have included all known contributions and modifications to the signal, including the total thermal noise. However, there may be unknown contributions, hence we describe Equation 1 as an approximation. All components are summarised in Table 1 for reference.

[a]Hamaker et al. (1996) use **C** to primarily account for a rotating feed; however, we incorporate the antenna response as well. We also expect and model polarisation correlation in the Jones matrices (Section 3.2), whereas Hamaker et al. (1996) does not.

Table 1. Brief definitions of the variables used within the signal path framework, organised by type. There are four types, (**B**) interactions that occur before the antenna elements, (**S**) interactions that occur at the antenna elements, (**E**) interactions that occur after the antenna elements, and (**v**) visibility-related variables.

Type	Variables	Definition
B	**F**	Faraday rotation
	O	Source position offsets
	U	Unmodelled RFI
S	**P**	Parallactic rotation
	X	Cross-talk antenna correlations
	C	Antenna element response
	D	Errors in nominal configuration
E	**R**	Gain amplitude and phase
	T	Temperature changes
	K	Cable reflections
	A	Frequency correlations
v	$\mathbf{I}^{\text{true}}$	True sky
	$\mathbf{v}^{\text{meas}}$	Measured visibilities
	noise	Thermal noise

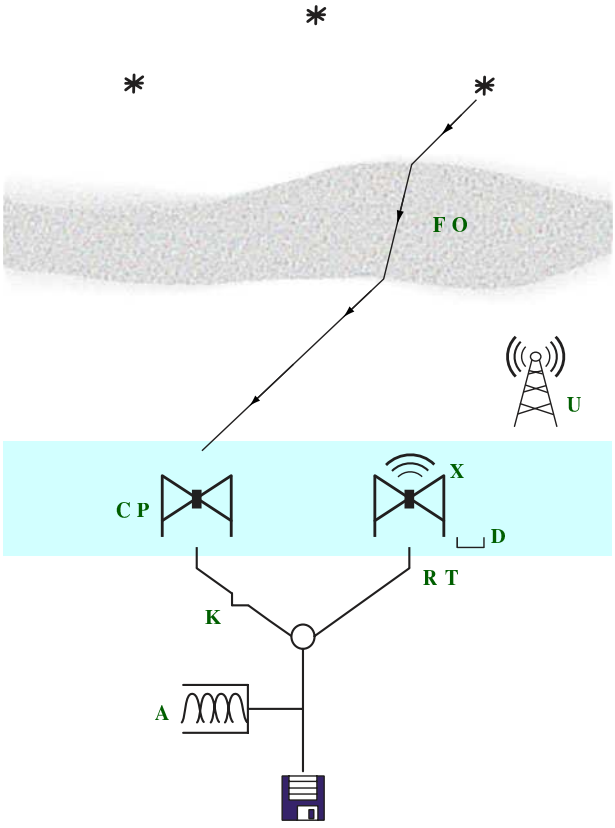

Figure 1. The signal path through the instrument. There are three categories of signal modification: before antenna, at antenna (coloured light blue), and after antenna. Each modification matrix (green) is detailed in the text and Table 1.

The visibilities can be condensed into vectors since the measurements are naturally independent over the $[f, t, P, AB]$ dimensions. However, the modification matrices can introduce correlations across the dimensions, and thus cannot be reduced without assumptions.

One of the goals of an imaging EoR analysis is to reconstruct the true sky visibilities, $\mathbf{v}^{\text{true}}$, given the measured sky visibilities, $\mathbf{v}^{\text{meas}}$. This is achieved through the process of calibration. We will use our generalised framework described in this subsection to detail the assumptions and methodology for the FHD/εppsilon pipeline.

3. Fast Holographic Deconvolution

FHD[b] is an open-source radio analysis package that produces calibrated sky maps from measured visibilities. Initially built to implement an efficient deconvolution algorithm (Sullivan et al. 2012), its purpose is now to serve as a vital step in EoR power spectrum analysis.[c]

FHD is a tool to analyse radio interferometric data and has a series of main functions at its core: (1) creating model visibilities from sky catalogues for calibration and subtraction, (2) gridding calibrated data, and (3) making images for analysis and integration.

Using the signal path framework described in Section 2, we will describe the steps in building a gridding kernel (Section 3.2), forming model visibilities (Section 3.3), calibrating data visibilities (Section 3.4), and producing images (Section 3.5).

3.1. Pre-pipeline flagging

Before any analysis can begin, the data must be RFI-flagged. RFI, particularly from FM radio and digital TV, can contaminate the data. However, RFI has characteristic signatures in time and frequency which allow it to be systematically removed by trained packages.

We use the package AOFLAGGER[d] to RFI-flag the data (Offringa et al. 2015). This removes bright line-like emission, but has difficulty removing faint, broad emission like TV. We completely remove any observations that have signatures of TV. Therefore, we remove contributions from $\mathbf{U}$ by avoidance.

As demonstrated in Offringa, Mertens, & Koopmans (2019), averaging over flagged channels can cause bias. However, this does not affect the nominal FHD/εppsilon pipeline because it avoids inverse-variance weighting of the visibilities. Any future incorporation of inverse-variance weighting will need to take this into account.

3.2. Generating the beam

The measurement collecting area of an antenna element, $\mathbf{C}$, is commonly referred to as the primary beam. A deep knowledge of the beam is critical for precision measurements with widefield interferometers (Pober et al. 2016). Since our visibility measurements are correlations of the voltage response between elements, we must understand the footprint of each element's voltage response to reconstruct images (Morales & Matejek 2009).

We build the antenna element response from finely interpolated beam images. This happens in the instrument's coherency domain, or instrumental polarisation. We assume each element has two physical components p and q which are orthogonal, so each visibility correlation between elements a and b will have a $P = \{p_a p_b, p_a q_b, q_a p_b, q_a q_b\}$ response in the coherency domain (Hamaker et al. 1996). This is calculated from the two elements' polarised response patterns. For example,

$$\begin{aligned}\mathbf{C}_{p_a q_b} &= \mathbf{C}_{q_a p_b} \\ &= \Bigg(\left(\mathbf{J}_{x,p_a} \mathbf{J}^*_{x,p_b} + \mathbf{J}_{y,p_a} \mathbf{J}^*_{y,p_b} \right) \\ &\qquad \left(\mathbf{J}_{x,q_a} \mathbf{J}^*_{x,q_b} + \mathbf{J}_{y,q_a} \mathbf{J}^*_{y,q_b} \right) \Bigg)^{\frac{1}{2}}, \end{aligned} \tag{2}$$

where $\mathbf{C}_{p_a q_b}$ is the beam response for a p component in element a and a q component in element b, $\mathbf{J}$ is the vector field for an element (also known as the Jones matrix), and the subscript x, p_a is the contribution of the p component in element a for the coordinate x (and likewise for the other subscripts) (Sutinjo et al. 2015). Each matrix is a function of spatial coordinates and frequency, and each operation is done element-by-element.

The Jones matrices describe the transformation needed to account for $\mathbf{P}$, or parallactic rotation. Due to the wide field-of-view and the lack of moving parts for most EoR instruments, this is a natural requirement. Known inter-dipole mutual coupling, a form of $\mathbf{X}$, is also captured in the Jones matrices. We generate $\mathbf{C}_P$ for a pair of elements, which can be applied to all other identical element pairs.

We then take the various beams in direction cosine space $\{l, m\}$ and Fourier transform them to get beams in $\{u, v\}$. Since we plan on using the beam as a gridding kernel later on, the beam must vary as smoothly as possible. This is achievable by hyperresolving beyond the usual uv-grid resolution of $\frac{1}{2}\lambda$, down to typically $\frac{1}{7000}\lambda$. We create this beam once to build a highly resolved reference table.

FHD does have the flexibility to generate unique beams for each element given individual element metadata; however, this quickly increases computing resources. Instead, we can build a coarse beam per baseline with phase offsets in image space to account for pixel centre offsets (Line 2017), which is built on-demand rather than saved as a reference table. This corrects for one form of $\mathbf{D}$, or individual element variation and error.

When we generate model visibilities, we want to be as instrumentally accurate as possible. This necessitates a kernel which represents the instrumental response to the best of our knowledge. However, we can choose a separate, modified kernel for uv-plane generation in power spectrum analysis, similar to a Tapered Gridded Estimator (Choudhuri et al. 2014, 2016). As long as the proper normalisations are taken into account in power estimation, a modified gridding kernel acts as a weighting of the instrumental response. This image-space weighting will correlate pixels; we investigate contributions from pixel correlations in the power spectrum in Section 6.1.

3.3. Creating model visibilities

We must calibrate our input visibilities. Due to our wide field-of-view and accuracy requirements, FHD simulates *all* reliable sources out to typically 1% beam level in the primary lobe and the sidelobes to build a nearly complete theoretical sky. For the MWA Phase I, the 1% beam level includes the first sidelobe out to a field-of-view of approximately 100°. By comparing these model visibilities to the data visibilities, we can estimate the instrument's contribution.

Generating accurate model visibilities is an important step in calibration. Therefore, the estimate of the sky must be as complete

[b] https://github.com/EoRImaging/FHD
[c] Deconvolution in FHD is generally only used in building sky models for calibration and subtraction.
[d] https://sourceforge.net/p/aoflagger/wiki/Home/

as possible, including source positions, morphology, and brightness. For example, we can use GLEAM (Hurley-Walker et al. 2017), an extragalactic catalogue with large coverage and high completeness, to model sources for observations in the Southern sky. For a typical MWA field off the galactic plane, we model over 50 000 known sources using GLEAM. As for a typical PAPER observation, we model over 10 000 known sources above 1 Jy (Kerrigan et al. 2018).

Not all sources are unpolarised. There are cases where a source can be linearly or circularly polarised. Only ten out of the thousands of sources seen in a typical MWA field off the galactic plane are known to be reliably polarised (Riseley et al. 2018). Therefore, we assume there are no polarised sources in making model visibilities. With this assumption, we avoid complications from **F**, or Faraday rotation in the polarisation components due to the ionosphere. Implicitly, we also assume the ionosphere is not structured, therefore **F** does not affect unpolarised sources.

We disregard source position offsets **O**, a contamination resulting from ionospheric distortions. We minimise this contribution by excluding data which is significantly modified by ionospheric weather. We determine the quality of the observation using various metrics described in Beardsley et al. (2016), which can be further supplemented by ionospheric data products (Jordan et al. 2017; Trott et al. 2018).

Not all sources are unresolved. We also can optionally include extended sources by modelling their contribution as a series of unresolved point sources (Carroll 2016). This can also be done for creating models of diffuse synchrotron emission (Beardsley 2015). However, the diffuse emission is significantly polarised (Lenc et al. 2016) and difficult to include.

With these assumptions, we now have a reliable sky model that we can use to generate model visibilities. For each point source in our catalogue, we perform a discrete Fourier transform using the RA/Dec floating-point location and the Stokes I brightness. This results in a discretised, model *uv*-plane for each source *without* instrumental effects; typically at $\frac{1}{2}\,\lambda$ resolution. All *uv*-planes from all sources are summed to create a model *uv*-plane of the sky, and this process is repeated for each observation to minimise *w*-terms associated with the instantaneous measurement plane (see Section 4 for more discussion).

Once a model *uv*-plane is created with all source contributions, we simulate what the instrument actually measures. The hyperresolved *uv*-beam from Section 3.2 is the sensitivity of the cross-correlation of two elements. We calculate the *uv*-locations of each cross-correlated visibility, and multiply the model *uv*-plane with the *uv*-beam sampling function. The sum of the sensitivity multiplied by the model at the sampled points is the estimated measured value for that cross-correlated visibility.

