Calibration Requirements for Detecting the 21 cm Epoch of Reionization Power Spectrum and Implications for the SKA

N. Barry,1⋆ B. Hazelton,2,1 I. Sullivan,3 M. F. Morales,1 J. C. Pober 4
1University of Washington, Department of Physics  
2University of Washington, eScience Institute  
3University of Washington, Department of Astronomy  
4Brown University, Department of Physics

ABSTRACT

21 cm Epoch of Reionization observations promise to transform our understanding of galaxy formation, but these observations are impossible without unprecedented levels of instrument calibration. We present end-to-end simulations of a full EoR power spectrum analysis including all of the major components of a real data processing pipeline: models of astrophysical foregrounds and EoR signal, frequency-dependent instrument effects, sky-based antenna calibration, and the full PS analysis. This study reveals that traditional sky-based per-frequency antenna calibration can only be implemented in EoR measurement analyses if the calibration model is unrealistically accurate. For reasonable levels of catalogue completeness, the calibration introduces contamination in otherwise foreground-free power spectrum modes, precluding a PS measurement. We explore the origin of this contamination and potential mitigation techniques. We show that there is a strong joint constraint on the precision of the calibration catalogue and the inherent spectral smoothness of antennae, and that this has significant implications for the instrumental design of the SKA and other future EoR observatories.

Key words: dark ages, reionization, first stars; techniques: interferometric; methods: data analysis; instrumentation: interferometers

1 INTRODUCTION

Observations of the Epoch of Reionization (EoR) promise to reveal a wealth of information about the dynamics and evolution of the universe. The 21 cm hyperfine transition line of neutral hydrogen is one of the best probes of the EoR (see Furlanetto, Peng Oh & Briggs (2006); Morales & Wyithe (2010) for reviews) and several experiments are currently seeking or will seek a power spectrum (PS) measurement of this faint cosmological signal. The low-frequency interferometers attempting to make these measurements include the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER; Parsons et al. 2010)1, the LOw Frequency ARray (LOFAR; Yatawatta et al. 2013; van Haarlem et al. 2013)2, the Murchison Widefield Array (MWA; Lonsdale et al. 2009; Tingay et al. 2013; Bowman et al. 2013)3, the Hydrogen Epoch of Reionization Array (HERA; Pober et al. 2014)4, and the Square Kilometre Array (SKA; Mellema et al. 2013; Koopmans et al. 2015)5.

However, astrophysical foregrounds are 4–5 orders of magnitude brighter than the expected cosmological signal. Suppressing this foreground contamination in the PS requires unprecedented precision in instrumental calibration. We simulate sky-based calibration on the EoR PS signal to explore the techniques necessary to suppress the foreground contamination. Our end-to-end simulations include chromatic instrumental effects common to current EoR experiments and several experiments are currently seeking or will seek a power spectrum (PS) measurement of this faint cosmological signal. The low-frequency interferometers attempting to make these measurements include the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER; Parsons et al. 2010), the Low Frequency ARray (LOFAR; Yatawatta et al. 2013; van Haarlem et al. 2013), the Murchison Widefield Array (MWA; Lonsdale et al. 2009; Tingay et al. 2013; Bowman et al. 2013), the Hydrogen Epoch of Reionization Array (HERA; Pober et al. 2014), and

Variance and convergence statistics of calibration in image space have been studied in detail, including the application of ionospheric changes (van der Tol et al. 2007; Mitchell et al. 2008; Wijnholds & van der Veen 2009; Datta et al. 2009), variation of dif-

1 Contact e-mail: nbarry@uw.edu  
2 http://eor.berkeley.edu  
3 http://www.lofar.org  
4 http://www.mwatelescope.org  
5 http://reionization.org

© 2016 The Authors

arXiv:1603.00607v2 [astro-ph.IM] 7 Jun 2016
fuse structure scales (Mitchell et al. 2008), addition of polarized components (Mitchell et al. 2008), inclusion of small source position offsets (Ng & See 1996; Wijnholds & van der Veen 2009; Datta et al. 2009), and the imperfection of source models (Datta et al. 2009, 2010). This includes the approach that we label as traditional, where calibration solutions are solved for each antenna and each frequency from a least squares analysis of the visibilities (Salvini & Wijnholds 2014). Calibration techniques and effects have been rigorously tested via variance statistics. However, only a few studies have investigated calibration effects on the PS (e.g. Trott & Wayth 2016; Switzer et al. 2015). The EoR measurement is to occur in PS space, so the standard of the effectiveness for calibration should also be established in PS space.

