https://doi.org/10.3847/1538-4357/aad3c3

Comparing Redundant and Sky-model-based Interferometric Calibration: A First Look with Phase II of the MWA

W. Li[1], J. C. Pober[1], B. J. Hazelton[2], N. Barry[2], M. F. Morales[2], I. Sullivan[2], A. R. Parsons[3], Z. S. Ali[3], J. S. Dillon[3],
A. P. Beardsley[4], J. D. Bowman[4], F. Briggs[5], R. Byrne[2], P. Carroll[2], B. Crosse[6], D. Emrich[6], A. Ewall-Wice[7], L. Feng[7],
T. M. O. Franzen[6], J. N. Hewitt[7], L. Horsley[6], D. C. Jacobs[4], M. Johnston-Hollitt[6,8], C. Jordan[6], R. C. Joseph[6,9,10],
D. L. Kaplan[11], D. Kenney[6], H. Kim[12], P. Kittiwisit[4], A. Lanman[1], J. Line[12], B. McKinley[12], D. A. Mitchell[10,13],
S. Murray[6], A. Neben[7], A. R. Offringa[14], D. Pallot[9], S. Paul[15], B. Pindor[12], P. Procopio[12], M. Rahimi[12,16], J. Riding[12],
S. K. Sethi[15], N. Udaya Shankar[15], K. Steele[6], R. Subrahmanian[15], M. Tegmark[7], N. Thyagarajan[4,17,19],
S. J. Tingay[6,10,18], C. Trott[6,10], M. Walker[6], R. B. Wayth[6,10], R. L. Webster[10,12,16], A. Williams[6], C. Wu[9], and
S. Wyithe[12]

[1] Department of Physics, Brown University, Providence, RI 02912, USA; wenyang_li@brown.edu
[2] Department of Physics, University of Washington, Seattle, WA 98195, USA
[3] University of California, Berkeley, Astronomy Dept., 501 Campbell Hall #3411, Berkeley, CA 94720, USA
[4] School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287, USA
[5] Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia
[6] International Centre for Radio Astronomy Research, Curtin University, Bentley, WA 6102, Australia
[7] MIT Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[8] Peripety Scientific Ltd., P.O. Box 11355 Manners Street, 6142 Wellington, New Zealand
[9] ICRAR University of Western Australia, Crawley, WA 6009, Australia
[10] ARC Centre of Excellence for All-sky Astrophysics (CAASTRO), Australia
[11] Department of Physics, University of Wisconsin–Milwaukee, Milwaukee, WI 53201, USA
[12] School of Physics, The University of Melbourne, Parkville, VIC 3010, Australia
[13] CSIRO Astronomy and Space Science (CASS), P.O. Box 76, Epping, NSW 1710, Australia
[14] ASTRON, The Netherlands Institute for Radio Astronomy, Postbus 2, 7990 AA, Dwingeloo, The Netherlands
[15] Raman Research Institute, Bangalore 560080, India
[16] ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia
[17] National Radio Astronomy Observatory, 1003 Lopezville Rd., Socorro, NM 87801, USA
[18] Osservatorio di Radio Astronomia, Istituto Nazionale di Astrofisica, 40123 Bologna, Italy

Abstract

Interferometric arrays seeking to measure the 21 cm signal from the epoch of reionization (EOR) must contend with overwhelmingly bright emission from foreground sources. Accurate recovery of the 21 cm signal will require precise calibration of the array, and several new avenues for calibration have been pursued in recent years, including methods using redundancy in the antenna configuration. The newly upgraded Phase II of Murchison Widefield Array (MWA) is the first interferometer that has large numbers of redundant baselines while retaining good instantaneous UV coverage. This array therefore provides a unique opportunity to compare redundant calibration with sky-model-based algorithms. In this paper, we present the first results from comparing both calibration approaches with MWA Phase II observations. For redundant calibration, we use the package OMNICAL and produce sky-based calibration solutions with the analysis package Fast Holographic Deconvolution (FHD). There are three principal results: (1) We report the success of OMNICAL on observations of ORBComm satellites, showing substantial agreement between redundant visibility measurements after calibration. (2) We directly compare OMNICAL calibration solutions with those from FHD and demonstrate that these two different calibration schemes give extremely similar results. (3) We explore improved calibration by combining OMNICAL and FHD. We evaluate these combined methods using power spectrum techniques developed for EOR analysis and find evidence for marginal improvements mitigating artifacts in the power spectrum. These results are likely limited by the signal-to-noise ratio in the 6 hr of data used, but they suggest future directions for combining these two calibration schemes.

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

1. Introduction

21 cm observations of the Epoch of Reionization (EOR) have the potential to reveal a wealth of information about the formation of the first stars and galaxies by measuring the 3D power spectrum (PS) and full tomographic maps of the neutral intergalactic medium (Morales & Wyithe 2010; Furlanetto 2016). However, these observations are technically very challenging owing to bright astrophysical foregrounds, the complex frequency dependence of the instrumental response of radio interferometers, radio frequency interference (RFI), and the effects of the ionosphere.

Recent work has highlighted the critical role precision instrument calibration will play in disentangling the faint cosmological signal from the bright foregrounds (Barry et al. 2016; Patil et al. 2016; Trott & Wayth 2016; Ewall-Wice et al. 2017). Current calibration efforts for EOR observations largely fall into two camps: sky-based calibration using deep foreground catalogs and forward modeling of the instrument

---

visibilities (Dillon et al. 2015; Beardsley et al. 2016; Carroll et al. 2016; Hurley-Walker et al. 2016; Patil et al. 2016; Trott et al. 2016; Intema et al. 2017; Procopio et al. 2017), and redundant calibration that forgoes a sky model but requires that the antennas be placed on a regular grid (Wieringa 1992; Liu et al. 2010; Zheng et al. 2014).

To date it has been impossible to directly compare the efficacy of the two calibration approaches on real data. Redundant arrays tend to have very poor UV coverage and are thus hard to calibrate with sky-based approaches (Parsons et al. 2012a; Zheng et al. 2016), and arrays with good imaging performance do not have the regular antenna layout necessary for redundant calibration.

Using new observations with Phase II (R. B. Wayth et al. 2018, in preparation) of the Murchison Widefield Array (MWA; Bowman et al. 2013; Tingay et al. 2013), we report on the first direct comparison of sky and redundant calibration with an EOR instrument. During Phase I, the MWA consisted of 128 antenna tiles in a pseudo-random layout designed for excellent instantaneous UV coverage. Phase II added 128 additional tiles (for a total of 256), but only 128 can be correlated simultaneously. Phase II therefore operates in two modes: a compact array and an extended array, each consisting of a subset of the 256 total available tiles. In the compact array new tiles were added in two hexagonal cores (see Section 2), providing a hybrid data set with both redundant baselines and the excellent imaging characteristics of the existing MWA array (Beardsley et al. 2012). We use data from this unique array to directly compare redundant and sky-based calibration.

The structure of the remainder of this paper is as follows. In Section 2, we further describe the compact array of Phase II of the MWA and the observations used in our analysis. In Section 3, we describe the calibration techniques used to perform both sky-based and redundancy-based calibration. We also develop and present new tools needed to map between the calibration approaches (Section 3.3.2). In Section 4, we present the results of applying redundant calibration to observations of the ORBComm satellite system, and in Section 5, we directly compare sky-based and redundant calibration solutions derived from observations of an EOR target field. In Section 6, we explore ways of combining the calibration results and compare the resulting EOR PSs. We discuss potential shortcomings of our analysis in Section 7, and we conclude in Section 8.