These visibilities represent our best estimate of what the instrument should have measured, disregarding any source position offsets **O**, polarised Faraday rotation **F**, and diffuse emission. Any deviations from these model visibilities (whether instrumental or not) will manifest as errors in the comparison between the data and model during calibration.

Our model visibilities can be represented in the signal path framework as

$$\mathbf{m} \sim \mathbf{S}\,\mathbf{I}^{\text{true}}, \tag{3}$$

where $\mathbf{m}$ are the estimated model visibilities, $\mathbf{I}^{\text{true}}$ is the true sky, and $\mathbf{S}$ are contributions from the plane of the measurement. We have not included any modifications in the signal path from before the instrument, **B**, and have instead chosen to use an avoidance technique for affected data.

3.4. Calibration

We are left with one type of modification to the signal that has not been accounted for by the model visibilities or by avoidance: **E**, the electronic response. This is what we classify as our calibration.

At this point in the analysis, we have a measurement equation that looks like

$$\mathbf{v}^{\text{meas}} \sim \mathbf{E}\,\mathbf{m} + \text{noise}. \tag{4}$$

We assume the electronic response **E** varies slowly with time, and thus does not change significantly over an observation (e.g. 2 min for the MWA). Due to our model-based assumptions, we do not have any non-celestial time correlations, antenna correlations, or unknown polarisation correlations. The electronic response is thus simply a time-independent gain **G** per observation which is independent per element and polarisation.

We begin by rewriting Equation 4 using these assumptions. The measured cross-correlated visibilities are a function of frequency, time, and polarisation. Individual elements are grouped into the sets $A = \{a_0, a_1, a_2, \ldots, a_n\}$ and $B = \{b_0, b_1, b_2, \ldots, b_n\}$, where a and b iterate through antenna pairs. A visibility is measured for each polarisation $P = \{p_a p_b, p_a q_b, q_a p_b, q_a q_b\}$, where p and q iterate through the two orthogonal instrumental polarisation components for each element a and b.

The resulting relation between the measured visibilities and the model visibilities is

$$\begin{aligned}\mathbf{v}_{ab,pq}([f_o, t]) &\sim \mathbf{G}_{a,p}(f_o, f_i)\mathbf{G}^{*}_{b,q}(f_o, f_i)\mathbf{m}_{ab,pq}([f_i, t]) \\ &\quad + \mathbf{n}_{ab,pq}([f_o, t]),\end{aligned} \tag{5}$$

where $\mathbf{v}_{ab,pq}([f_o, t])$ are the measured visibilities and $\mathbf{m}_{ab,pq}([f_i, t])$ are the model visibilities. Both are frequency and time vectors $[f, t]$ of the visibilities over all A and B element pairs and over all P polarisation products. $\mathbf{G}_{a,p}(f_o, f_i)$ is a frequency matrix of gains given input frequencies f_i which affect multiple output frequencies f_o for instrumental polarisations p for elements in A (and likewise for q and B). Thermal noise $\mathbf{n}$ is independent for each visibility. All variables used in this section are summarised in Table 2 for reference.

Our notation has been specifically chosen. Naturally discrete variables (element pairs and polarisation products) are described in the subscripts. Naturally continuous variables (frequency and time) are function arguments. We group frequency and time into a set $[f, t]$ in the visibilities to create vectors. Since frequency and time are independent, this notation is more compact. In contrast, the gain matrices are not independent in frequency. A full matrix must be used to accurately capture frequency correlation due to **A**, which includes aliasing from common electronics like polyphase filter banks or bandpasses.

To reduce Equation 5 significantly, we make the assumption that the frequencies are independent in **G**. This forces the frequency correlation contribution $\mathbf{A} = \mathbb{I}$, giving

$$\mathbf{G}_{a,p}(f_o, f_i) \sim \text{diag}(\mathbf{g}_{a,p}(f)). \tag{6}$$

The instrumental gains $\mathbf{g}$ are now an independent vector of frequencies for elements in A per instrumental polarisations p (and likewise for q and B). We flag frequency channels which are most affected by aliasing to make this assumption viable. The effect of this flagging is dependent on the instrument; for the MWA, flagging every 1.28 MHz to remove aliasing from the polyphase filter banks introduces harmonic contamination in the power spectrum. Thus, enforcing this assumption usually has consequences in the power spectrum space.

Table 2. Definitions of the variables used for calibration, organised by type.

Type	Variables	Definition
Observation & Scale Parameters	f	Measured frequencies of the observation.
	p, q	Orthogonal instrumental polarisations of the elements in the array.
	a, b	Elements in the array.
	t	Time steps within an observation.
Sets	$A = \{a_0, a_1, a_2, ...a_n\}$	All elements of the array. A and B can be iterated separately to form cross-correlated element pairs.
	$B = \{b_0, b_1, b_2, ...b_n\}$	
	$P = \{p_a p_b, ...q_a q_b\}$	Polarisations in the coherency domain between two elements.
	$C = \{c_n, ...c_0,$	Coefficients of a low-order amplitude and phase polynomial across the
	$\phi_n, ...\phi_0\}$	frequency band.
	$L = \{l_0, l_1, l_2, ...l_n\}$	Sets of elements associated with cable lengths and types.
	$D = \{c, \tau, \phi\}$	Amplitude, mode, and phase of a cable reflection fit across the frequency band.
	$T = \{\rho_0, \rho_1, \rho_2, ...\rho_n\}$	Observation timing sets of physical time separations, such as pointings.
	$[f, t]$	A combined set of all frequencies and times.
Groups, Matrices, & Vectors	$\boldsymbol{\alpha}_L$	An element grouping, where parameters are per element group rather than per element.
	$\boldsymbol{\theta}_T$	A grouping of observation times, where parameters are per timing group rather than per observation time.
	$\mathbf{G}_{a,p}(f_o, f_i)$	The full gain matrix for each element in group A per p where input and output frequencies are correlated.
	$\mathbf{g}_{a,p}(f)$	A vectorised approximation of the gains $\mathbf{G}$ for each element in the group A per P over frequency.
	$\mathbf{m}_{ab,pq}([f, t])$	A vector of the simulated model visibilities from a model sky with frequency-dependent beam effects for each element pair ab and polarisation product pq over the set $[f, t]$.
	$\mathbf{n}_{ab,pq}([f, t])$	A vector of the thermal noise for each element pair ab and polarisation product pq over the set $[f, t]$.
	$\mathbf{v}_{ab,pq}([f, t])$	A vector of the uncalibrated data visibilities for each element pair ab and polarisation product pq over the set $[f, t]$.
	$\mathbf{a}_{a,p}(f)$	The calculated auto-gain for each element in the group A per p over frequency.
	$\boldsymbol{\eta}_{a,p}(f)$	Scaling relation of the discrepancy between the measured sky and the calibration model in the cross-correlations.
	$\hat{\mathbf{a}}_{a,p}(f)$	The calculated auto-gain for each element in the group A per p over frequency which has been scaled to match the cross-correlations.
Functions	$\mathcal{P}(\,)$	A polynomial fit as a function of frequency of the input.
	$\mathcal{R}\langle\,\rangle$	A resistant mean of the input vector over element set L (and optionally a time group θ_T). Outliers beyond 2σ are excluded in the average.
Discrete Fourier Transforms	k	The Nyquist frequency index of the Fourier dual of frequency.
	κ	The hyperfine sub-Nyquist frequency index of the Fourier dual of frequency, with resolution at $1/20^{\text{th}}$ of k.
	n	The index of the frequency.
	N	The total number of frequency channels.
	τ_κ	The Fourier dual of frequency: a timing delay in the detection of the waveform between one element and another. The κ index indicates it runs over the hyperfine index.

We can now fully vectorise the variables in Equation 5:

$$\mathbf{v}_{ab,pq}([f,t]) \sim \mathrm{diag}(\mathbf{g}_{a,p}(f))\mathrm{diag}(\mathbf{g}^*_{b,q}(f))\mathbf{m}_{ab,pq}([f,t]) + \mathbf{n}_{ab,pq}([f,t]). \quad (7)$$

The gains calculated from solving Equation 7 will encode differences between the model visibilities and the true visibilities. This is a systematic; the gains will be contaminated with a non-instrumental contribution. We can reduce this effect with long-baseline arrays by excluding short baselines from calibration since we lack a diffuse-emission model. However, this can potentially introduce systematic biases at low λ (Patil et al. 2016). For the remainder of Section 3.4, we describe our attempts to remove the systematics encoded within the gains due to an imperfect model.

3.4.1. Per Frequency Solutions

Equation 7 can be used to solve for the instrumental gains for all frequencies and polarisations independently. This allows the use of Alternating Direction Implicit (ADI) methods for fast

and efficient solving of $\mathcal{O}(N^2)$ (Mitchell et al. 2008; Salvini & Wijnholds 2014). Due to this independence, parallelisation can also be applied. Noise is ignored during the ADI for simplicity; the least-squares framework is the maximum likelihood estimate for a Gaussian distribution. However, if any noise has a non-Gaussian distribution, it will affect the instrumental gains.

We begin solving Equation 7 by estimating an initial solution for $\mathbf{g}^*_{b,q}(f)$ to force the gains into a region with a local minimum. Reliable choices are the average gain expected across all elements or scaled auto-correlations.

With an input for $\mathbf{g}^*_{b,q}(f)$, Equation 7 can then become a linear least-squares problem.

$$\chi^2_{a,p}([f,t]) = \sum_b \Big| \mathbf{v}_{ab,pq}([f,t]) - \mathrm{diag}(\mathbf{g}_{a,p}(f))\,\mathrm{diag}(\mathbf{g}^*_{b,q}(f))\,\mathbf{m}_{ab,pq}([f,t]) \Big|^2, \quad (8)$$

where $\mathbf{g}_{a,p}(f)$ is found given a minimisation of $\chi^2_{a,p}([f,t])$ for each element a and instrumental polarisation p. All time steps are used to find the temporally constant gains over the observation. For computation efficiency, we have also assumed $P = \{p_a p_b, q_a q_b\}$ (e.g., XX and YY in linear polarisation) since these contributions are most significant. Optionally, a full polarisation treatment can be used given polarised calibration sources.

The current estimation of $\mathbf{g}^*_{b,q}(f)$ is then updated with knowledge from $\mathbf{g}_{a,p}(f)$ by adding together the current and new estimation and dividing by 2. By updating in partial steps, a smooth convergence is ensured. The linear least-squares process is then repeated with an updated $\mathbf{g}^*_{b,q}(f)$ until convergence is reached.[e]

3.4.2. Bandpass

The resulting gains from the least-squares iteration process are fully independent in frequency, element, observation, and polarisation. This is not a completely accurate representation of the gains. It was necessary to make this assumption for the efficient solving technique in Section 3.4.1, but we can incorporate our prior knowledge of the nature of the instrument and its spectral structure *ex post facto*.

For example, we did not account for noise contributions during the per-frequency ADI fit; this adds spurious deviations from the gain's true value with mean of zero. Historically, we accounted for these effects by creating a global bandpass,

$$\left|\mathbf{g}_p(f;\boldsymbol{\alpha})\right| = \left\langle \left|\mathbf{g}_{a,p}(f)\right| \right\rangle_{\boldsymbol{\alpha}}, \quad (9)$$

where the normalised amplitude average is taken over all elements $\boldsymbol{\alpha}$ as a function of frequency to create a global bandpass $\left|\mathbf{g}_p(f;\boldsymbol{\alpha})\right|$ independent of elements. Figure 2 shows an example of the global bandpass alongside the noisy per-frequency inputs for the MWA.

This methodology drastically reduces noise contributions to the bandpass when there are many elements, and it reduces spectral structure contributions due to imperfections in the model (Barry et al. 2016). However, this averaging implicitly assumes that all elements are identical. We must use other schemes to capture more instrumental parameters while maintaining spectral smoothness in the bandpass.