In addition, it is imperative to use a realistically imperfect calibration catalogue for sky-based calibration when simulating resultant effects. Models of sources generated from knowledge of the sky are used to calibrate the instrument (Mitchell et al. 2008), but no catalogue is perfect. There will always be unmodelled faint sources, either due to confusion limits or the inability to resolve morphology, and catalogues will always include small errors in flux, position, or compactness of sources. These imperfections in the sky model will affect calibration, and subsequently the EoR PS.

This study explores the impact of calibrating against an incomplete source model on the EoR PS measurement. We show that unmodelled, faint sources can interact with calibration to mix foregrounds into unrelated modes. Spectral structure due to the incomplete source model propagated via traditional per-frequency antenna calibration couples bright foreground power from unimportant Fourier modes into the most sensitive EoR modes. Consequently, the EoR measurement is impossible without the development of new calibration techniques beyond traditional methods.

Descriptions of the simulations, software packages, catalogue data, and PS space are given in §2. The effects on the PS of using traditional per-frequency calibration techniques are shown in §3. Mitigation techniques to avoid fitting spectral structure from faint sources are demonstrated in §4, which also highlights the importance of a spectrally smooth instrumental response. Approaches to faithfully reconstruct true instrumental spectral structure, while minimizing the effect of unmodelled faint sources, are described in §5 and the implications for the SKA and other future EoR machines are discussed in §6.

We note that this work concentrates on the “imaging” EoR PS analysis approach, and the calibration requirements for “delay” PS analyses with redundant arrays may be different (Parsons et al. 2012a). A full study of the calibration requirements of delay spectra is left for future work.

2 METHODS AND MEASUREMENT SPACE

Our calibration simulations employ a suite of packages designed for MWA EoR analysis. These full end-to-end simulations demonstrate the effect of calibration errors in the two-dimensional power spectrum (2D PS) — a primary figure of merit for the EoR measurement. In this section, we describe the 2D EoR PS figure of merit and our simulation methods.

2.1 The 2D power spectrum

The 21 cm hyperfine transition of neutral hydrogen is a narrow emission line. This allows the measured frequency of the emission to map closely to its line-of-sight distance. Measurements of the EoR are inherently three dimensional, with two angular dimensions and one frequency dimension represented in the volume \( \{\theta_\parallel, \theta_\perp, \nu\} \). Using the angular diameter and line-of-sight distances, the observations can be mapped to cosmological coordinates \( \{x, y, z\} \) in comoving Mpc (Hogg 1999). Statistical measurements of the EoR show the most promise for a robust detection in wavenumber space (Morales & Hewitt 2004), represented by \( \{k_\parallel, k_\perp, k_z\} \) and accessible through Fourier transforms.

The distribution of hydrogen in the universe is isotropic and homogeneous to first order. This spherical symmetry can be harnessed to achieve greater sensitivity. Averaging the squared measurements in spherical shells within the volume \( \{k_\parallel, k_\perp, k_z\} \) allows a transformation into a one-dimensional power spectrum (1D PS). While this aids in the measurement of the EoR, it removes the ability to view the k-space distributions of foreground and calibration effects. We will therefore present results in the 2D PS, achieved through averaging squared measurements along only the angular wavenumbers \( \{k_\parallel, k_\perp\} \). This creates the PS as a function of modes perpendicular to the line-of-sight (\( k_z\)) and modes parallel to the line-of-sight (\( k_z\)) as shown in Figure 1. Axes are displayed in units of Hubble constant (\( b\)) times inverse megaparsec (Mpc\(^{-1}\)).