2. Observations

2.1. Phase II of the MWA

The MWA Phase II compact array consists of 128 tiles. Each tile includes 16 dual-polarization dipoles, as shown in Figure 1. A total of 72 of the tiles are configured into two hexagons with high redundancy for redundant calibratability and PS sensitivity. The other 56 tiles are arranged with minimal redundancy; 8 of these tiles are located at 200–300 m from the core to provide extended baselines for better imaging and survey capabilities.

The top panel of Figure 2 shows the configuration of all 128 tiles of MWA Phase II; the bottom panel shows the north hexagon, with tile numbers labeled. All tiles in the north hexagon are labeled from 1001 to 1036 (bottom panel in Figure 2), and tiles in the south hexagon are labeled from 1037

to 1072. Due to ground conditions at the MWA site, one of the tiles in the south hexagon (tile 1037) could not be placed at the position where the corner of the hexagon should be, so it is flagged, leaving 71 hexagon tiles and 56 nonhexagon tiles. The hexagon-shaped configuration is designed for two reasons: increased sensitivity on short baselines for PS measurements (Parsons et al. 2012a), and opportunities for redundant calibration.

2.2. The Data

The data we processed in this work are from MWA Phase II compact array observations of the EOR0 field (R.A. = $0°$, decl. = $-27°$) at frequencies of 167–197 MHz (corresponding to a 21 cm redshift of 7.5–6.2). Observations were taken on 2016 November 17, from 11:26 to 13:10, 2016 November 19, from 11:18 to 13:02, and 2016 November 21, from 11:19 to 12:54 (UTC), as well as 2 minutes of ORBComm satellite observations at 134–164 MHz on 2016 September 21 at 18:43 (UTC). The total band was divided into 24 1.28 MHz sub-bands; each sub-band is further subdivided into 32 fine channels with a frequency resolution of 40 kHz. The time of observation per data file was 112 s, with a time resolution of 0.5 s. The data were preprocessed by the COTTER pipeline (Offringa et al. 2015), which uses AOFlagger[20] to flag RFI, reduces data volume by averaging in time and frequency, and converts data into the uvfits format. In this work, we average the EOR0 data to 2 s time integrations and 80 kHz frequency resolution. The ORBComm data were averaged into 4 s time integrations and 40 kHz frequency resolution. We choose the EOR0 high-band observation because this is one of the best-studied fields with the MWA (Beardsley et al. 2016; Carroll et al. 2016). It has low sky temperature and relatively few bright, resolved sources, which leads to better EOR sensitivity (Jacobs et al. 2016). The ORBComm observation is for testing redundant calibration because of its high signal-to-noise ratio (S/N).

In the MWA, an analog beamformer can steer the main lobe of the tile primary beam to change the field being observed. For EOR observations, we use the "drift-and-shift" method, where we observe specific pointings with the beamformer and allow the sky to drift overhead for some duration before repointing. The EOR0 observations we used in this work include five pointings, and each pointing spans 30 minutes.[21] The 2-minute ORBComm observation consists of a single pointing toward zenith.

3. Calibration Techniques

Discrepancies between measured data visibilities and true visibilities can have different causes: instrumental gains, cross talk between tiles, RFI, thermal noise, tile pointing error, ionosphere distortion, etc. In this work, we only consider the contribution from the analog/digital electronics of each tile and mainly focus on the complex antenna-based instrumental gain calibration. In this section, we will briefly show the basic mathematical background of both sky calibration and redundant calibration and describe the specific software packages we use to perform them.

---

[20] http://aoflagger.sourceforge.net/doc/api/
[21] The pointings are labeled as −2, −1, 0, 1, and 2, where 0 corresponds to a zenith pointing. Pointings −2 and 2 have less data, i.e., less than 30 minutes.

Figure 1. MWA Phase II tiles in the Murchison Radio-astronomy Observatory in Western Australia. (Taken by Greg Rowbotham in 2016 June, when Phase II was under construction.)
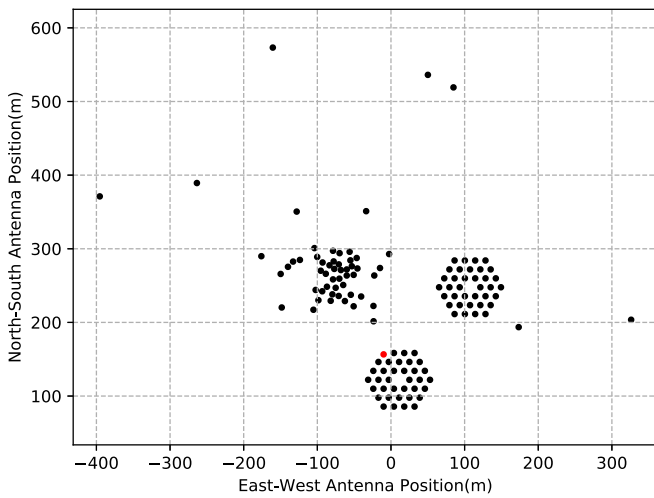
Figure 2. Top: MWA Phase II configuration. Tile 1037 (red) in the upper left corner of the south hexagon was flagged because the site terrain prevented its placement. Bottom: tile positions of the north hexagon, with tile numbers labeled.

3.1. Assumptions

The instrumental calibration is assumed to be tile based. At a given polarization p, given frequency channel ν, and given time step t, the basic assumption of the relation between the measured visibility v_{ij} recorded by the baseline ij (the baseline formed by tile i and tile j) and the true visibility y_{ij} is described by Equation (1), where g_i and g_j are the complex gains of tile i and tile j, respectively, and n_{ij} is a random noise term

$$v_{ij}(t, \nu, p) \approx g_i(t, \nu, p) g_j^*(t, \nu, p) y_{ij}(t, \nu, p) + n_{ij}. \quad (1)$$

In the case of MWA, the tile gains vary from pointing to pointing owing to the change in tile beams. As we observed from real data calibration (using both sky-based calibration and redundant calibration), gains of the same pointing also vary from day to day but are relatively stable over time within one pointing (30 minutes). Barry (2018) has demonstrated that the gain amplitudes are stable if the ambient temperature does not change. Therefore, we assume that 30 minutes of a single pointing is the longest timescale within which we can consider the instrumental gains to be time independent.

Our goal is to solve for the gain per time, per frequency channel, per polarization for each tile using two different methodologies: (1) generate model visibilities based on the combination of our best models for the sky, array layout, and tile primary beam, and then minimize the difference between model visibilities and data (*sky calibration*); and (2) using redundancy, minimize the differences among the measurements from redundant baselines (*redundant calibration*).

3.2. Fast Holographic Deconvolution (FHD) Sky-based Calibration

FHD[22] (Sullivan et al. 2012) is a software package that provides interferometric data simulation, calibration, and imaging. In this paper, we will use the FHD framework as our method to do sky calibration.

In Equation (1), the true visibility y_{ij} consists of foregrounds and EOR signal. We neglect the EOR term because it is orders of magnitude smaller than the foreground term. If we have reasonable knowledge of the foreground sources, we can generate model visibilities m_{ij} and replace y_{ij} with m_{ij} in Equation (1), as Equation (2) shows:

$$v_{ij} \approx g_i g_j^* m_{ij} + n_{ij}. \quad (2)$$

We use a sky model developed by Carroll et al. (2016) specifically for the EOR0 field, which contains about 11,000 point sources in the field of view. We then solve for the gains

---

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

by evaluating χ^2 in Equation (3), making it a least-squares problem, with $2 \times N_{\text{tiles}} - 1$ parameters[23] to solve

$$\chi^2 = \sum_{ij} \frac{|v_{ij} - g_i g_j^* m_{ij}|^2}{\sigma_{ij}^2}. \tag{3}$$

Here σ_{ij}^2 is the noise variance of baseline ij.[24] We solve for each g_i per polarization per frequency channel by feeding an initial guess of the gain solutions (generally all ones by default), fixing all other $g_j (j \neq i)$, and minimizing the χ^2 to get a new guess for g_i, and then we average it with the previous guess of g_i; this average is treated as the solution for g_i, and we then run the previous process iteratively until the solutions converge (Salvini & Wijnholds 2014).