The level of accuracy required in the calibration to feasibly detect the EoR is 1 part in 10^5 as a function of frequency

[e]We have found that allowing the first 10 iterations to only update the phase of $\mathbf{g}^*_{b,q}(f)$ helps to converge faster.

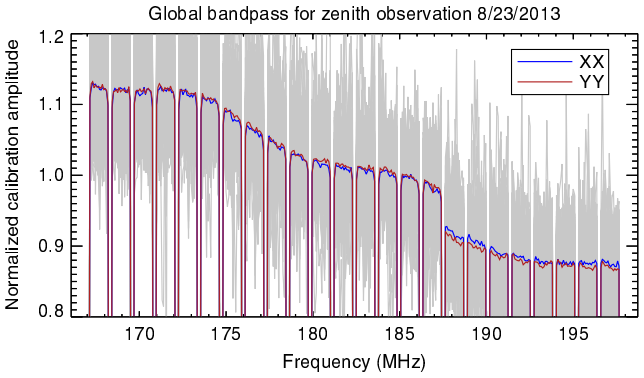

Figure 2. The MWA global bandpass for the zenith observation of 2013 August 23 for polarisations $pp = XX$ (blue) and $qq = YY$ (red). All the per-frequency antenna solutions used in the global bandpass average are shown in the background (grey). This historical approach greatly decreased expected noise on the solutions.

(Barry et al. 2016). As seen in Figure 2, instruments can be very spectrally complicated. Therefore, more sophisticated calibration procedures must be used.

Usually, these schemes are instrument-specific. For example, in the MWA, sets of elements experience the same attenuation as a function of frequency due to cable types, cable lengths, and whitening filters. We group these elements into different cable length sets $L = \{l_0, l_1, l_2, ...l_n\}$. We get

$$\left|\mathbf{g}_{l,p}(f;\boldsymbol{\alpha}_l)\right| = \mathcal{R}\left\langle \left|\mathbf{g}_{a\in L,p}(f)\right|, 2\sigma \right\rangle, \quad (10)$$

where $\mathcal{R}$ is the resistant mean function[f] calculated over each element set $\boldsymbol{\alpha}_L$ for each polarisation and frequency. We choose the resistant mean because outlier contributions are more reliably reduced than median calculations. The variable change from a to $\boldsymbol{\alpha}$ indicates one parameter per group of elements.

If more observations are available, we follow a similar averaging process over time. If the instrument is stable in time, a normalised bandpass per antenna should be nearly identical from one time to the next excluding noise contributions and potential Van Vleck quantisation corrections (Vleck & Middleton 1966). Gains from different local sidereal time (LST) will have different spectral structure from unmodelled sources, and thus an average will remove even more of this effect.

We create a time set $T = \{\rho_0, \rho_1, \rho_2, ...\rho_n\}$ where times are grouped by physical time separations based on the instrument. For example, we group the MWA observations by pointings[g] because they sample different beam errors, and thus averaging between pointings would remove this instrumental feature. Whenever possible, we use

$$\left|\mathbf{g}_{l,p}(f;\boldsymbol{\alpha}_l,\boldsymbol{\theta}_\rho)\right| = \mathcal{R}\left\langle \left|\mathbf{g}_{a\in L,p}(f)\right|, t\in T, 2\sigma \right\rangle, \quad (11)$$

where $\boldsymbol{\theta}_\rho$ runs over observations within physical time separations in the set T and over as many days as applicable. The variable change from t to $\boldsymbol{\theta}$ indicates one parameter per group of times.

[f]The resistant mean function calculates the distribution of the amplitudes of a similar element set $\boldsymbol{\alpha}_l$ for each frequency and polarisation and then calculates the mean of that distribution after Gaussian 2σ outliers have been excluded.

[g]A pointing defines a group of observations with the same electronic delay. As the sky rotates throughout the night, different electronic delays are used to roughly point the instrument to the same location in the sky.

3.4.3. Low-order polynomials

An overall amplitude due to temperature dependence **T** in the amplifiers must still be accounted for within the gains. These differ from day to day and element to element, therefore they cannot be included in the average bandpass. For well-behaved amplifiers, this varies slightly as a function of frequency and is easily characterised with a low-order polynomial for each element and observation.

In addition to fitting polynomials to the amplitude as a function of frequency, we must also account for the phase. There is an inherent per-frequency degeneracy that cannot be accounted for by sky-based calibration. Therefore, we reference the phase to a specific element to remove this degeneracy, which makes the collective phases smooth as a function of frequency for well-behaved instruments. We have found that using a per-element polynomial fit as the calibration phase solution has been a reliable estimate.

For the amplitude, we fit

$$c_n f^n + ... + c_1 f + c_0 = \mathcal{P}\left(\frac{|\mathbf{g}_{a,p}(f)|}{|\mathbf{g}_p(f;\boldsymbol{\alpha})|}\right), \tag{12}$$

where the bandpass contribution, $|\mathbf{g}_p(f;\boldsymbol{\alpha})|$, is removed before the fit and c_n are the resulting coefficients. Any bandpass contribution can be used; $|\mathbf{g}_p(f;\boldsymbol{\alpha})|$ is just an example. For the phase, we fit

$$\phi_n f^n + \cdots + \phi_1 f + \phi_0 = \mathcal{P}\left(\arg \mathbf{g}_{a,p}(f)\right), \tag{13}$$

where the polynomial fit is done over the phase of the residual and ϕ_n are the resulting coefficients. Due to phase jumps between $-\pi$ and π, special care is taken to ensure the function is continuous across the π boundary.[h] We can create a set of these coefficients, $C = \{c_n, ..., c_1, c_0, \phi_n, ..., \phi_1, \phi_0\}$, for easy reference.

In all cases, we choose the lowest-order polynomials possible for these fits on the amplitude and phase. For example, the MWA is described by a 2nd-order polynomial fit in amplitude and a linear fit in phase. As a general rule, we do not fit polynomials which have modes present in the EoR window.

3.4.4. Cable reflections

Reflections due to a mismatched impedance must also be accounted for within the gains (contribution **K** in the signal path framework). Even though instrumental hardware is designed to meet engineering specifications, residual reflection signals are still orders of magnitude above the EoR and are different for each element signal path. Averaging the gains across elements and time in Section 3.4.2 artificially erased cable reflections from the solutions, thus we must specifically incorporate them. Currently, we only fit the cable reflection for elements with cable lengths that can contaminate prime locations within the EoR window in power spectrum space (Beardsley et al. 2016; Ewall-Wice et al. 2016).

We find the theoretical location of the mode using the nominal cable length and the specified light travel time of the cable. We then perform a hyperfine discrete Fourier transform of the gain around the theoretical mode

$$\mathbf{g}_{a,p}(\tau_\kappa) = \sum_{m=0}^{M-1}\left(\frac{\mathbf{g}_{a,p}(f_m)}{|\mathbf{g}_p(f;\boldsymbol{\alpha})|} - \left(c_n f_m^n + ... + c_0\right) e^{i\left(\phi_n f_m^n + ... + \phi_0\right)}\right) e^{-2\pi i \kappa \frac{m}{M}}, \tag{14}$$

where τ_κ is the delay, $[m, M] \in \mathbb{Z}$, and κ is the hyperfine index component. Typically, we set the range of κ to be $[k_{\tau_0} - \frac{1}{20}k, k_{\tau_0} + \frac{1}{20}k]$, where k_{τ_0} is the index of the theoretical mode and k is the index in the range of $[0, M-1]$ (Beardsley 2015). Again, any bandpass contribution can be used; $|\mathbf{g}_p(f;\boldsymbol{\alpha})|$ is just an example.

The maximum $|\mathbf{g}_{a,p}(\tau_\kappa)|$ around k_{τ_0} is chosen as the experimental cable reflection. The associated amplitude c, phase ϕ, and mode τ are then calculated to generate the experimental cable reflection contribution $ce^{-2\pi i \tau f + i\phi}$ to the gain for each observation. Optionally, these coefficients can be averaged over sets of observations and times, depending on the instrument. We can create a set of these coefficients, $D = \{c, \tau, \phi\}$, for easy reference.

3.4.5. Auto-correlation bandpass

The fine-frequency bandpass is a huge potential source of error in the power spectrum (Barry et al. 2016). As such, we have tried averaging along various axes to reduce implicit assumptions while maintaining spectral smoothness in Section 3.4.2. However, there always remains an assumption of stability along any axis that we average over, and this may not be valid enough for the level of required spectral accuracy.

We can instead use the auto-correlations to bypass these assumptions. An auto-correlation should only contain information about the total power on the sky, instrumental effects, and noise. No information about structure on the sky is encoded, so spectral effects from unmodelled sources do not contribute. However, there are two main issues with using auto-correlations as the bandpass: improper scaling and correlated noise.

- The scale of the auto-correlations are dominated by the largest modes on the sky. Currently, we do not have reliable models of these largest modes. Either a large-scale calibration model must be obtained, or the scaling must be forced to match the cross-correlations.
- Thermal noise is correlated in an auto-correlation and will contribute directly to the solutions. Truncation effects during digitisation and channelisation can artificially correlate visibilities. This will result in bit noise, which will also contribute to the auto-correlations.

Whether or not auto-correlations can effectively be used for the bandpass depends on the importance of their errors in power spectrum space and our ability to mitigate these errors. We solve for the auto-correlation gains via the relation,

$$\mathbf{a}_{a,p}(f) = \sqrt{\frac{\left\langle \mathbf{v}_{aa,pp}([f,t])\right\rangle_t}{\left\langle \mathbf{m}_{aa,pp}([f,t])\right\rangle_t}}, \tag{15}$$

where $\mathbf{a}_{a,p}(f)$ is the auto-gain for element a and polarisation p as a function of frequency.

To correct for the scaling error and for the correlated thermal noise floor, we define a cross-correlation scaling relation,

$$\boldsymbol{\eta}_{a,p}(f) = \frac{\mathcal{P}\left(|\mathbf{g}_{a,p}(f;\boldsymbol{\alpha}_l, \boldsymbol{\theta}_\rho, [C], [D])|\right)}{\mathcal{P}\left(\mathbf{a}_{a,p}(f)\right)}, \tag{16}$$

where $\mathcal{P}$ is a polynomial fit to the gains, usually limited to just a linear fit. This captures the discrepancy between the measured sky and the calibration model in the space of one

[h] We 'unwrap' the phase by creating a new continuous plane from the Riemann sheets. We then solve and 'rewrap'. If the phase varies quickly, there can be ambiguity in which Riemann sheet to place the phase, but this is not an issue with most instruments.

cross-correlation gain. We then rescale the auto-gains to match the cross-correlations,

$$\hat{\mathbf{a}}_{a,p}(f) = \boldsymbol{\eta}_{a,p}(f)\,\mathbf{a}_{a,p}(f), \tag{17}$$

where $\hat{\mathbf{a}}_{a,p}(f)$ is the scaled auto-gain for element a and polarisation p as a function of frequency.

We have reduced scaling errors and correlated thermal noise on the auto-gain solutions. However, we have not corrected for bit noise, or a bias caused by bit truncation at any point along the signal path. These errors are instrument-specific and will persist into power spectrum space. If the bits in the digital system are not artificially correlated due to bit truncation, then the auto-correlation can yield a better fine-frequency bandpass.