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; Peng Oh & Mack 2003). Only small \( k_z\) values are theoretically contaminated by bright, spectrally smooth astrophysical foregrounds. Since the foreground power is restricted to only a few low \( k_z\) modes, larger \( k_z\) values tend to be free of “intrinsically 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 \( k_z\) values into larger \( k_z\) values as demonstrated in Figure 1 (Datta et al. 2010; Morales et al. 2012; Vedantham et al. 2012; Parsons et al. 2012b; Trott et al. 2012; Hazeldon et al. 2013; Thyagarajan et al. 2013; Pober et al. 2013; Liu et al. 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 \( k_z\), the most sensitive measurements are expected to be in the lower, left-hand corner of the EoR window.

PS from our end-to-end simulation in Figure 2 show the standard features described in the Figure 1 schematic. The left plot is the 2D PS of the calibrated simulation containing foregrounds and EoR signal. The middle plot, which looks nearly identical, shows the instrument calibration model. This contains a subset of the foregrounds to simulate an incomplete knowledge of the sky. We can decrease the contamination in the PS by subtracting this model from calibrated data to possibly reveal the EoR signal in a wider range of modes, depending on completeness of the model. Taking the difference yields the “residual” 2D PS in the right plot of Figure 2, which reveals unmodelled foregrounds with their instrumental effects and the EoR signal. This subtraction is implemented in the three dimensional measurement cube before constructing the PS. In addition to the typical 2D PS effects, foregrounds also contaminate higher \( k_z\) as a consequence of decreased baseline coverage of the instrument at those scales, which is specific and intrinsic to each array. For the remainder of this paper, we will use the residual 2D PS space to explore the effects of calibration.
2.2 Simulation methods

All of the calibration simulations in this paper utilize the MWA antennae and positions in the frequency band 167–198 MHz. We use a precursor to the KGS catalogue (Carroll et al. in review) as our foreground model and the Fast Holographic Deconvolution\(^6\) (FHD) software package to implement our simulations, calibration, and imaging (Sullivan et al. 2012). To create PS, we use the Error Propagated Power Spectrum with InterLeaved Observed Noise\(^7\) (epsilon) software package. This simulates the full end-to-end MWA PS analysis detailed in Jacobs et al. (2016).

A model of 6950 compact sources seen by the MWA and compiled in the KGS catalogue were used as simulated input data, along with the addition of a simulated Gaussian EoR signal in the visible.

\(^6\) FHD software package is available at https://github.com/ EoRImaging/FHD

\(^7\) epsilon software package is available at https://github.com/ EoRImaging/epsilon
Figure 2. From left to right: the calibrated data from a calibration simulation, the model given known foregrounds and instrumental effects, and the residual after the model is subtracted from the calibrated data. The residual 2D PS has the potential to reveal more modes to the EoR PS measurement given the accuracy and completeness of the model.

Figure 3. The result of the calibration simulation pipeline with a perfect sky model. Modelling, calibrating, and subtracting all of the 6950 KGS sources using FHD and epsilon recovers the added simulated EoR signal. The EoR signal that we recover does not experience signal loss regardless of calibration technique used. The color scale has been fixed throughout this paper to provide order of magnitude reference.

approach was an extension from single-frequency radio astronomy to small-bandwidth multi-frequency instrumentation (Fomalont & Perley 1999; Sullivan 2009). While this method involves solving for many variables, the number of degrees of freedom in the data from the MWA is orders of magnitude larger than the number of parameters used and thus theoretically constrained. This calibration method has remained a stalwart in the community as the field has advanced. Our calibration simulations first examine the traditional radio astronomy calibration technique and how it effects the EoR PS measurement.

To capture realistic differences between the true sky and the calibration catalogue, we simulate the sky as 6950 sources but only use the brightest 4000 to predict the visibilities for use in calibration. This introduces small differences between the sky and calibration visibilities that can affect the per-frequency antenna calibration solutions. We apply the antenna calibration solutions to the input sky visibilities, and then subtract the brightest 4000 sources used in the calibration model for one observation. This residual 2D PS is shown as the left plot of Figure 4. The 2950 unmodelled faint sources populate the foreground wedge as expected.