Following this per-tile, per-frequency, per-polarization sky-based calibration, FHD reduces the number of calibration parameters by computing an average bandpass over subsets of the tiles and then only allowing tile-to-tile deviations from this average solution to be smooth in frequency (Barry et al. 2016; Beardsley et al. 2016). The exact form of the final calibration solutions $g_i(\nu)$ for tile i is given by

$$g_i(\nu) = B_c(\nu)[(\alpha_{0,i} + \alpha_{1,i}\nu + \alpha_{2,i}\nu^2)e^{2\pi i(\beta_{0,i} + \beta_{1,i}\nu)} + R_i(\nu)]. \tag{4}$$

$B_c(\nu)$ is a tile-independent bandpass amplitude calculated by averaging the amplitude gains over all tiles that share a cable type. In the MWA Phase II design, each tile has one of four distinct lengths of cable leading from its beamformer to the receiver: 90, 150, 230, or 320 m. This design leads to four subtly different bandpass responses $B_c(\nu)$ owing to different filters used on different cable types and imperfect terminations (Beardsley et al. 2016; Barry 2018). For each tile, deviations from the per-cable type bandpass $B_c(\nu)$ are fit with low-order polynomials in frequency, with coefficients $\alpha_{0,i}$, $\alpha_{1,i}$, $\alpha_{2,i}$, $\beta_{0,i}$, $\beta_{1,i}$ (α for the amplitude and β for the phase). The final parameter, $R_i(\nu)$, is the strongest sinusoidal cable reflection mode found for tile i, which is a complex number. In this work, we only fit $R_i(\nu)$ for 150 m cables, which have the strongest reflection (Barry et al. 2016; Beardsley et al. 2016). The motivation for this fitting methodology is to mitigate frequency-dependent errors introduced by an incomplete sky model, which can lead to foreground contamination of an EOR signal (Barry et al. 2016; Beardsley et al. 2016).

3.3. Redundant Calibration (OMNICAL)

Mathematically, redundant calibration requires sufficient baselines to measure the same Fourier mode of the sky emission so that the there are more measurements than the number of unknown visibilities and tile gains (Liu et al. 2010). In Phase II data, only the two hexagonal subarrays are redundantly calibratable. For one time step, one frequency channel, and one polarization, the unknown parameters consist of tile gains (for those tiles that participate in a minimum number of redundant baselines) and the visibilities themselves for each unique type of baseline. If only redundant baseline

groups containing at least two baselines are considered, there are 71 tiles and 181 unique baseline types, and therefore 252 free parameters to fit, while the number of measurements is 2477.[25]

In this paper, we use the package OMNICAL[26] (Zheng et al. 2014) for redundant calibration. When running OMNICAL on the data, we load uvfits data files using the open source python module pyuvdata[27] (Hazelton et al. 2017).

OMNICAL consists of two algorithms: a logarithmic method (logcal) and a linearized method (lincal; Liu et al. 2010; Zheng et al. 2014). To interpret the algorithms, we express the gain in the form of $g_i = e^{\eta_i + i\phi_i}$, with e^{η_i} the amplitude and ϕ_i the phase of tile i.

In logcal, we linearize the equations by taking the logarithm of Equation (1), where the noise contribution is represented by $\omega_{ij} = \ln\left(1 + \frac{n_{ij}}{g_i g_j^* y_{ij}}\right)$. This gives

$$\ln(v_{ij}) = \eta_i + \eta_j + i(\phi_i - \phi_j) + \ln(y_{ij}) + \omega_{ij}. \tag{5}$$

By separating the real and imaginary parts of Equation (5), the amplitude terms and phase terms are separated. We solve for the gains by minimizing Equations (6) and (7), which are the linear least-squares equations for the amplitudes and phases, respectively.

$$\sum_{ij} [\ln|v_{ij}| - \eta_i - \eta_j - \ln|y_{ij}|]^2 \tag{6}$$

$$\sum_{ij} [\arg(v_{ij}) - \phi_i + \phi_j - \arg(y_{ij})]^2. \tag{7}$$

However, the logcal method is biased. The noise is assumed to be Gaussian and to have zero mean in real/imaginary space, while this is not the case in amplitude/phase space (Liu et al. 2010). To address this issue, lincal is introduced.

In lincal, we perform a Taylor expansion on Equation (1) about some fiducial guess g_i^0 for the gains and y_{ij}^0 for the true visibilities, which leads to Equation (8):

$$v_{ij} \approx g_i^0 g_j^{0*} y_{ij}^0 + g_j^{0*} y_{ij}^0 \Delta g_i + g_i^0 y_{ij}^0 \Delta g_j^* + g_i^0 g_j^{0*} \Delta y_{ij}, \tag{8}$$

where $\Delta g_i = g_i - g_i^0$ and $\Delta y_{ij} = y_{ij} - y_{ij}^0$. This expansion linearizes Equation (1) so that we can employ a least-squares fit to solve for Δg_i and Δy_{ij}. The initial fiducial guess is required to be in a local minimum around the true solution; we use the logcal solutions as the initial guesses for lincal. After we have the solutions for g_i and y_{ij}, we take them as our new fiducial guess and feed them into lincal, and we run this process iteratively. lincal solves in real and imaginary space, so if the noise level for all baselines is the same, the least-squares fit is unbiased.

Before we start the calibration, we have to deal with a phase-wrapping problem: there is ambiguity between 0 and 2π in phase. For example, the difference between a phase of 359° and 1° is 358° instead of 2°. If there is no pre-calibration before logcal, these calibration procedures can potentially take a small difference in phase and drive it in the opposite direction instead of further minimizing it. As a result, the calibration is not handled properly and the solutions do not converge. To

---

[23] The gains are complex, so the number is multiplied by 2. The overall absolute phase parameter is constrained by picking a phase reference tile.
[24] In the FHD framework, the noise does not contribute to the linear least-squares solver, i.e., $\sigma_{ij} \equiv 1$, assuming that all baselines have the same noise level.

[25] All parameters are complex numbers; thus, the number of fitted parameters, as well as the number of measurements, is multiplied by 2.
[26] https://github.com/jeffzhen/omnical
[27] https://github.com/HERA-Team/pyuvdata

overcome phase wrapping, we introduce `firstcal` as our pre-calibration method, i.e., to get an initial estimate of phase solutions.

3.3.1. Firstcal Method

We use the `firstcal` module developed by the Hydrogen Epoch of Reionization Array (HERA) team to find a per-tile delay to provide an initial phase solution[28] using array redundancy, without any reference to the sky. `firstcal` takes visibility pairs v_{ij} and v_{kl} from the same redundant baseline group and calculates the product of v_{ij} and v_{kl}^*. If complex gains of all four tiles differed only by a single per-tile delay, τ_i, then the time average of this quantity, R_{ijkl}, is given by

$$
\begin{aligned}
R_{ijkl}(\nu) &= \langle v_{ij}(\nu, t) v_{kl}^*(\nu, t) \rangle_t \\
&= A^2(\nu) \exp(2\pi i \nu (\tau_i - \tau_j - \tau_k + \tau_l)).
\end{aligned}
\tag{9}
$$

Here ν is frequency and A is visibility amplitude. Multiplying v_{ij} by v_{kl}^* cancels out the frequency structure of the visibilities, leaving only the exponential of the four tile delays. The Fourier transform of Equation (9) along the frequency axis (i.e., the delay transform; Parsons & Backer 2009) should be peaked at

$$
\tau_{\max} = \tau_i - \tau_j - \tau_k + \tau_l.
\tag{10}
$$

With enough visibility pairs, we can produce a set of coupled linear system equations like Equation (10) so that we can solve for all τ simultaneously.[29] Multiplying each v_{ij} by $e^{-2\pi i \nu (\tau_i - \tau_j)}$ flattens the phase across the band. This gives us a reasonably accurate starting point for later calibration that effectively avoids phase wrapping. Since all we require is a reasonable starting point for `OMNICAL`, it is unnecessary to include all redundant baseline pairs into the calculation. The number of all redundant baseline pairs is large (27,032 pairs), while a subset of baseline pairs can be sufficient. We only include baseline type (1001, 1005) and (1001, 1006) (see Figure 2), so that the number of baseline pairs is reduced by a factor of 10 (2970 pairs), which is more computationally efficient.