3.4.6. Final calibration solutions

Our final calibration solution using only cross-correlations is

$$\mathbf{g}_{a,p}(f;\boldsymbol{\alpha}_l,\boldsymbol{\theta}_\rho,[C],[D]) = \underbrace{\left|\mathbf{g}_{l,p}(f;\boldsymbol{\alpha}_l,\boldsymbol{\theta}_\rho)\right|}_{\text{bandpass Section 3.4.2}} \underbrace{\left((c_n f^n + \ldots + c_0)e^{i(\phi_n f^n + \ldots + \phi_0)}\right)}_{\text{low-order polynomials Section 3.4.3}} + \underbrace{ce^{-2\pi i\tau f + i\phi}}_{\text{reflection Section 3.4.4}}, \tag{18}$$

for elements with fitted cable reflections, and

$$\mathbf{g}_{a,p}(f;\boldsymbol{\alpha}_l,\boldsymbol{\theta}_\rho,[C]) = \underbrace{\left|\mathbf{g}_{l,p}(f;\boldsymbol{\alpha}_l,\boldsymbol{\theta}_\rho)\right|}_{\text{bandpass Section 3.4.2}} \underbrace{(c_n f^n + \ldots + c_0)e^{i(\phi_n f^n + \ldots + \phi_0)}}_{\text{low-order polynomials Section 3.4.3}}, \tag{19}$$

for all other elements. The bandpass amplitude solution is generated over a set of elements with the same cable/attenuation properties ($\boldsymbol{\alpha}_l$) and includes many observations within a timing set ($\boldsymbol{\theta}_\rho$) covering many days. The same bandpass is applied to all elements of the appropriate type and all observing times from the respective time set. In contrast, the polynomials and the cable reflection are fit independently for each observation and per element. We divide the data visibilities by the applicable form of $\mathbf{g}_a\mathbf{g}_b^*$ to form our final, calibrated visibilities.

Alternatively, our final calibration solution harnessing the auto-correlations is

$$\mathbf{g}_{a,p}(f;\boldsymbol{\alpha}_l,\boldsymbol{\theta}_\rho,[C],[D],\hat{\mathbf{a}}) = \underbrace{\hat{\mathbf{a}}_{a,p}(f)}_{\text{auto-gain Section 3.4.5}} \underbrace{e^{i\arg \mathbf{g}_{a,p}(f;\boldsymbol{\alpha}_l,\boldsymbol{\theta}_\rho,[C],[D])}}_{\text{phase from Eq. 18}}, \tag{20}$$

where the amplitude is described by the scaled auto-gains and the phase is described by the cross-gains.

For EoR science, we want to reduce spectral structure as much as possible but still capture instrumental parameters. Our auto-gain calibration solution performs the best *in the most sensitive, foreground-free regions*. Therefore, we currently use Equation 20 in creating EoR upper limits with the MWA. This is an active area of research within the EoR community; reaching the level of accuracy required to detect the EoR is ongoing.

3.5. Imaging

The final stage of FHD transforms calibrated data visibilities into a space where they can be combined across observations. To begin this process, we perform an operation called gridding on the visibilities.

We take each complex visibility value and multiply it by the corresponding *uv*-beam in the coherency domain calculated in Section 3.2 through a process called Optimal Map-Making. In essence, this takes visibility values integrated by the instrument and estimates their original spatial *uv*-response given our knowledge of the beam. These are approximations of the instrument power response for each baseline to a set of regular gridding points (e.g. Myers et al. 2003; Bhatnagar et al. 2008; Morales & Matejek 2009; Sullivan et al. 2012; Dillon et al. 2014; Shaw et al. 2014; Zheng et al. 2017).

We separately perform gridding to individual *uv*-planes for calibrated data visibilities, model visibilities, and residual visibilities generated from their difference. By gridding each separate data product, we can make a variety of diagnostic images. For example, images generated from residual visibilities help to ascertain the level of foreground removal in image space, and thus are important for quality assurance.

We also grid visibilities of value 1 with the beam gridding kernel to create natural *uv*-space weights. This generates a sampling map which describes how much of a measurement went into each pixel. In addition, we separately grid with the beam-squared kernel to create a variance map. The variance map relates to the uncertainty for each *uv*-pixel, which will be a vital component for end-to-end error propagation in εppsilon.

We then have three types of *uv*-plane products: the sampling map, the variance map, and the data *uv*-planes. From these, we can create weighted-data *uv*-planes using the sampling map and data *uv*-planes. All these products are transformed via 2D FFTs to image space in slant orthographic projection.[i]

At this point, the various images are made for two different purposes: (1) the sampling map, variance map, and data image planes are for power spectrum packages and (2) the weighted-data image plane is for diagnostic images per observation. The result of calibrated data and residual snapshot images for a zenith observation with Phase I of the MWA is shown in Figure 3.

Our slant orthographic images are in a basis that changes with LST. Therefore, we perform a bilinear interpolation to HEALPix pixel centres[j] (Górski et al. 2005), which are the same for all LSTs. We interpolate the calibrated data, model, sampling map, and variance map to HEALPix pixel centres *separately* for use in εppsilon. If we want to combine multiple observations, we will need to do a weighted average. Therefore, we keep the numerator (data) separate from the denominator (sampling map) for this purpose.

3.6. Interleaved cubes

The images that will be used for power spectrum analysis are split by interleaved time steps, grouped by even index and odd index. The even–odd distinction is arbitrary; what really matters is that they are interleaved. While this doubles the number of Fourier transforms to perform, it allows for crucial error analysis.

The sky should not vary much over sufficiently small integration intervals. Any significant variation can be attributed to either RFI (which has been accounted for in Section 3.1) or thermal noise on the observation. By subtracting the even–odd groupings and

[i] Slant orthographic projection: a flat projection of the sky that is slanted to be parallel with the measurement plane.

[j] HEALPix: the Hierarchical Equal Area isoLatitude Pixelisation of a sphere.

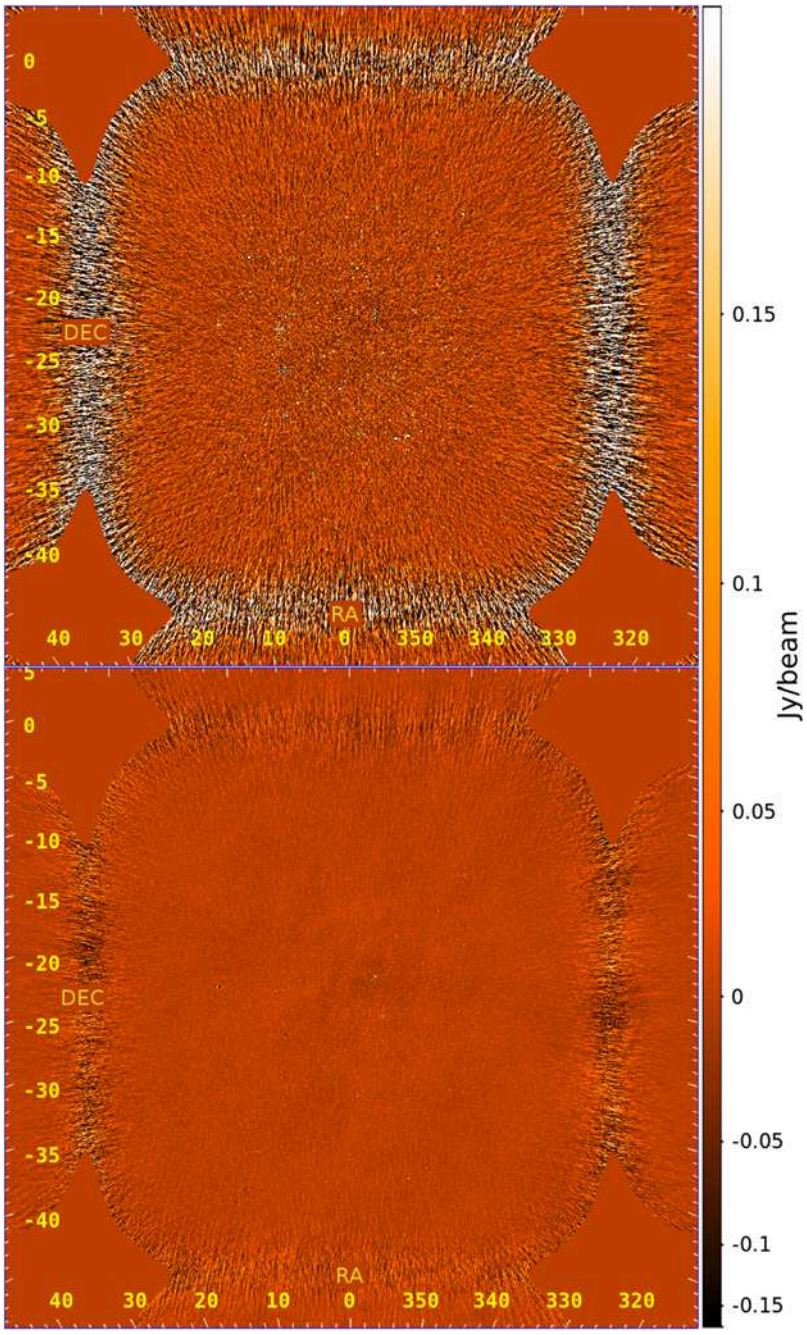

Figure 3. An example of images output from FHD: calibrated data (top) and residual (bottom) Stokes I images from the MWA for the zenith observation of 2013 August 23. There is significant reduction of sources and point spread functions in the residual. However, the diffuse synchrotron emission can be seen in the residual because it was not in the subtraction model.

enforcing consistent flagging, we should be left with the thermal noise contribution to the observation.

We carry even–odd interleaved cubes throughout the power spectrum analysis and check at various stages against noise calculations. It is a robust way to ensure that error propagation and normalisation have occurred correctly, and it allows us to build more diagnostic data products while still building the cross-power spectrum (see Section 5.4).

Thus, there are at least eight images per polarisation product output from FHD for use in a power spectrum analysis: the calibrated data, model, sampling map, and variance map for each even–odd set.

4. Integration

In order to reduce noise to reach the signal-to-noise required to detect the EoR, we must now integrate. This requires integration of thousands of observations (Beardsley et al. 2013).

FHD outputs a variety of data cubes that we must integrate together before input into our next package, εppsilon (described in Section 5). These RA/Dec frequency slices are numerators and denominators which create meaningful maps in *uv*-space. The various cubes are:

- **Calibrated data cubes**: HEALPix images for each frequency of the unweighted calibrated data.

- **Model data cubes**: HEALPix images for each frequency of the unweighted generated model.
- **Sampling map cubes**: HEALPix images for each frequency of the sampling estimate.
- **Variance map cubes**: HEALPix images for each frequency of the variance estimate.

These cube types are split by instrumental polarisation and by an interleaved time sampling set of even and odd time indices (detailed in Section 3.6). Each even–odd polarisation grouping is added across observations, pixel by pixel, using the HEALPix coordinate system. Averaging *in image space*, rather than in *uv*-space, is a crucial aspect of the analysis. Either approach can theoretically be used, but the computational requirements vary greatly.

The measurement plane is not parallel to the tangent plane of the sky for non-zenith measurements. The instrument's measurement plane appears tilted compared to the wave's propagation direction; this tilt causes a measurement delay. Since each measurement plane will measure different phases of the wave propagation, different *uv*-planes cannot be coadded. This is a classic decoherence problem in Fourier space.

There are two methodologies to account for this decoherence. The first is to project the measurement *uv*-plane to be parallel to the tangent plane of the sky. This is called *w*-projection since it projects $\{u, v, w\}$-space to $\{u, v, w = 0\}$-space (Cornwell, Golap, & Bhatnagar 2008). Unfortunately, the projection requires the propagation of a Fresnel pattern for every visibility to reconstruct the wave on the $w = 0$ plane. This requires intense computational overhead, but has been successfully implemented by other packages (Trott et al. 2016).

The second methodology is to integrate in image space. We create an image for each observation by performing a 2D spatial FFT of the *uv*-plane, which results in a slant orthographic projection of the sky. This inherently assumes the array is coplanar; the integration time must be relatively small and altitude variations must be insignificant or absorbed into unique kernel generation (Section 3.2). For the MWA Phase I, a mean *w*-offset of 0.15 λ in the imaged modes is small enough to not be a dominating systematic for current analyses.