We can also calculate a residual 2D PS with a perfect calibration and with the same 4000 source foreground subtraction to provide a reference for the observation. This is shown in the middle panel of Figure 4 (which is the same residual 2D PS shown in the righthand panel of Figure 2). Unmodelled faint sources also populate the foreground wedge; however, it is apparent that the 2D PS using traditional per-frequency antenna calibration has relatively high amounts of power in the EoR window.

The calibration simulation can be used to quantify the shape and amount of excess power in PS space. Direct subtraction between the traditional calibration simulation 2D PS (the left plot of Figure 4) and the reference 2D PS (the middle plot of Figure 4) provides this information. The result is a difference 2D PS, where red indicates relative excess power and blue indicates relative depressed power. Evident in the difference 2D PS is the excess power contamination of the entire EoR window by as much as $10^7 \text{mK}^2 \ h^{-3} \text{Mpc}^3$. It is important to note that this level of calibration error would make the EoR measurement impossible. Using traditional per-frequency antenna calibration in a PS measurement would require a highly accurate calibration catalogue.

Qualitatively, allowing independent calibration parameters for each frequency channel and antenna allows small deviations from the true solutions on small spectral scales. These amplitude and phase deviations are caused by the point spread functions (PSF) of unmodelled sources which modify the observed fluxes of true sources, as seen in real data by Ofrińga et al. (2016). This effect is frequency dependent and its magnitude depends on the completeness of the sky model and the natural PSF of the array. The resulting
calibration errors are only on the order of 1 part in $10^7$ in this simulation. However, this varied spectral structure in the calibration solutions is enough to couple power from the bright, intrinsic foregrounds to the Fourier modes in the EoR window. This fills every possible EoR measurement mode with foreground power.

Not only are sensitive regions of the EoR window dominated by coupled power from intrinsic foregrounds, but there is a corresponding depression of power in the foreground wedge as well. This is also the result of small spectral deviations captured in the calibration solutions. The measured fluxes of modelled sources do not accurately reflect the true fluxes due to the residual PSF of unmodelled sources. Allowing calibration solutions to be modified by this residual structure results in overfitting and over-subtraction.

Using the modulation theorem, we can quantitatively associate the level of contamination seen in the PS with the observed calibration errors. Data that is modified by spectrally variant calibration solutions is Fourier transformed into PS space, and the modulated fluxes of modelled sources do not accurately reflect the true fluxes due to the residual PSF of unmodelled sources. Allowing calibration solutions to be modified by this residual structure results in overfitting and over-subtraction.

Excess power can be estimated given a modulated signal

$$h(\nu) = f(\nu) (1 + \Delta g \cos \eta_0 \nu),$$

where $h(\nu)$ is the modulated instrumental response as a function of frequency, $f(\nu)$ is the original instrumental response as a function of frequency, $\eta_0$ is the Fourier dual of a mode in the amplitude deviations of the calibration gain, and $\Delta g$ is the amplitude deviation associated with the frequency mode $\eta_0$. The modulation theorem results in the Fourier transform

$$H(\eta) = \frac{\Delta g}{2} F(\eta - \eta_0) + \frac{\Delta g}{2} F(\eta + \eta_0) + F(\eta).$$

Fourier transforms of the original signal $f$ constructs signal at $\eta - \eta_0$, and $\eta + \eta_0$. Equation 2 is squared to obtain the PS, and cross-terms between $F(\eta)$ and $F(\eta \pm \eta_0)$ can be neglected since overlap is small for an $\eta_0$ in the EoR window. An order of magnitude estimate of the positive power spectrum of this modified signal is

$$O[(H(\eta))^2] \approx O[(F(\eta))^2] + O\left(\frac{\Delta g}{2} F(\eta \pm \eta_0)\right)^2.$$  

As a result, the modulated power response $O[(H(\eta))^2]$ has power contributions as a function of $\eta$ and, to a lesser extent, $\eta \pm \eta_0$. When all $\eta$ and $\eta_0$ values are considered, the result is equivalent to the convolution of the foregrounds with the Fourier transform of the calibration deviations.