3.3.2. Degeneracy Projection (DP)

Since redundant calibration does not rely on any information from the sky, there are four intrinsic degeneracy parameters per frequency per polarization that `OMNICAL` cannot constrain: one overall amplitude, which depends on the sky flux density; one absolute phase, which depends on the absolute timing of incoming plane waves; and two rephasing parameters, which correspond to the tip and tilt of the array, or equivalently, the location of the phase center on the sky (Liu et al. 2010; Zheng et al. 2014; Dillon et al. 2017). `OMNICAL` can only be performed in the redundant subset of the MWA Phase II array (71 tiles), and without the degeneracy parameters determined, `OMNICAL` alone cannot provide an absolute calibration. To perform absolute calibration after `OMNICAL`, we use the FHD

---

[28] https://github.com/HERA-Team/hera_cal

[29] This system of equations has a degenerate additive offset (an overall phase) that cannot be solved without an additional constraint. This is equivalent to increasing the length of the cables connecting each tile by the same length. Since this term drops out of any difference $\tau_i - \tau_j$, it is not physically meaningful and can be fixed arbitrarily (e.g., by demanding that all delays average to 0).

calibration solutions as references to constrain the degeneracy parameters. Since FHD calibration is based on a sky model, we take the knowledge of the sky flux density and sky center of FHD as a fiducial guess. We then look for the best-fit four degeneracy parameters per frequency per polarization for the whole array for the `OMNICAL` solutions, which makes them comparable to FHD results. This fitting process is defined as DP.

The fitting for the amplitude parameter is straightforward: for `OMNICAL`, multiplying all g_i by an arbitrary positive constant and simultaneously dividing y_{ij} by the square of that constant does not change the amplitude of $g_i^* g_j y_{ij}$. We correct the amplitude degeneracy parameter by multiplying each `OMNICAL` gain by e^{δ}, where

$$
\delta = \frac{1}{N_{\text{tiles}}} \left(\sum_i \eta_i^{\text{FHD}} - \sum_i \eta_i^{\text{OMNICAL}} \right).
\tag{11}
$$

To illustrate phase degeneracies, we evaluate Equation (12), which is the phase part of Equation (1):

$$
\gamma_{ij} = \phi_i - \phi_j + \theta_{ij},
\tag{12}
$$

where $\gamma_{ij} \equiv \arg(v_{ij})$, $\theta_{ij} \equiv \arg(y_{ij})$. We can add a linear field $\boldsymbol{\Phi} \cdot \boldsymbol{r}_i + \psi$ to ϕ_i and simultaneously subtract $\boldsymbol{\Phi} \cdot (\boldsymbol{r}_i - \boldsymbol{r}_j)$ from θ_{ij}, to get a new set of solutions as defined in Equation (13):

$$
\begin{cases}
\phi_i' = \phi_i + \boldsymbol{\Phi} \cdot \boldsymbol{r}_i + \psi \\
\theta_{ij}' = \theta_{ij} - \boldsymbol{\Phi} \cdot (\boldsymbol{r}_i - \boldsymbol{r}_j).
\end{cases}
\tag{13}
$$

Under this transformation, γ_{ij} in Equation (12) is invariant, as Equation (14) shows (Zheng et al. 2016):

$$
\begin{aligned}
\phi_i' - \phi_j' + \theta_{ij}' &= (\phi_i + \boldsymbol{\Phi} \cdot \boldsymbol{r}_i + \psi) - (\phi_j + \boldsymbol{\Phi} \cdot \boldsymbol{r}_j + \psi) \\
&\quad + (\theta_{ij} - \boldsymbol{\Phi} \cdot (\boldsymbol{r}_i - \boldsymbol{r}_j)) \\
&= \phi_i - \phi_j + \theta_{ij}.
\end{aligned}
\tag{14}
$$

Here $\boldsymbol{r}_i$ is the ideal position of tile i, i.e., tile positions with perfect redundancy. We assume that all tiles are coplanar, $\boldsymbol{r}_i$ is a 2D vector, and thus $\boldsymbol{\Phi}$ is 2D. The absolute phase parameter is given by ψ, and the two rephasing parameters are given by the 2D vector $\boldsymbol{\Phi}$.

We define $\Delta\Psi_i \equiv \arg(g_i^{\text{FHD}}/g_i^{\text{OMNICAL}})$. Equation (15) shows the relation between calibration solutions and phase degenerate parameters

$$
\Delta\Psi_i = \Phi_x x_i + \Phi_y y_i + \psi,
\tag{15}
$$

where $(x_i, y_i) = \boldsymbol{r}_i$. Equation (15) is a function of a plane. The basic idea of solving for (Φ_x, Φ_y, ψ) is to fit a plane in $(x_i, y_i, \Delta\Psi_i)$ space. This is the process of phase DP. The fitting details are described in Appendix A.