We can then easily interpolate from the slant orthographic projection to the HEALPix projection (Ord et al. 2010), which is a constant basis. We do not need to propagate waves in image space in order for coherent integration, thus it is computationally efficient in comparison.

5. Error propagated power spectrum with interleaved observed noise

Error Propagated Power Spectrum with Interleaved Observed Noise, or εppsilon,[k] is an open source power spectrum analysis package designed to take integrated images and create various types of diagnostic and limit power spectra. It was created as a way to propagate errors into power spectrum space, rather than having *estimated* errors. Other image-based power spectrum analyses exist for radio interferometric data (Paciga et al. 2011; Shaw et al. 2014; Patil et al. 2014; Shaw et al. 2015; Dillon et al. 2015; Ewall-Wice et al. 2016; Patil et al. 2017); however, the end-to-end error propagation of εppsilon and CHIPS (Trott et al. 2016) is uncommon.

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

The main functions of εppsilon are to transform integrated images into $\{u, v, f\}$-space (Section 5.1), calculate observed noise using even–odd interleaving (Section 5.2), transform frequency to *k*-space (Section 5.3), and average *k*-space voxels together for diagnostic and limit power spectra (Section 5.4). Detailing these processes will build the groundwork needed to describe 2D power spectra, 1D power spectra, and their respective uncertainty estimates in Section 6.

5.1. From integrated images to uv-space

The input products of εppsilon are integrated image cubes per frequency. While this space was necessary for integration, the error bars in image space are complicated; every pixel is covariant with every other pixel. Our measurements were inherently taken in *uv*-space, and that is where our uncertainties on the measurement are easiest to propagate.

We then Fourier transform the integrated image back into *uv*-space using a direct Fourier transform between the curved HEALPix sky and the flat, regularly spaced *uv*-plane. Since this happens once per observation integration *set*, rather than once per observation, the slower direct Fourier transform calculation is computationally feasible. The modified gridding kernel described in Section 3.2 acts as a window in the HEALPix image. In order to propagate our uncertainties correctly, this must be done during the generation of the kernel.

After transforming into *uv*-space, we calculate the resulting $\{k_x, k_y\}$-values for each pixel. The *uv*-space is related to the wavenumber space by the simple transforms (Morales & Hewitt 2004)

$$k_x = \frac{u2\pi}{D_M(z)} \qquad k_y = \frac{v2\pi}{D_M(z)}, \tag{21}$$

where $D_M(z)$ is the transverse comoving distance dependent on redshift (Hogg 1999).

In addition, we also perform the 3D pixel-by-pixel subtraction of the integrated model from the integrated data to create residual cubes. We could instead perform this after we have transformed f to k_z to save on computation time, but it is helpful to make diagnostic plots of the residual cube in $\{u, v, f\}$-space. By creating diagnostics at every important step in the analysis, we can better understand our systematics.

5.2. Mean and noise calculation

We now have twenty various *uv*-products: calibrated, model, residual, sampling map, and variance map data, for each polarisation product *pp* and *qq*, and for each interleaved even–odd time sample set. We have kept the numerators and denominators (data and weights) separate up until this point so that we can perform variance-weighted sums and differences.

First, we weight the calibrated data, model, and residual by the sampling map, thereby upweighting well-measured modes and downweighting poorly measured modes. Second, we weight the variance map by the square of the sampling map to scale our uncertainty estimates with our choice of weighting scheme. We then perform sums and differences using our sampling-map-weighted uncertainty estimates as the weights,

$$\hat{\mu} = \frac{\frac{x_e}{\sigma_e^2} + \frac{x_o}{\sigma_o^2}}{\frac{1}{\sigma_e^2} + \frac{1}{\sigma_o^2}} \qquad \hat{n} = \frac{\frac{x_e}{\sigma_e^2} - \frac{x_o}{\sigma_o^2}}{\frac{1}{\sigma_e^2} + \frac{1}{\sigma_o^2}}, \tag{22}$$

where $\hat{\mu}$ is the mean, $\hat{n}$ is the noise, $\{e, o\}$ are the interleaved even–odd sets, x is the sampling-map-weighted data (calibrated data, model, or residual) for a given even–odd set, and σ^2 is the sampling-map-weighted variance map for a given even–odd set. The calculated mean and noise are the maximum likelihood estimates for a weighted Gaussian probability distribution, hence using the variables $\hat{\mu}$ and $\hat{n}$ instead of μ and n.

Finally, we also calculate our uncertainty estimates using the maximum likelihood estimation:

$$\hat{\epsilon}^2 = \frac{1}{\frac{1}{\sigma_e^2} + \frac{1}{\sigma_o^2}}, \tag{23}$$

where $\{e, o\}$ are the interleaved even–odd sets and σ^2 is the sampling-map-weighted variance map for a given even–odd set.

These analytical uncertainty estimates are propagated through the analysis to create our uncertainty on the cross power spectrum. We assume that there are no cross-correlations in the noise; each noise pixel in uv-space is assumed to be independent. Various techniques, including spatial/frequency windowing, increases pixel-to-pixel correlation. We investigate the strength of the correlations by comparing the analytic noise propagation to the observed noise in the power spectrum in Section 6.1.

5.3. Transforming frequency to k-space

We now have a variety of cubes in $\{k_x, k_y, f\}$-space. In order to go to power spectrum space, we must perform a spectral transform in the frequency direction to go from f to k_z. Due to the nature of the data, this includes several steps.

While the binned data are regularly spaced as a function of frequency, the sampling distribution is not constant. We have flagged channels due to RFI in Section 3.1 and due to channeliser aliasing in Section 3.4. The sampling of the uv-plane is also inherently non-regular as a function of frequency due to baseline length evolving with frequency. As a result, the sin and cos basis functions of the Fourier transform are not orthogonal to the noise distribution on our sampling; the noise is not independent in the sin and cos basis. However, we can find a basis which is orthogonal.

We use the Lomb–Scargle periodogram to find this basis (Lomb 1976; Scargle 1982). A rotation phase is found for each spectral mode that creates orthogonal cos-like and sin-like eigenfunctions in the noise distribution. This periodogram effectively erases the phase of the data, though this is suitable given that our final goal is to create a power spectrum (a naturally phase-less product).

The spectral transformation of a relatively small, finite set of data will cause leakage. We mitigate this by multiplying the data by a Blackman–Harris window function in frequency. This will decrease the leakage by about 70 dB at the first sidelobe, but at the cost of half the effective bandwidth.

The uneven sampling as a function of frequency can also cause leakage from the bright foregrounds into higher k_z. To reduce leakage from the uneven sampling and from the finite spectral transform window, we remove the mean of the data before applying the window function. To preserve the DC term in the power spectrum, we add this value back to the zeroth mode after our spectral transform. This preserves power while improving dynamic range.

We do not expect this mean subtraction and reinsertion to affect our measured EoR power spectrum because the vast majority of the DC-mode power is from foregrounds. In addition, the zeroth k_z-mode is not included in our EoR estimates. All our signal loss simulations include this technique to verify that there is no unexpected effect on higher k_z-modes.

The periodogram dual, η, is related to wavenumber space through

$$k_z \approx \frac{2\pi H_0 f_{21} E(z)}{c(1+z)^2}\eta, \tag{24}$$

where c is the speed of light, z is redshift, H_0 is the Hubble constant in the present epoch, f_{21} is the frequency of the 21 cm emission line, $E(z)$ describes how H_0 evolves as a function of redshift, and k_z is the wavenumber along the line-of-sight. This is an approximation; it is valid for all reasonable parameters and avoids a frequency kernel (Morales & Hewitt 2004).

We have chosen our power estimator to be along k_z rather than along the time delay of the electric field propagation between elements k_τ. This subsequently determines *how* the foreground contaminates our final power spectrum (Morales et al. 2018). Given our choice of basis, we are constructing an *imaging* (or reconstructed sky) power spectrum.

5.4. Power spectrum products

We now have the power spectra estimates for the mean, the noise, and the uncertainty estimates as a function of $\{k_x, k_y, k_z\}$.

We construct our power spectra by subtracting the power of the even–odd difference from the power of the even–odd summation, and dividing by 4:

$$p = \frac{p_{\hat{\mu}} - p_{\hat{n}}}{4} = \frac{\hat{\mu}^2 - \hat{n}^2}{4} = \frac{\frac{x_e}{\sigma_e^2}\frac{x_o}{\sigma_o^2}}{\frac{1}{\sigma_e^2} + \frac{1}{\sigma_o^2}}. \tag{25}$$

This gives us the same result as the cross power between the even and odd cubes, and the additional products also allow us to build more diagnostics and to carry the noise throughout the pipeline.

We would like to perform averages over these cubes to generate the best possible limits and to generate diagnostics. To do so, we must assume that the EoR is spatially homogeneous and isotropic, which allows for spherical averaging in Fourier space (Morales & Hewitt 2004). Much like the creation of even–odd sum and difference cubes, we perform the weighted average of pixels. The only difference is that the variances σ^2 are no longer Gaussian, but rather Erlang variances which can be propagated from the original variances.

We can now create various power spectrum products in 2D and 1D with our 3D power spectrum cube. Uncertainty estimates, measured noise contributions, and expectation values that we have generated in εppsilon will also be used in our diagnostics.

6. Power spectrum diagnostics

We create a variety of diagnostic power spectrum plots using various averaging schemes. These are essential to understanding contributions to the power spectrum and are vital to assessing changes to the analysis. In this section, we use MWA Phase I data as a specific example; however, all of these diagnostic plots are part of any analysis with FHD/εppsilon.

6.1. 2D power spectrum

The most useful diagnostic we have is the 2D power spectrum. We average our $\{k_x, k_y, k_z\}$-power measurements along only the angular wavenumbers $\{k_x, k_y\}$ in cylindrical shells. The resulting

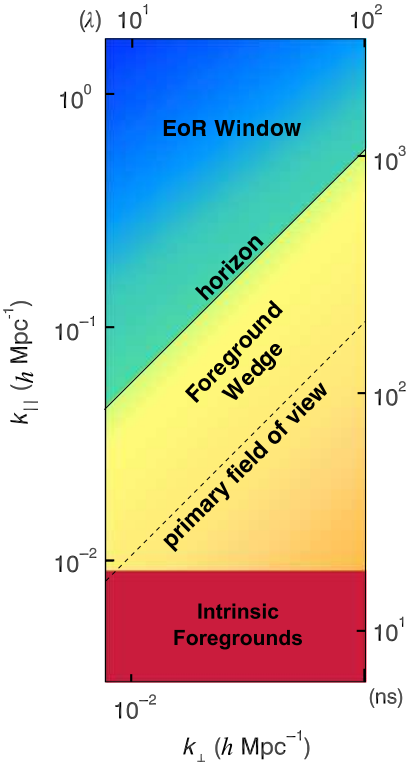

Figure 4. A schematic representation of a 2D power spectrum. Intrinsic foregrounds dominate low $k_\parallel$ (modes along the line-of-sight) for all $k_\perp$ (modes perpendicular to the line-of-sight) due to their relatively smooth spectral structure. Chromaticity of the instrument mixes foreground modes up into the foreground wedge. The primary-field-of-view line and the horizon line are contamination limits dependent on how far off-axis sources are on the sky. Foreground-free measurement modes are expected to be in the EoR window.

power spectrum is a function of modes perpendicular to the line-of-sight ($k_\perp$) and modes parallel to the line-of-sight ($k_\parallel$) shown in Figure 4. The $\{k_\parallel, k_\perp\}$-axes have been converted into $\{\tau, \lambda\}$ on the right and top axes—delay in nanoseconds and baseline length in wavelengths, respectively.