For small $\eta$ values, intrinsic foregrounds dominate. Power will be modulated from these intrinsic foregrounds into any frequency mode $\eta_0$ captured in the amplitude deviations in calibration. Given simulation values of the intrinsic foregrounds ($O(P_{\text{ff}}) \approx 10^{14} \text{ mK}^2 h^{-3} \text{ Mpc}^3$) and the amplitude deviations ($O(P_{\text{ff}}) \approx 10^{-7}$, or a $\Delta g$ of order 1 part in $10^7$), the excess contamination in frequency mode $\eta_0$ of the PS is estimated to be $10^7 \text{ mK}^2 h^{-3} \text{ Mpc}^3$. This agrees with the level of contaminated power in Figure 4 generated by calibration simulations.

The satisfactory performance of traditional per-frequency antenna calibration depends on a highly accurate calibration catalogue. When we use the same sources to generate the sky and calibration models — even with an added EoR signal — the resulting calibration and foreground suppression in the PS is excellent, as seen in Figure 3. However, this is not a realistic situation for current and planned EoR observatories. When the calibration catalogue is not perfect, traditional per-frequency antenna calibration distributes spectral power and overwhelms the faint cosmological signal as seen in Figure 4. This sets very strong constraints on the accuracy of the calibration catalogue if the traditional calibration approach is to be used for EoR measurements.

4 MITIGATION BY SMOOTH CALIBRATION SOLUTIONS

Spectral contamination in the EoR window from traditional calibration techniques necessitates mitigation. If the instrument is spectrally-smooth across the frequency band, we can use this as a prior that must be met in our calibration solutions. We explore constraining the spectral variation of the calibration to be smooth relative to the band size to avoid contamination of the EoR window. However, non-smooth spectral features of the instrument must be
incorporated into the calibration, and therefore we also investigate the consequences of fitting specific instrumental features.

### 4.1 Constraining smooth instrumental response

If an antenna has a naturally smooth bandpass, its response can be modeled with low-order polynomials or other slowly varying functions. With this restriction, we avoid the fine-scale spectral structure in the calibration solutions that causes the contamination of the EoR window seen in Section 3. Perfect calibration solutions are flat in our simulation, so polynomials applied to the frequency band would only model the level of error expected with polynomial fitting.

We calculate best-fit polynomials over the whole frequency band from the traditional calibration solutions generated in Section 3 with a calibration catalogue of the brightest 4000 of the 6950 input sources. Five calibration parameters for the frequency band are allowed and are chosen to represent typical instrumental variation. Three amplitude parameters create a second-order bandpass-like polynomial, and two phase parameters create a smooth ramp in frequency.

Figure 5 shows the difference 2D PS between the smooth calibration solution PS and the reference PS for one observation. The EoR window from the smooth mitigation technique and the reference are strikingly similar, leading to very little difference. The level of the difference is also noise-like and far below the EoR signal. This will neither affect an EoR measurement to a significant degree nor bias the result.

Power from intrinsic foregrounds is not coupled to the EoR window due to the restriction of smoothness relative to EoR spectral modes. Spectral contamination on the scale of the polynomials still occurs; however, this contamination only occurs within the foreground wedge and will not hinder the EoR measurement. The bold line in Figure 5 highlights the highest $k_L$ with significant contamination caused by the low-order polynomial fitting, which is below the EoR window.

Differences in power in the region of high $k_L$ and high $k_⊥$ in Figure 5 are also apparent. Poor baseline coverage couples the foreground wedge to this region as described in Section 2.1. Since spectral contamination did occur in the foreground wedge, power changes occur in the region affected by poor baseline coverage. EoR measurements will not be made in $k$-space areas with poor baseline coverage, so power changes in this area are not a large concern. The dotted line in Figure 5 indicates the largest $k_⊥$ with high baseline coverage. A significant foreground-free EoR window to the upper left remains.