4. Observations of ORBComm

As a first test of `OMNICAL` on MWA Phase II data, we investigate observations at 137.1 MHz, where the ORBComm satellite system transmits (Neben et al. 2015, 2016; J. L. B. Line et al. 2018, in preparation), because this data set has extremely high S/N. Since the MWA has a wide field of view, it is difficult to point to a patch of sky with a flux density dominated by one bright point source. However, an ORBComm satellite provides a good opportunity to observe a "point source" because its signal
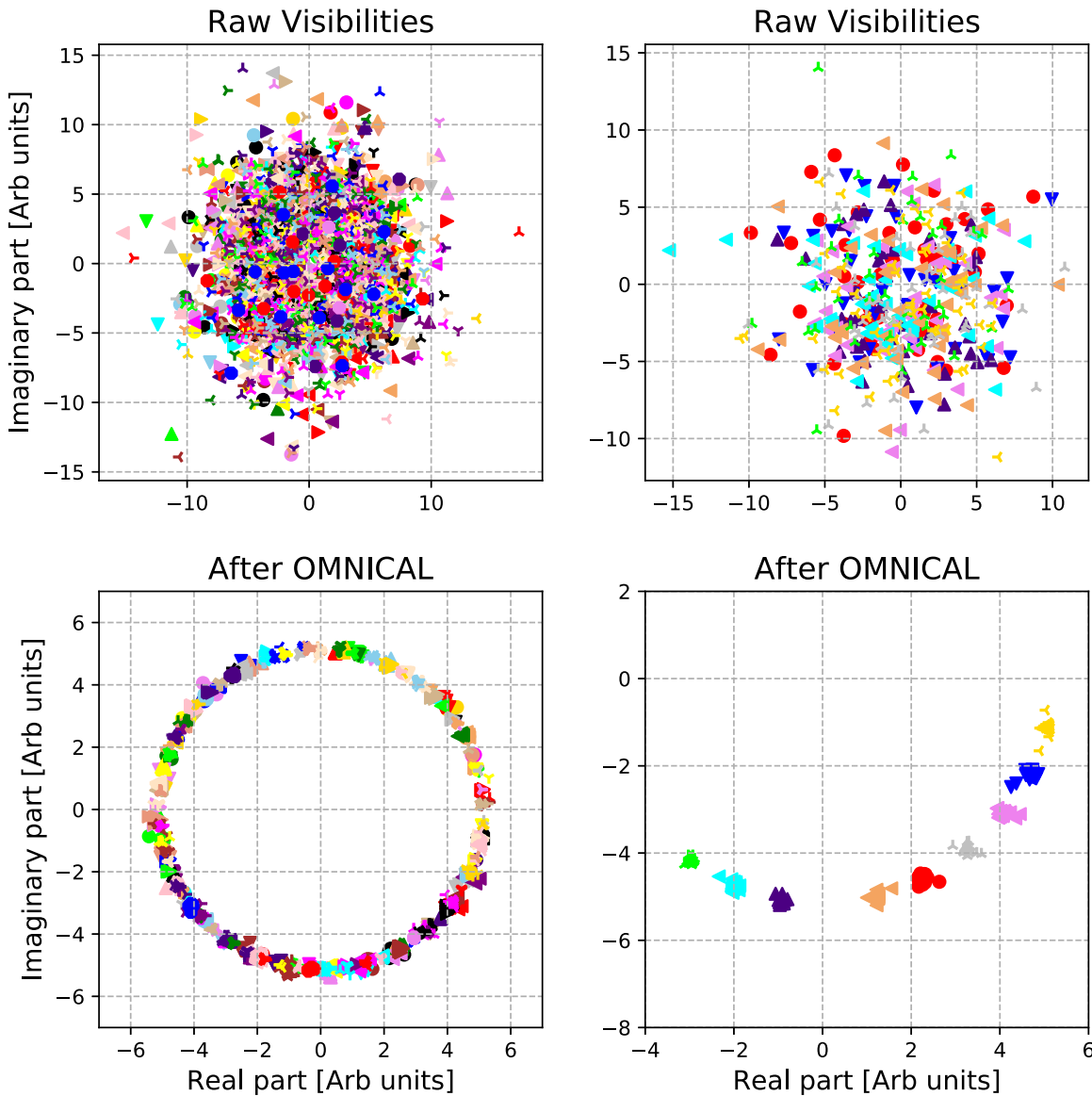
Figure 3. Complex visibility plots of ORBComm observation at 137.1 MHz, 4 s of data. Each unique combination of color and symbol represents visibility measurements from a unique baseline type. Top left: raw visibilities from all redundant baseline groups; bottom left: calibrated visibilities from all redundant baseline groups; top right: raw visibilities from nine baseline types; bottom right: calibrated visibilities from nine baseline types. The units are arbitrary because no absolute calibration is performed.

is orders of magnitude brighter than any other sources in the sky at 137.1 MHz. The near-infinite S/N measurements on ORB-Comm are an excellent opportunity to quantify the uncertainties in the redundant calibration procedure (Zheng et al. 2014).

Figure 3 shows the OMNICAL results on observations of an ORBComm satellite with the MWA Phase II hexagons on 2016 September 21. Each unique combination of color and symbol represents visibilities measured by a redundant baseline group. The top left panel shows the complex visibilities from all redundant baseline groups before OMNICAL, and the bottom left panel shows the same set of data after calibration.

The constant amplitude of visibilities in the bottom left panel indicates a delta function in the image domain, which agrees with our point-source expectation of ORBComm. We pick nine unique baseline groups as representatives from the left column in Figure 3 and show the uncalibrated (top right) and calibrated (bottom right) visibilities in the right column. This illustrates

that OMNICAL makes visibility measurements from baselines with the same length and orientation cluster together, i.e., OMNICAL is performing as expected: it makes visibilities from physically redundant baselines agree with each other. The level of the standard deviation within each redundant visibility group is 1% comparing to their magnitudes, which is possibly due to the nonperfectly gridded antenna positions. This quantifies the systematic uncertainty of the redundant calibration procedure for the MWA Phase II array, or in other words, this level of disagreement is the best that redundant calibration can achieve.

5. Comparison between FHD and OMNICAL

In this section, we will take MWA Phase II observations targeting the EOR0 field as an example to show the comparison between FHD sky calibration and redundant calibration. All
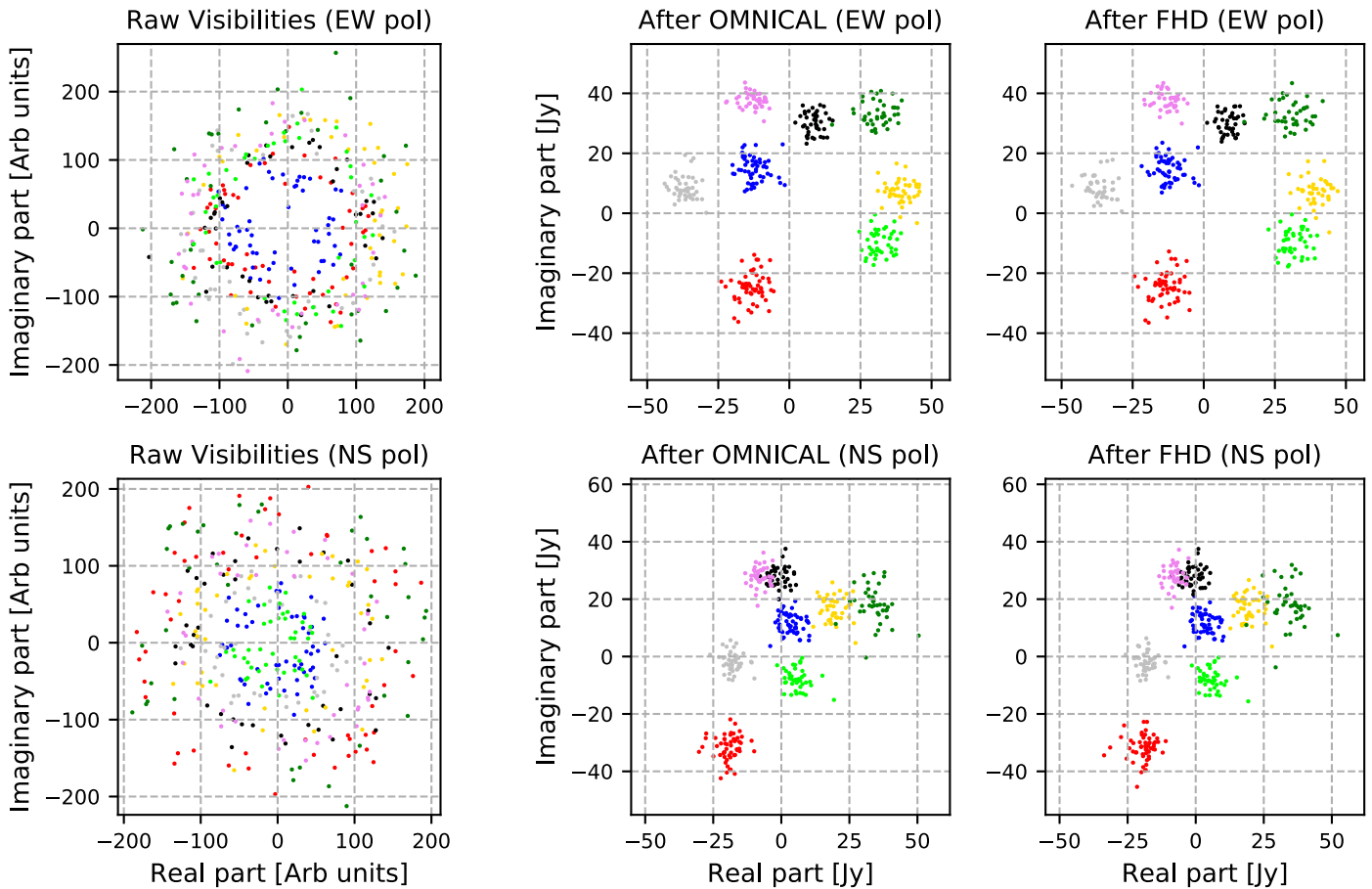
Figure 4. 112 s averaged complex visibility plots of EOR0 data at 191 MHz from eight redundant baseline groups. Each color represents one redundant baseline group. Left column: raw visibilities, with arbitrary units; middle column: visibilities after OMNICAL, with degeneracy parameters projected (units: Jy); right column: visibilities after FHD sky calibration (units: Jy). Top row: east–west polarization; bottom row: north–south polarization.