Wavenumber space is crucial for statistical measurements due to the spectral characteristics of the foregrounds. Diffuse synchrotron emission and bright radio sources, while distributed across the sky, vary smoothly in frequency (e.g. Matteo et al. 2002; Oh & Mack 2003). Only small $k_\parallel$-values are theoretically contaminated by bright, spectrally smooth astrophysical foregrounds. Since the foreground power is restricted to only a few low $k_\parallel$-modes, larger $k_\parallel$-values tend to be free of these intrinsic foregrounds in wavenumber space.

However, interferometers are naturally chromatic. This chromaticity distributes foreground power into a distinctive foreground wedge due to the mode-mixing of power from small $k_\parallel$-values into larger $k_\parallel$-values as illustrated in Figure 4 (Datta, Bowman, & Carilli 2010; Morales et al. 2012; Vedantham, Shankar, & Subrahmanyan 2012; Parsons et al. 2012; Trott, Wayth, & Tingay 2012; Hazelton, Morales, & Sullivan 2013; Thyagarajan et al. 2013; Pober et al. 2013; Liu, Parsons, & Trott 2014). The primary field-of-view line and the horizon line are the expected contamination limits caused by measured sources in the primary field-of-view and the sidelobes, respectively. The remaining region, called the EoR window, is expected to be contaminant-free. Because the power of the EoR signal decreases with increasing $k_{||}$, the most sensitive measurements are expected to be in the lower, left-hand corner of the EoR window.

The intrinsic foregrounds, the foreground wedge, the primary field-of-view line, the horizon line, and the resulting EoR window all have characteristic shapes in 2D power spectrum space. Thus, it is a very useful diagnostic space for identifying contamination in real data. We generate a 2D power spectrum as a function of $k_\perp$ and $k_\parallel$ for each of the calibrated data, model, and residual data sets in polarisations $pp = XX$ and $qq = YY$. Shown in Figure 5 is an example 2D power spectrum panel from a MWA integration of 2013 August 23.

We use this 2D power spectrum panel to help determine if expected contamination occurred in expected regions. In addition to the features shown in Figure 4, there is harmonic $k_\parallel$ contamination in the EoR window which is constant in $k_\perp$. This is caused from regular flagging of aliased frequency channels due to the polyphase filter banks in the MWA hardware, which creates harmonics in k-space.

In order to verify our error propagation, we also calculate observed and expected 2D noise power spectra. The expected noise, observed noise, propagated error, and noise ratio 2D power spectra are shown in Figure 6. The observed noise is the power spectrum of the maximum likelihood noise in Equation 22, so it is the realisation of the noise in the data. The expected noise and error are the expectation value and the square root of the variance, respectively, of the analytically propagated error distributions. We calculate these by propagating the initial Gaussian distributions through the full analysis:

$$\mathrm{Var}[N] = \frac{1}{\sum_{i=0}^{n} \frac{1}{4\sigma_i^4}} \qquad \mathrm{E}[N] = \frac{\sum_{i=0}^{n} \frac{1}{2\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{4\sigma_i^4}}, \tag{26}$$

where $\mathrm{Var}[N]$ is the variance on the noise, $\mathrm{E}[N]$ is the expected noise, n is the number of pixels in the average, and σ^2 is the original Gaussian variance.

The final plot in Figure 6 is the ratio of our observed noise to our expected noise. This investigates our assumption of uncorrelated pixels in uv-space during noise propagation. There will be fluctuations given different noise realisations on the observed noise, but the ratio should fluctuate around one. In practice, this is approximately true; the mean of the noise ratio from 10 to 50 λ is $\sim$0.9. The propagated noise is therefore slightly higher, indicating that we overestimate our noise due to assuming a lack of correlation.

6.2. 1D power spectrum

Averaging to a 1D power spectrum harnesses as much data as possible, thereby making the limits with the lowest noise. However, the 1D power spectrum also has the ability to be an excellent secondary diagnostic after the 2D power spectrum. While characteristic locations of contamination are easier to distinguish on 2D plots, 1D diagnostics are more able to distinguish subtleties.

The most simplistic 1D power spectrum is an average in spherical shells of all voxels in the 3D power spectrum cube which surpass a low-weight cutoff (see Appendix A for more details). This includes all areas of contamination explored in Section 6.1 which can obscure low-power regions. Therefore, a typical secondary diagnostic is a 1D power spectrum generated only from pixels which fall within 10 and 50 λ in $k_\perp$, shown in Figure 7.

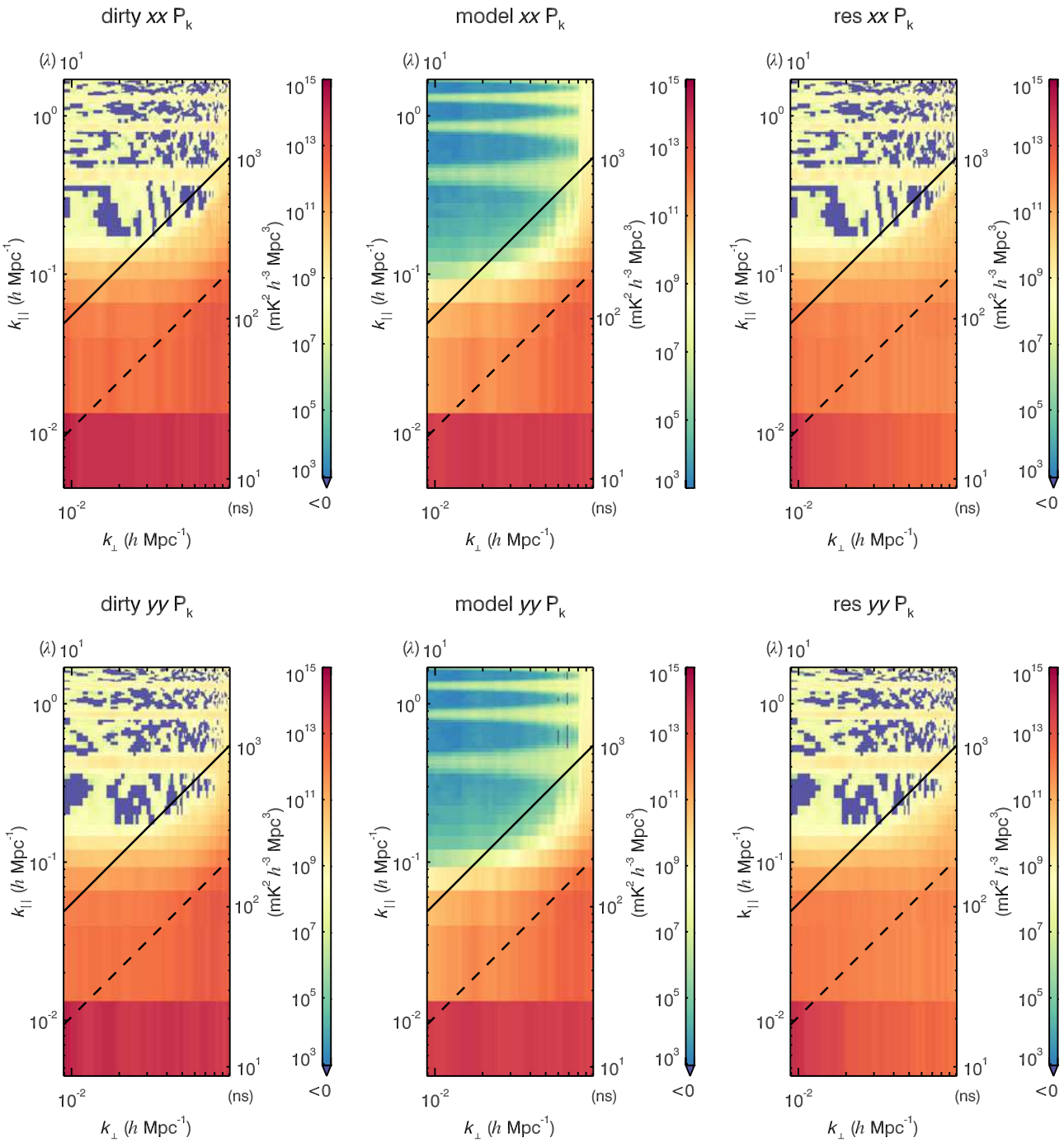

Figure 5. The 2D power spectra for the calibrated data, model, and residual for polarisations *XX* and *YY* of an integration of 64 MWA observations (∼2 h of data) from 2013 August 23. The characteristic locations of contamination are very similar to Figure 4, with the addition of contamination at $k_\parallel$ harmonics due to flagged frequencies with channeliser aliasing. Voxels that are negative due to thermal noise are dark purple-blue.

This will include intrinsic foregrounds and some of the foreground wedge, but will avoid contaminated $k_\perp$-modes with poor *uv*-coverage for the MWA. The exact $k_\perp$-region for this diagnostic will depend on the instrument.

The typical characteristic contamination shapes are present in Figure 7, like the intrinsic foregrounds, foreground wedge, and flagging harmonics, all of which are expected given the 2D power spectra. However, a new contamination feature occurs at about 0.7 h Mpc^{-1}. This contamination is from a cable reflection in the elements with 150 m cables. It was significantly reduced due to the calibration technique in Section 3.4.4, but spectral structure from unmodelled sources limit the precision (Barry et al. 2016). Much like with the aliased channel flagging, spectrally repetitive signals will appear as a bright contamination along $k_\perp$, translating to a near-constant 1D contribution if high in $k_\parallel$.

This highlights the potential of utilising multiple types of power spectra. The 2D power spectrum is a powerful diagnostic due to characteristic locations of contamination. However, it is