By restricting the instrumental response to be smooth with respect to EoR spectral modes, we significantly reduced the excess power in the EoR window caused by spectral contamination. Now, the EoR window has no power bias and what little contamination there is appears noise-like. Measuring the EoR signal in a 2D PS that utilizes the smooth mitigation technique in the calibration solutions will be essentially unaffected by an imperfect catalogue. Instrumentation that is spectrally smooth can avoid contaminating the EoR window using this technique.

### 4.2 Calibration parameters in spectral modes

Instrumental responses are not always smooth across the frequency band. Any spectral features in the instrument need to be fit so that the calibration is physically true. We simulate the effect of calibrating cable reflections as an example of the consequences of fitting for instrumental structure on a per-antenna basis.

Receiver-to-beamformer cable reflections with amplitudes of $\sim 1\%$ of the signal are apparent in MWA data, creating a characteristic frequency ripple in the antenna gain at the corresponding light travel time (Dillon et al. 2015; Beardsley 2015; Ewall-Wice et al. 2016a; Jacobs et al. 2016). Cable reflections can vary dramatically from antenna to antenna, and must be fit individually. Our calibration simulation uses three parameters (mode, amplitude, and phase) to describe the spectral ripple from a hypothetical 150 m cable reflection in a subset of the antennas. The 150 m cable reflection does not actually exist in the simulated data, thus the fit responds to only unmodelled spectral structure from faint sources. For clarity, no other calibration parameters are included. The observed error in the reflection calibration parameters is less than 1 part in $10^3$.

The effect of coupling the intrinsic foregrounds and the imperfect fitted mode is shown in Figure 6. The difference 2D PS between the reference PS and the PS with a cable reflection calibration shows a clear excess of power at $k_L \sim 0.7 \, h \, $Mpc$^{-1}$, or the $k_L$ associated with a 150 m spectral ripple.

The amount of excess power, about $10^3 \, mK^2/h^{1.5}$Mpc$^3$, is not a coincidence. The accuracy levels of the fit and the amplitude deviations in traditional per-frequency calibration in §3 were similar. Whether the calibration is described per-frequency (§3) or per-spectral-mode ($\times k_L$), the same number of calibration terms are required to cover the bandwidth and the same level of contamination results.
5 MITIGATION BY AVERAGING CALIBRATION SOLUTIONS

Calibrating fine-scale instrumental frequency structure requires accurately modelling the antenna response while avoiding spurious spectral structure from unmodelled sources. This faint, unmodelled structure can be largely independent of non-redundant arrays, the spurious spectral structure from unmodelled sources varies from antenna to antenna. If the antennae are identical in manufacture, then averaging their calibration solutions to form a common bandpass can reduce the calibration amplitude and phase deviations that cause spectral contamination. This may not be true for arrays with redundant layouts or layouts with strong symmetries, where the antenna PSFs can be very similar.

Additionally, if the antenna calibrations are very stable in time, subsequent calibration solutions can be averaged effectively. This relies on rotation of the antenna PSFs with LST to provide semi-independent contamination from unmodelled sources.

We simulate a hypothetical array with the MWA layout (very random distribution, Beardsley et al. 2012, 2013; Tingay et al. 2013) with mechanically identical and stable antennae. Using traditional calibration techniques, we calculate solutions per frequency channel for each of the 128 antennae in the MWA every 2 minutes for a 30 minute observation traversing zenith. The resulting 1920 solutions per frequency (128 x 15) are then averaged, excluding outliers beyond a 2σ cut. The final averaged per-frequency calibration solution is then applied to all antennae for only one observation, allowing direct comparison to the other simulations.