calibrations are performed per data file (every 112 s), and the gains are assumed to be time independent within a data file. In FHD calibration, the sky model is a point-source catalog specifically developed for the EOR0 field (Carroll et al. 2016). All time steps (2 s integrations) are fed into the linear least-squares solver, which minimizes the difference between data and model and returns one set of time-independent calibration solutions per file.

In OMNICAL, we average the data along the time axis of each data file, i.e., average every 2 minutes of data before calibrating for two technical purposes: increasing S/N for better redundant calibration performance (Liu et al. 2010), and excluding sparse flagged data samples without dramatically increasing computational expense.[30]

We also exclude the baseline type (1001, 1002) (the index refers to Figure 2), which is the 14 m east–west baseline type, because we have seen significant systematics from that baseline group. The visibility variances in this group are about 6 times larger than other redundant baseline groups. This could be due to strong cross talk between tiles, or because the Galactic plane aligns with these baselines (see Thyagarajan et al. 2015), enhancing the effect of tile-to-tile beam variations across the array (Noorishad et al. 2012). The reason is still unclear, but it is a topic to be investigated in future work.

---

[30] In OMNICAL, explicitly excluding flagged baseline samples per time and frequency requires generating distinct linear equations per time and frequency instead of per data file, which is computationally infeasible.

5.1. Visibility Clustering

The redundant baselines should measure the same Fourier mode of the sky regardless of the calibration procedures involved. Evaluating how visibilities measured by redundant baselines agree with each other (visibility clustering) is an approach to evaluating calibration methods. Figure 4 shows 112 s averaged complex visibilities at 191 MHz observed on 2016 November 21. We plot the visibilities for eight types of baselines with lengths below 20 wavelengths, which are of most importance for EOR sensitivity, at 180 MHz. Visual inspection shows substantial agreement between the two methods. Quantitatively, visibilities after OMNICAL (middle column) are in better agreement than FHD (right column) (about 6%–30% reduction in the standard deviation of a cluster). One explanation for this effect is that in FHD calibration, baselines shorter than 50 wavelengths at 180 MHz are omitted (due to the difficulties in modeling diffuse emission; Patil et al. 2016). Thus, short baselines (like those plotted here) have less weight in FHD calibration; OMNICAL uses the information of these short baselines. OMNICAL also explicitly minimizes the variance within redundant visibilities; thus, it should lead to better visibility clustering than alternative methods. Although this metric does not necessarily indicate a better calibration, it shows that it is possible to put more weight on the most EOR-sensitive baselines, instead of calibrating with only long baselines with low EOR sensitivity as is currently required for sky calibration.
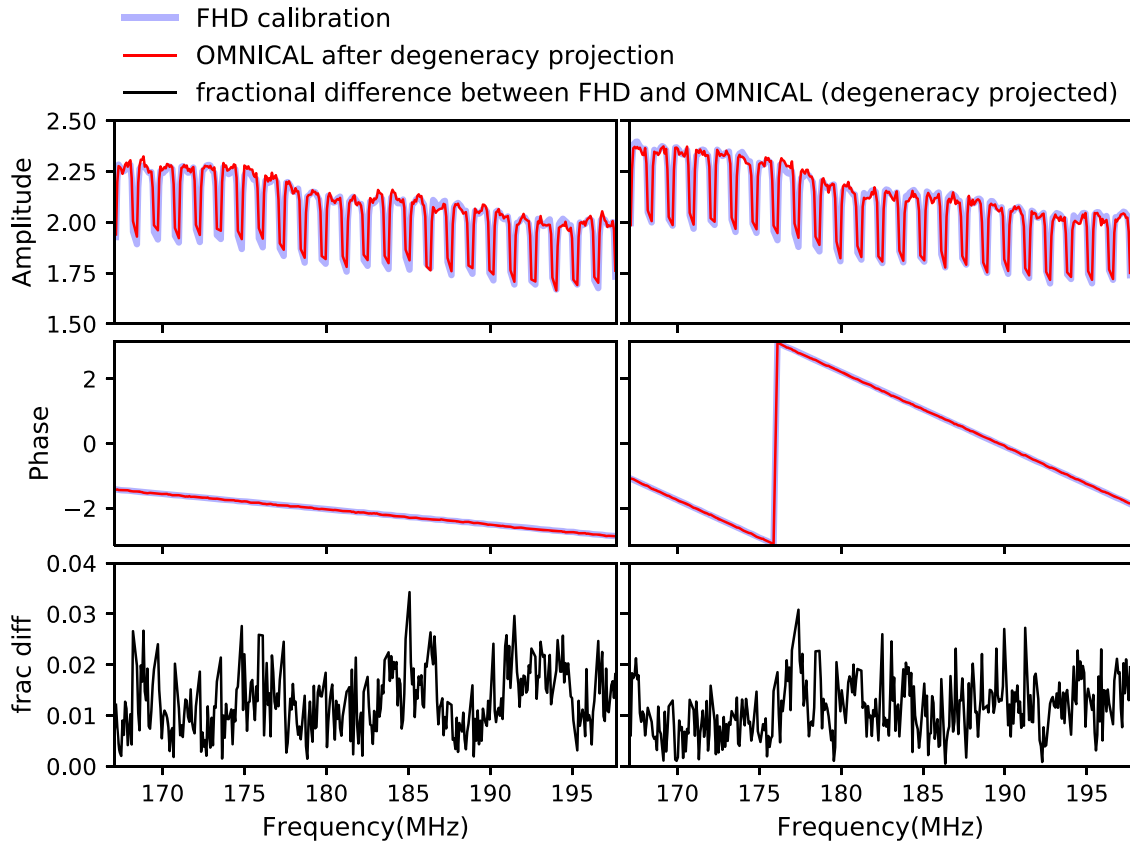
Figure 5. 30-minute averaged gain solutions of tile 1024 (left column) and tile 1064 (right column) from zenith pointing, east–west polarization. Top: gain amplitude; middle: gain phase; bottom: fractional difference between FHD solution and OMNICAL solutions with degeneracy projected. Blue: FHD solutions; red: OMNICAL solutions after projecting degeneracy. The fractional difference in the bottom panels is calculated by dividing the amplitude of the complex difference between the two by the amplitude of FHD solutions.

5.2. Direct Comparison

Figure 5 shows a direct comparison between FHD solutions and OMNICAL solutions after DP. The data are from a 30-minute zenith pointing on the EOR0 field, and calibration solutions have been averaged over the entire pointing. Figure 5 shows solutions for tile 1024 and tile 1064 in gain amplitudes (top) and phases (middle) over frequency, as well as fractional difference between solutions from these two approaches (bottom).