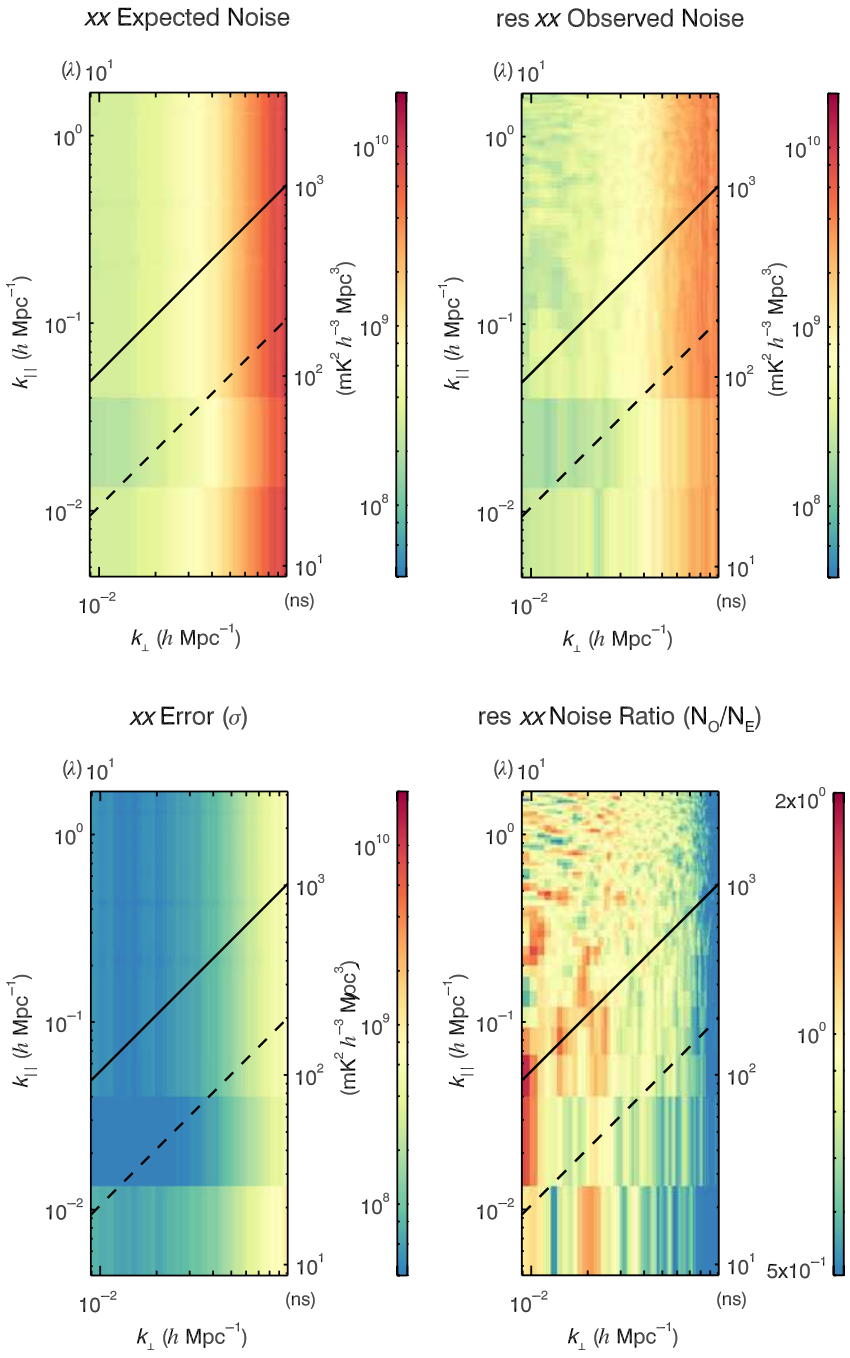

Figure 6. The 2D power spectra for expected noise, observed noise, error bars, and noise ratio of an integration of 64 observations from 2013 August 23 in instrumental *XX* for the MWA. The observed noise (N_O) and expected analytically propagated noise (N_E) have a ratio near 1, indicating our assumptions are satisfactory. The error bars are related to the observed noise via Equation 26.

useful to also use 1D diagnostic power spectra to more quantitatively measure contamination, which helps in discerning smaller contributions.

The cable reflections and flagged aliased channels lead to bright contamination due to their modulation as a function of frequency. This is an important revelation; we must minimise all forms of spectrally repetitive signals during power spectrum analysis. If a spectral mode is introduced in the instrument or in the processing, that mode cannot be used to detect the EoR.

6.3. 2D difference power spectrum

Often, we would like to compare a new data analysis technique to a standard to assess potential improvements. We can do this with a side-by-side comparison of 2D power spectra, but are limited by the large dynamic range of the colour bar. Instead, we create 2D difference power spectra.

We take a 3D bin-by-bin difference between two $\{k_x, k_y, k_z\}$-cubes and then average to generate a 2D difference power spectrum. The reference or standard is subtracted *from* the new run, and the 2D difference power spectrum varies positive and negative depending on power levels. We choose a red–blue log-symmetric colour bar to indicate sign, where red indicates an increase in power relative to the reference and blue indicates a decrease in power.

Figure 8 shows an example of a 2D difference power spectra. The new data analysis run is the left panel, the reference is the middle panel, and the 2D difference is the right panel. For this

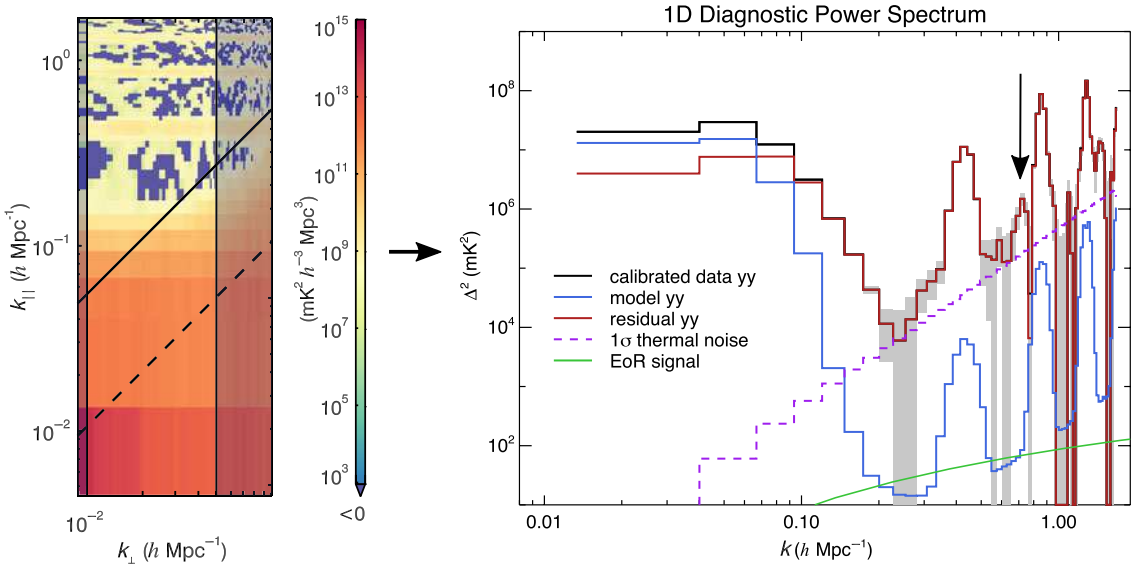

Figure 7. 1D power spectra as a function of k for calibrated data (black), model (blue), residual (red), 2σ uncertainties (shaded grey), theoretical EoR (for comparison, green), and the thermal noise (dashed purple) of an integration of 64 MWA observations from 2013 August 23, for instrumental *YY*. The 2D power spectrum highlights the bins that went into the 1D averaging, which we can modify to exclude the foreground wedge when making limits. A cable-reflection contamination feature at 0.7 h Mpc^{-1} is more obvious in this 1D power spectrum, which highlights the importance of using 1D space as a secondary diagnostic.

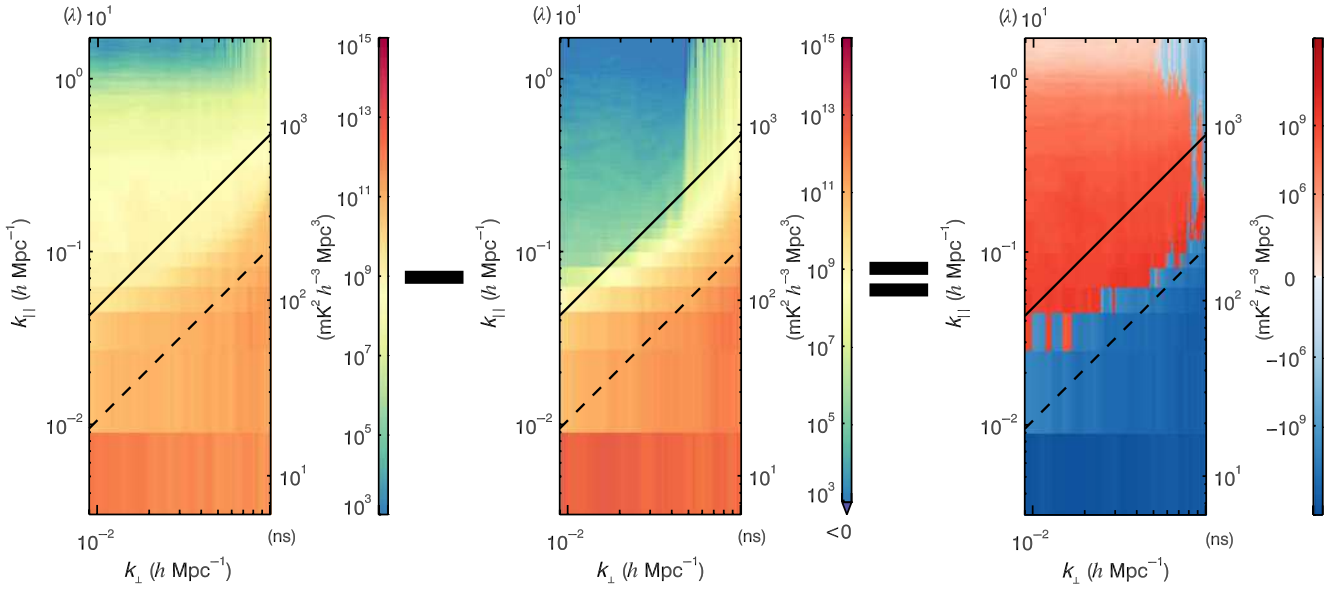

Figure 8. The subtraction of a residual 2D power spectrum (left) and a reference residual 2D power spectrum (middle) to create a difference 2D power spectrum (right). Red indicates a relative excess of power, and blue indicates a relative depression of power. This diagnostic is helpful in determining differences in plots that inherently cover twelve orders of magnitude.

example, we have chosen a new data analysis run that had excess power in the window and a decrease in power in the foreground wedge compared to the reference. In general, we make 2D difference power spectra for dirty, model, and residual, for both *XX* and *YY* polarisations to match the six-panel plot in Figure 5.

7. Signal loss simulation

The FHD/εppsilon pipeline can be used as an *in situ* simulation to test for signal loss. This validates our final power spectra results, allowing for confidence in our EoR upper limits.

FHD is an instrument simulator by design; we create model visibilities from all reliable point sources and realistic beam kernels. We can use these model visibilities as the base of an input simulation of the instrument and sky. A statistical Gaussian EoR is Fourier-transformed to *uv*-space, encoded with instrumental effects via the *uv*-beam, and added to point source visibilities to create data that we can test for signal loss. These *in situ* simulation visibilities are input into the FHD/εppsilon pipeline. They are treated like real data and are subject to all real data analyses.

We perform four *in situ* simulations to validate our pipeline:

- Compare input EoR signal to output EoR signal with no additional foregrounds or calibration. This simple test demonstrates consistency in power spectrum normalisation and signal preservation.

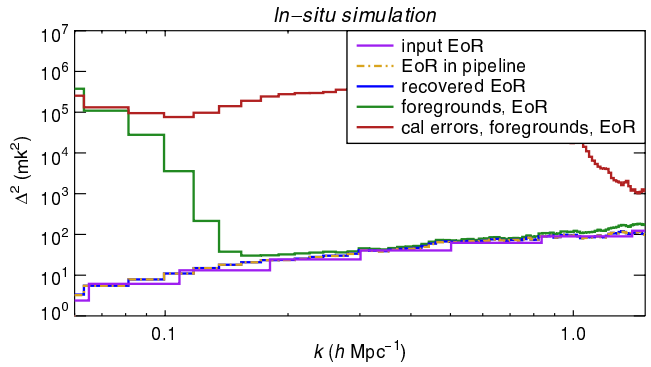

Figure 9. The signal loss simulations of the FHD/εppsilon pipeline. We recover the input EoR signal (purple) for most k modes if: (1) an EoR signal is propagated through the pipeline without foregrounds or calibration (orange, dashed), (2) all foregrounds are used for calibration and are perfectly subtracted (blue), and (3) all foregrounds are used for calibration but are not perfectly subtracted (green). We do not recover the EoR if there are calibration errors (red); however, these errors are not indicative of signal loss.

- Allow all sources to be used in the calibration and subtraction model and compare the output power to the input EoR signal. By recovering the input EoR signal, we demonstrate that the addition of foregrounds and calibration does not result in signal loss.
- Use an imperfect model for calibration and subtraction and compare the output power to the input EoR signal. Particularly, we calibrate with a global bandpass (Equation 9) in the comparison. The calibration errors caused by spectral structure from unmodelled sources cause contamination in the EoR window (see Barry et al. 2016), but there is no evidence of signal loss.
- Use an imperfect model for subtraction but a perfect model for calibration, and compare the output power in the EoR window to the input EoR signal. We can recover the EoR signal in some of the EoR window, demonstrating that we can theoretically detect the EoR even with unsubtracted foregrounds.