Figure 7 presents the difference 2D PS between a reference PS and the averaged calibration solution PS. The amount of excess power in the EoR window has decreased by over three orders of magnitude compared to Figure 4 with traditional per-frequency and per-antenna calibration. However, relative excess power still remains and lines of constant $k_\perp$ contaminate the EoR window, indicating that spectral structure from unmodelled faint sources is still present in the average bandpass solution. The excess power level is similar to the expected power of the EoR. Whether or not an EoR measurement is feasible will depend on the completeness of the sky model, similarity of the antennae across elements and time, and the instrument’s design.

Figure 8 compares all difference PS in this work to the expected level of the EoR. We average $k_\parallel$ from 10 to 20.0 to generate a 1D PS as a function of $k_\perp$ in the EoR window. The level of contamination should be significantly below the EoR in order to realistically detect it with all other possible sources of error not explored in this work. We find that the contamination from the maximally-averaged calibration solution over all possible LSTs and antennae using a 4000 source sky model is at the level of the EoR, and therefore not a practical solution for the MWA. In contrast, we find that using low-order polynomials described in §4 is the best calibration method; it is lower than the expected EoR by one to two orders of magnitude. Current efforts to calibrate MWA EoR data use antenna and time averaging in conjunction with the smooth characteristics of the antenna to reduce the level of calibration contamination (Beardsley 2015). Other instruments may be able to achieve practical levels of spectral contamination with only averaged calibration solutions if more LST or antenna samples can be used.

In practice, thermal noise will also affect the PS and the calibration solutions. An additional set of simulations explored the effect of thermal noise on the calibration solutions, and showed the expected additional spectral contamination in the EoR window. The thermal noise contribution is uncorrelated in time, and averaging calibration solutions night to night proves effective in removing this contribution (see Trott & Wayth 2016 for detailed analysis). However, spectral contamination from faint, unmodelled sources still remains as a systematic error due to the limited number of antennae and observing LSTs. The systematic contribution of the calibration solution is largely independent for non-redundant arrays, the spurious spectral structure from unmodelled sources varies from antenna to antenna. The associated “antenna PSF” captures the effect of unmodelled sources on that specific calibration solution. Since each antenna’s baseline coverage is
§3 explored the source of this contamination, identifying the PSFs of the faint, unmodelled sources as the root cause. The chromatic PSFs of the unmodelled sources lead to small calibration errors on the order of $10^{-4}$ that couple to the bright foregrounds and distribute spectral contamination throughout the EoR window (Figure 4). §4 & §5 then explored more advanced calibration techniques that could mitigate this contamination.

The lesson for SKA, HERA, the MWA upgrade, the LOFAR upgrade, and other future EoR instruments is that any spectral features of the antennae that are calibrated will lead to contamination at the corresponding location in the EoR window. This is most clearly seen in Figure 6. This source of contamination via calibration errors places strong constraints on the instruments and planned observational programmes. To avoid the contamination identified in this paper, we identify four potential solutions:

1. Create a nearly perfect calibration catalogue. As the quality of the calibration catalogue improves, the amplitude of the associated calibration errors and modulated PS contamination both decrease. The necessary precision of the catalogue depends on the details of the array (in particular the PSF over the calibration period), and can be simulated using the techniques developed in §2. This solution requires that a high fidelity foreground catalogue be created before EoR analysis can start, and for some arrays creating a catalogue to the necessary precision may not be possible.

2. Use antennae with very smooth spectral responses. It is possible to design an analog and digital system that is naturally very spectrally smooth, with no calibration of features within the EoR window needed. The PS of such a hypothetical antenna is shown in Figure 4. In practice, this means there can be no spectral features larger than $\sim 10^{-3}$ in the antenna or receiver system with spectral scales faster than $\sim 8$ MHz (125 ns). This thinking is driving the spectrally smooth antenna and receiver designs of HERA and the MWA upgrade (Neben et al. 2016; Ewall-Wice et al. 2016b; Thyagarajan et al. 2016).

3. Manufacture physically identical and stable antennae. If all the antennae have the same spectral features and their response is very stable in time, then the antenna-time averaging explored in §5 can be used to reduce the spectral smoothness and catalogue precision requirements. While this work focuses on sky-based calibration, redundant calibration techniques explicitly depend on the antenna responses being identical (Wieringa 1992; Liu et al. 2010; Parsons et al. 2012a; Noorishad et al. 2012; Zheng et al. 2014).