The first conclusion is that the bandpass structures from both approaches show consistent results at a level of 98%. However, the 2% level of difference between the two is not negligible as far as the EOR signal is concerned. For each 1.28 MHz sub-band, the frequency channels near the band edges appear to show relatively larger differences. The differences in solutions can come about not only because they are derived with different algorithms using different assumptions but also because they use different subsets of the data to perform the calibration, i.e., FHD uses data from long baselines, while OMNICAL uses data from redundant baselines. We will investigate the effects of this level of difference on EOR PS measurements in Section 6.

6. Combining FHD and OMNICAL

FHD performs well on calibrating EOR0 data with a well-developed point-source catalog (e.g., Beardsley et al. 2016; Barry 2018), but it also has shortcomings, including errors introduced by an incomplete sky model (Barry et al. 2016) and a loss of sensitivity from excluding short baselines owing to difficulty in modeling diffuse sources (Bowman et al. 2009;

Sullivan et al. 2012; Patil et al. 2016). OMNICAL is free of sky model error and is able to calibrate short baselines (although, as noted, we exclude the shortest 14 m east–west baselines from redundant calibration because they exhibit significantly larger scatter than other redundant baseline types), but it cannot solve for the degenerate parameters, and it can only calibrate a subset of the array. In addition, OMNICAL has the potential for error introduced by tile position inaccuracies and beam variation from tile to tile.

Their respective advantages and disadvantages, however, suggest that OMNICAL and FHD can be mutually complementary. We can possibly use the algorithms to mitigate both sky model and nonredundancy errors. These two methods also allow us to make use of more baselines for calibration, since FHD excludes short baselines and OMNICAL only can calibrate antennas in the redundant subset of the array.

With bad tiles excluded (tile 45 and tile 1037 are not operational in our data set), there are 71 hexagon tiles and 55 nonhexagon tiles. In FHD calibration, if we only calibrate baselines longer than 50 wavelengths at 180 MHz, the number of baselines we use is 5653. For the combined calibration, there are 2477 baselines involved in redundant calibration, 1235 of which are shorter than 50 wavelengths; thus, 6888 baselines can be used in calibration.

In this section, we propose two strategies to combine FHD with OMNICAL. As our metric for evaluating different approaches, we use the 2D $(k_\perp, k_\parallel)$ PS common to 21 cm EOR analyses.[31] A

---

[31] Our PS estimator (discussed below) uses all baselines, so it is necessary to combine both FHD and OMNICAL to get calibration solutions for both the hexagon and nonhexagon tiles. Hence, we do not use the PS metric to compare the independent solutions from FHD and OMNICAL in the previous section.
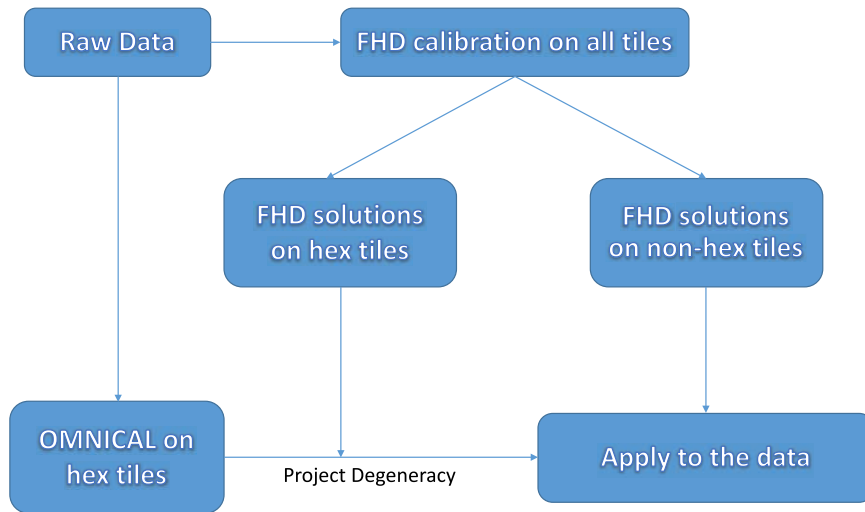
Figure 6. Flow diagram showing the procedure of OFcal.

schematic 2D PS is shown in the top left panel of Figure 7. The power in the lower red region in $k_\parallel$ for all $k_\perp$ is dominated by the intrinsically spectrally smooth foregrounds. The instrument chromaticity mixes foreground modes up to higher $k_\parallel$, forming into a "foreground wedge." The limit of the wedge depends on how far the sources are from the center of the field of view and increases on longer baselines (larger $k_\perp$). The solid line and dashed line represent the horizon limit and the primary field-of-view limit, respectively. The remaining "EOR window" is foreground free and expected to contain a wealth of information about the 21 cm signal (Datta et al. 2010; Morales et al. 2012; Parsons et al. 2012b; Trott et al. 2012; Vedantham et al. 2012; Hazelton et al. 2013; Pober et al. 2013; Thyagarajan et al. 2013; Liu et al. 2014; Barry et al. 2016). In the "EOR window," any observed excess of power is a contaminant, as the EOR signal is buried deep in the noise. Our metric of evaluating calibration techniques is to quantify their performances of mitigating power contamination in the "EOR window." Not only is this metric the quantity of interest (a major goal of MWA Phase II is to measure the PS of the EOR), it also highlights subtle differences between the calibration schemes due to its inherent sensitivity to spectral structure that can corrupt EOR measurements.

To create our PS, we use the software package Error Propagated Power Spectrum with Interleaved Observed Noise (ϵppsilon,[32] which calculates the PS using image cubes as input with errors propagated through the full analysis; Jacobs et al. 2016).

6.1. Strategies

We propose two simple strategies to combine OMNICAL with FHD by running them sequentially: "OMNICAL first, FHD second" and "FHD first, OMNICAL second." To simplify, we name OMNICAL first, FHD second as OFcal, and FHD first, OMNICAL second as FOcal. Since OMNICAL can only calibrate the subset of the array, no matter what strategy we propose, these hybrid approaches only change the calibration on hexagon tiles; the calibration of nonhexagon tiles remains the same as FHD calibration results.

---

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

6.1.1. OFcal

The OFcal approach is illustrated in Figure 6. The calibration procedure is as follows:

1. Run OMNICAL on raw visibilities measured by baselines within hexagon tiles.
2. Perform FHD calibration on raw visibilities measured by all baselines longer than 50 wavelengths at 180 MHz.
3. Average OMNICAL solutions for each pointing (30-minute time average).
4. Project degeneracy parameters of OMNICAL solutions to FHD solutions.
5. Apply degeneracy-projected, time-averaged OMNICAL solutions to tiles within the hexagons and apply FHD solutions to all other tiles.

When averaging calibration solutions from a single pointing, we first make sure this set of solutions has the same degeneracy parameters. We do this by picking one data file solution as a target, projecting degeneracy of solutions from other data files to this target, and then averaging.

6.1.2. FOcal

The description of FOcal is simpler:

1. Perform FHD calibration on raw visibilities measured by all baselines.
2. Apply FHD solutions to the raw data.
3. Run OMNICAL on FHD-calibrated visibilities (from baselines within the hexagons).
4. Average OMNICAL solutions for each pointing (30-minute time average).
5. Project degeneracy parameters of OMNICAL solutions to default values of 0.
6. Apply time-averaged OMNICAL solutions to FHD-calibrated data.