Figure 9 shows the output power from each test alongside the underlying, input EoR signal. For these simulations, we use the MWA Phase I instrument without channeliser effects. They perform as expected, recovering the EoR signal when possible. However, at high k-modes, the foreground simulation without calibration errors is contaminated due to a gridding resolution systematic (Beardsley et al. 2016).

If regular flagging of aliased channeliser effects is included, more modes are contaminated in the EoR window, as seen in the model data of Figure 7. This is a consequence of the lack of inverse-variance weighting during the spectral transform. Implementing a more complex weighting scheme and/or modifying the actual instrument will be necessary to recover those modes in the future.

It is important for an EoR power spectrum analysis to characterise potential signal loss in order for EoR limits to be valid (Cheng et al. 2018). These tests demonstrate that we are not subject to signal loss within our pipeline. None of our analysis methodologies, from FHD all the way through εppsilon, is artificially removing the EoR signal. In addition, these simulations can be expanded to include other effects in the future, such as channeliser structure, diffuse emission, and beam errors.

These signal loss simulations can only test *relative* forms of signal loss. For example, this cannot test the validity of the initial encoded instrument effects on the input (i.e. the *in situ* simulation or EoR visibilities), nor can it test the accuracy of the sky temperature used in normalisation.

8. Overview

We have built an analysis framework that takes large quantities of measured visibilities and generates images, power spectra, and other diagnostics. Four modularised components exist: pre-analysis flagging and averaging, calibration and imaging, integration, and error-propagated power spectrum calculations. The accuracy of each component is crucial due to the level of precision needed in an EoR experiment. In particular, the analysis handled by FHD and εppsilon is complicated and multifaceted, necessitating constant refinement and development to ensure accuracy, precision, and reproducibility.

Many data products are produced with our analysis pipeline. We make instrumental polarisation images for each observation for calibrated, model, and residual data. As for power spectra, we make 2D and 1D representations for the calibrated, model, and residual data from integrated cubes along with propagated noise and propagated uncertainty estimates. These form the basic outputs within the FHD/εppsilon pipeline.

All examples provided have been from MWA Phase I data; however, FHD/εppsilon is more generally applicable. It has been used to publish MWA Phase II data (Li et al. 2018) and PAPER data (Kerrigan et al. 2018), with plans to analyse HERA data. Flexibility within the pipeline allows for constant cross-validation and a wide variety of use cases.

It is crucial to have a fully verifiable pipeline, and this has been the major motivation behind FHD/εppsilon. All of our diagnostic outputs, error propagation products, and signal loss simulations are a part of a signal preservation narrative. By establishing confidence in our data analysis, we can pursue and publish credible EoR upper limits and measurements.

The FHD/εppsilon pipeline is readily available online and developed fully in public. Anyone can view our progress and critically analyse our methodologies. This is a resource for the community; FHD/εppsilon is an open-source example of the necessary precision techniques required for EoR power spectrum analysis.

Acknowledgements. 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 funded by National Science Foundation grants AST-1410484, AST-1506024, AST-1613040, AST-1613855, and AST-1643011. APB is supported by an NSF Astronomy and Astrophysics Postdoctoral Fellowship under #1701440.

References

Barry, N., Hazelton, B., Sullivan, I., Morales, M. F., & Pober, J. C. 2016, MNRAS, 461, 3135
Beardsley, A. P. 2015, PhD thesis, University of Washington
Beardsley, A. P., et al. 2013, MNRASL, 429, L5
Beardsley., A. P., et al. 2016, ApJ, 833, 102
Bhatnagar, S., Cornwell, T. J., Golap, K., & Uson, J. M. 2008, A&A, 487, 419
Bowman, J. D., et al. 2013, PASA - Publications of the Astronomical Society of Australia, 30, E031
Byrne, R., et al. 2019, ApJ, 875, 70
Carroll, P. A. 2016, PhD thesis, University of Washington
Cheng, C., 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, IEEE JSTSP, 2, 647
Datta, A., Bowman, J. D., & Carilli, C. L. 2010, ApJ, 724, 526
DeBoer, D. R., et al. 2017, PASP, 129, 045001
Dillon, J. S., et al. 2014, PRD, 89, 023002
Dillon, J. S., et al. 2015, PRD, 91, 123011
Ewall-Wice, A., et al. 2016, MNRAS, 460, 4320
Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., & Bartelmann, M. 2005, ApJ, 622, 759
Hamaker, J. P., Bregman, J. D., & Sault, R. J. 1996, A&ASS, 117, 137
Hazelton, B. J., Morales, M. F., & Sullivan, I. S. 2013, ApJ, 770, 156
Hogg, D. W. 1999, arXiv:astro-ph/9905116
Hurley-Walker, N., et al. 2017, MNRAS, 464, 1146
Jacobs, D. C., et al. 2016, ApJ, 825, 114
Jordan, C. H., et al. 2017, MNRAS, 471, 3974
Kerrigan, J. R., et al. 2018, ApJ, 864, 131
Koopmans, L., et al. 2014, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). SISSA, ed. T. L. Bourke, et al. (SISSA: Giardini Naxos, Italy)
Lenc, E., et al. 2016, ApJ, 830, 38
Li, W., et al. 2018, ApJ, 863, 170
Line, J. 2017, PhD thesis, University of Melbourne, http://minerva-access.unimelb.edu.au/handle/11343/191127
Liu, A., Parsons, A. R., & Trott, C. M. 2014, PRD, 90, 023018
Lomb, N. R. 1976, ASS, 39, 447
Matteo, T. D., Perna, R., Abel, T., & Rees, M. J. 2002, ApJ, 564, 576
Mellema, G., et al. 2013, Exp. Astron., 36, 235
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 JSTSP, 2, 707
Morales, M. F., & Hewitt J. 2004, ApJ, 615, 7
Morales, M. F., & Matejek, M. 2009, MNRAS, 400, 1814
Morales, M. F., Hazelton, B., Sullivan, I., & Beardsley, A. 2012, ApJ, 752, 137
Morales, M. F., Beardsley, A., Pober, J., Barry, N., Hazelton, B., Jacobs, D., & Sullivan, I. 2018, MNRAS, 483, 2207–2216
Myers, S. T., et al. 2003, ApJ, 591, 575
Offringa, A. R., et al. 2015, PASA, 32, e008
Offringa, A. R., Mertens, F., & Koopmans, L. V. E. 2019, MNRAS, 484, 2866
Oh, S. P., & Mack, K. J. 2003, MNRAS, 346, 871
Ord, S. M., et al. 2010, PASP, 122, 1353
Paciga, G., et al. 2011, MNRAS, 413, 1174
Parsons, A. R., et al. 2010, ApJ, 139, 1468
Parsons, A. R., Pober, J. C., Aguirre, J. E., Carilli, C. L., Jacobs, D. C., & Moore, D. F. 2012, ApJ, 756, 165
Patil, A. H., et al. 2014, MNRAS, 443, 1113
Patil, A. H., et al. 2016, MNRAS, 463, 4317
Patil, A. H., et al. 2017, ApJ, 838, 65
Pober, J. C., et al. 2013, ApJL, 768, L36
Pober, J. C., et al. 2014, ApJ, 782, 66
Pober, J. C., et al. 2016, ApJ, 819, 8
Riseley, C. J., et al. 2018, PASA, 35, E043
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, PRD, 91, 083514
Sullivan, I. S., et al. 2012, ApJ, 759, 17
Sutinjo, A., O'Sullivan, J., Lenc, E., Wayth, R. B., Padhi, S., Hall, P., & Tingay, S. J. 2015, Radio Sci., 50, 2014RS005517
Thyagarajan, N., et al. 2013, ApJ, 776, 6
Tingay, S. J., et al. 2013, PASA, 30, e007
Trott, C. M., Wayth, R. B., & Tingay, S. J. 2012, ApJ, 757, 101
Trott, C. M., et al. 2016, ApJ, 818, 139
Trott, C. M., et al. 2018, ApJ, 867, 15
van Haarlem, M. P., et al. 2013, A&A, 556, A2
Vedantham, H., Shankar, N. U. & Subrahmanyan, R. 2012, ApJ, 745, 176
Vleck, J. H. V., & Middleton, D. 1966, Proc. IEEE, 54, 2
Wayth, R. B., et al. 2018, PASA, 35, E033
Yatawatta, S., et al. 2013, A&A, 550, A136
Zheng, H., et al. 2017, MNRAS, 465, 2901

Appendix A. Power spectrum volume normalisation

The data products handed from FHD to εppsilon are reconstructions of the apparent sky—a 1 Jy source at the half-power point of the beam will appear as a $\frac{1}{2}$ Jy source. A volume factor is needed to convert the power spectrum of the apparent image into a properly normalised cosmological power spectrum. The normalisation factor needed for an area of the *uv*-plane estimated from an isolated baseline differs from the normalisation factor for an area estimated from many overlapping baselines. Formally, this is related to the covariance of the overlapping visibilities, as described in Liu et al. (2014). Carrying the full covariance through the pipeline is computationally intractable, so we separately normalise regions with dense and poor *uv*-coverage.

We demonstrate this effect with end-to-end simulations in FHD/εppsilon of a flat power spectrum signal (constant power as a function of k) and randomly located baselines with increasing density in the *uv*-plane. For each simulation, a baseline density is selected and baselines are randomly placed in a *uv*-plane. The corresponding visibilities are simulated for a stochastic signal with a flat power in k, and then gridded to calculate a 1D power spectrum. The reconstructed 1D power spectra are flat in k, but the normalisation relative to the calculated sparse normalisation depends on baseline density, as shown in Figure A.1. In the limit of very sparse baselines, there are almost no overlaps and the sparse normalisation is correct, but as the density increases the normalisation decreases quickly and asymptotes to half the sparse normalisation. This factor of 2 in the denominator can be understood as a doubling of the effective area of the *uv*-plane that contributes to the gridded value at any location in the gridded *uv*-plane.

Formally, there is a transition region where the normalisation is difficult to calculate without fully propagating the covariance. However, virtually all our sensitivity comes from regions of the *uv*-plane with very dense overlap. We do not discard the low-density regions in our 2D power spectra, because 2D spectra are primarily diagnostic in nature and are useful in helping us to identify systematics. However, for EoR upper limits and 1D power spectra we discard low density regions to avoid uncertainty in the normalisation. Discarding the sparse areas of the *uv*-plane does not significantly decrease our sensitivity and alleviates the need to fully propagate the covariance matrix.

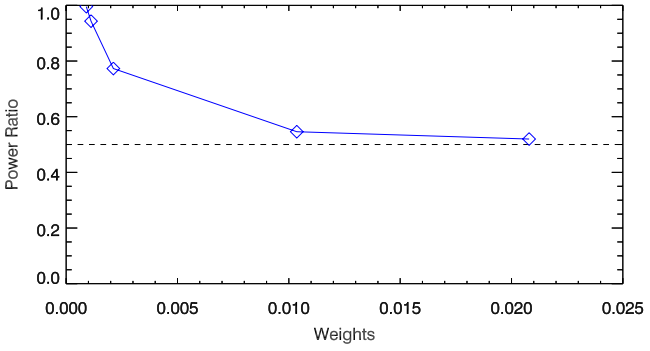

Figure A.1. Results of our end-to-end simulations through FHD/εppsilon of a flat power spectrum signal. The blue points give the ratio of the reconstructed 1D power spectra using the calculated sparse normalisation to the input power level (which is flat in k). The x-axis is a measure of baseline density—it gives the average baseline weight gridded to each *uv*-pixel in the simulation. In constructing limits and 1D power spectra, we only use regions of the *uv*-plane where the minimum weight (as a function of frequency) is greater than or equal to 1.