4. Develop an external calibration system. The coupling to an incomplete sky catalogue can be entirely avoided by using an external calibrator such as a drone, satellite, pulsar, or pulse injection system. It is still a challenge to reach the $\sim 10^{-3}$ calibration precision needed, but several groups have been pursuing this path (Newburgh et al. 2014; Patra et al. 2015; Neben et al. 2015).

Which combinations of these four approaches will work the best is jointly dependent on the instrument-specific antenna PSFs and the precision and depth of the calibration catalogue. Calibration simulations following the techniques developed in §2 must be explored for each instrument to calculate the necessary instrument specifications.

In our opinion, building antennae with a naturally smooth spectral response (2) is the lowest risk and most cost-effective approach. Basing analysis plans on the development of a nearly perfect calibration catalog (1) is risky because it is hard to predict the achievable precision and depth of a catalogue made with a new instrument. Manufacturing physically identical antennae (3) is expensive, particularly factoring in the logistics of maintaining identical antennae with the same physical and spectral features. Developing an external calibration system (4) is challenging but potentially the most straightforward approach to achieving a high fidelity calibration catalogue.
performance in the field. Similarly, developing external calibration systems of the requisite precision (4) is expected to be expensive.

Spectrally smooth antennae described in approach (2) are the practical solution to avoiding spectral contamination of the EoR PS window. Figure 8 shows that it is the least contaminated calibration method in sensitive EoR modes explored in this work. We recommend that EoR instruments aim to have no spectral features larger than $\sim 10^{-5}$ on scales faster than $\sim 8$ MHz (125 ns).

It has long been recognized that precision calibration is necessary to perform EoR PS observations. In this work, we have identified that traditional per-frequency calibration techniques with an imperfect calibration catalogue can lead to significant contamination of the EoR PS window. We feel this insight and the associated simulation techniques can help guide the design of the SKA and other future EoR machines.

ACKNOWLEDGEMENTS

We would like to thank Cathryn Trott and Aaron Ewall-Wice for their insightful remarks and guidance. This work was supported by National Science Foundation grants AST-1410484 and AST-1506024.

REFERENCES

Beardsley A., 2015, PhD thesis, University of Washington
Beardsley A. P., et al., 2012, Monthly Notices of the Royal Astronomical Society, 425, 1781
Beardsley A. P., et al., 2013, Monthly Notices of the Royal Astronomical Society: Letters, 429, L5
Bowman J. D., et al., 2013, PASA - Publications of the Astronomical Society of Australia, 30
Datta A., Bhatnagar S., Carilli C. L., 2009, The Astrophysical Journal, 703, 1851
Datta A., Bowman J. D., Carilli C. L., 2010, The Astrophysical Journal, 724, 526
Dillon J. S., et al., 2015, Physical Review D, 91, 123011
Ewall-Wice A., et al., 2016a, Monthly Notices of the Royal Astronomical Society
Ewall-Wice A., et al., 2016b, The HERA Dish II: Electromagnetic Simulations and Science Implications (arXiv:1602.06277)
Fomalont E. B., Perley R. A., 1999, in Taylor G. B., Carilli C. L., Perley R. A., eds, Synthesis Imaging in Radio Astronomy II. ASP Conference series, pp 79–110
Furlanetto S. R., Peng Oh S., Briggs F. H., 2006, Physics Reports, 433, 181
van Haarlem M. P., et al., 2013, Astronomy & Astrophysics, 556, A2
Hazelton B. J., Morales M. F., Sullivan I. S., 2013, Astrophysical Journal, 770, 156
Hogg D. W., 1999, Distance measures in cosmology, arXiv: astro-ph/9905116, http://arxiv.org/abs/astro-ph/9905116
Jacobs D. C., et al., 2016, The Astrophysical Journal
Koopmans L., et al., 2015, in Bourke T. L., et al., eds, Advancing Astrot-