Since the data before OMNICAL are already calibrated by FHD, the degeneracy is removed in a different way. By forcing the average of η values of all tiles to be 0 (similar to Equation (11), but with FHD terms excluded), the flux density scale set by FHD does not change, and by making the linear
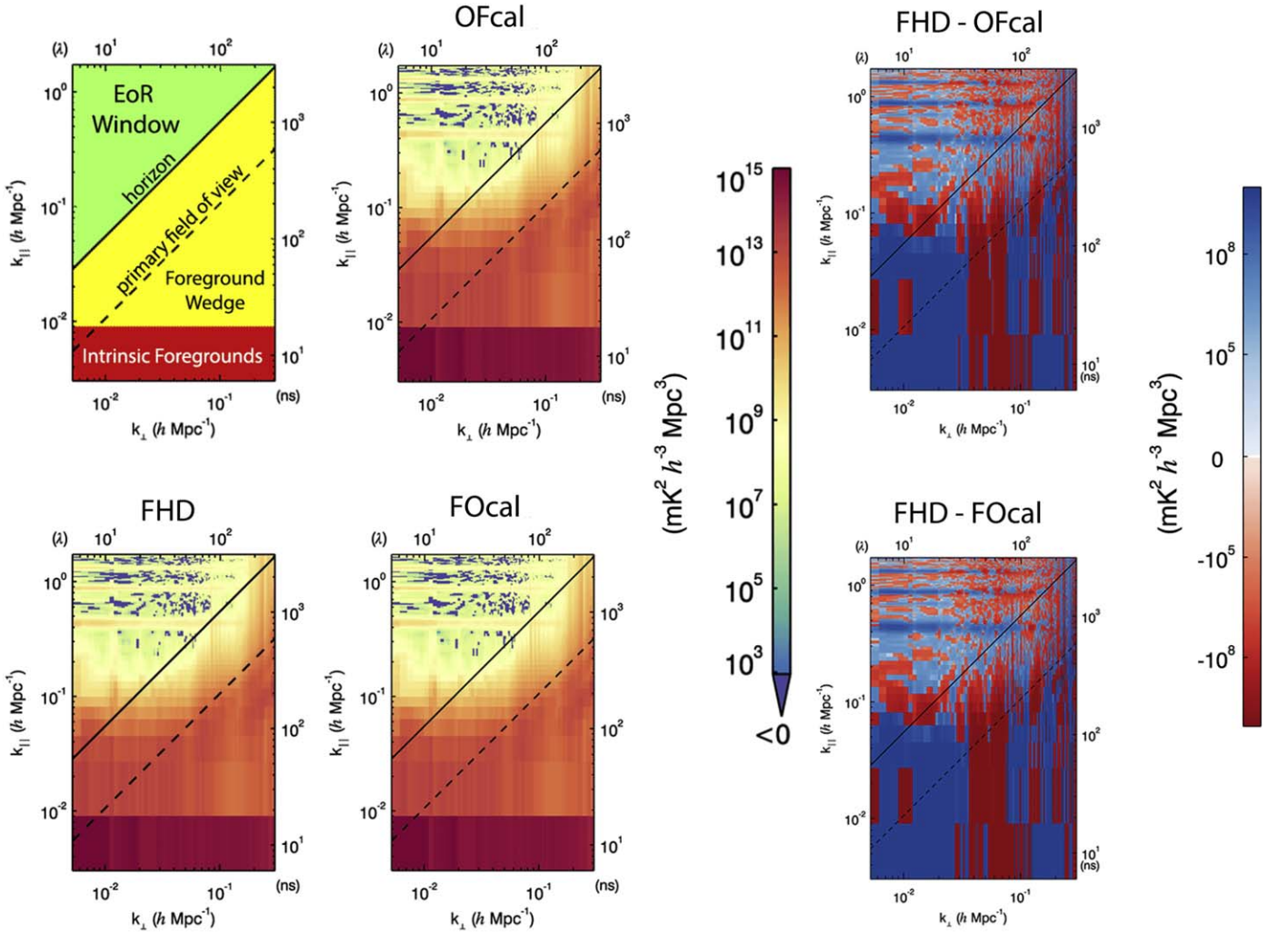
Figure 7. Top left: schematic plot of 2D cylindrical PS. Low-$k_\parallel$ modes are dominated by intrinsic foregrounds, and the chromaticity of the interferometric instrument smears foregrounds contamination up to high $k_\parallel$, leaving an "EOR window" that is foreground free. Bottom left: PS after FHD calibration. Top middle: PS after OFcal. Bottom middle: PS after FOcal. Top right: difference PS of FHD minus OFcal. Bottom right: difference PS of FHD minus FOcal. See text for details on the calibration methods.

field Φ have zero slope and setting the average phase to be 0, the sky center does not change. More details are described in Appendix A.

The basic difference between OF and FOcal is that each individual baseline has different weights in these two cases. When constructing χ^2 for `lincal`, with each individual term being $|v_{ij} - g_i g_j^* y_{ij}|^2$, we will see baselines with larger $|g|$ having larger noise level. In OFcal, to avoid any bias, we weight each v_{ij} by the reciprocal of the product of the square root of autocorrelations of tile i and tile j, which effectively cancels out the gain amplitude differences. In FOcal, since the amplitude calibration of FHD is already applied before `OMNICAL`, we do not apply this weighting. In the noiseless case, we expect both approaches to yield the same result, but in the presence of noise the best weighting for these methods is an open question.

6.2. Results

A PS comparison of OFcal and FOcal with the FHD-only calibration for the north–south polarization data is shown in Figure 7. The PS of the data with only FHD calibration applied

is shown in the bottom left panel. The middle column shows the PS of the data with the two new calibration schemes applied (OFcal on top, and FOcal on the bottom).

The three PSs from FHD, OFcal, and FOcal have common features and are nearly indistinguishable. The horizontal streaks of excess power shown in PS plots are harmonic modes due to flagged channels between every 1.28 MHz sub-band. The vertical streak at ~12 wavelengths is due to sparse sampling in $k_\perp$ space, or in other words, we do not have baselines sampling those modes.

To illustrate the difference between OFcal (or FOcal) and FHD, we make difference PS plots, shown in the right column of Figure 7. The difference plots are obtained by subtracting (in 3D k space) the PS of data with OFcal applied (top right) and FOcal applied (bottom right) from that of the FHD-only-calibrated data. In PS difference plots, red indicates an excess of power in the OFcal (or FOcal) strategy, and blue indicates a reduction of power when compared with FHD. From the difference plot, we can conclude that both OFcal and FOcal show lower power at sub-band harmonic modes than FHD. We expect this improvement at sub-band harmonic modes because
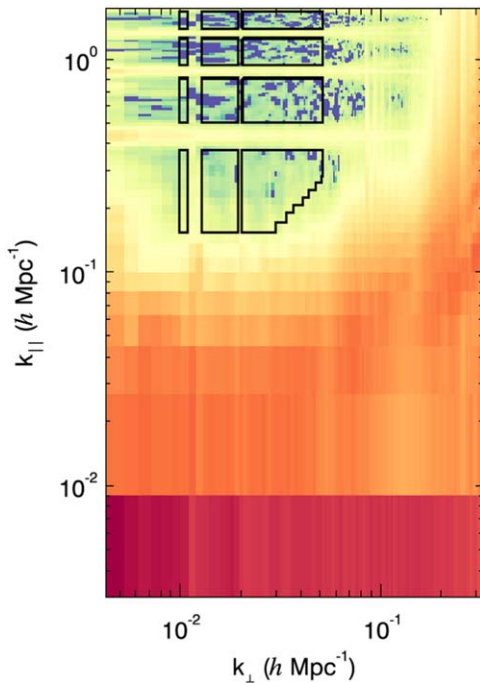
Figure 8. 2D PS using calibration from FHD only (Figure 7, bottom left) with contours to highlight modes that will be used for 1D power in $k_\parallel$ in Figure 10.

difference is shown in Figure 10 (green). We conclude that the differences between OFcal and FOcal are centered around zero and much less significant overall.

We note that in all cases these differences are *below* the thermal noise level in our measurements. However, in creating these difference PSs, we are subtracting the same data—with the exact same realization of the noise—only with different calibrations applied. If the calibrations were the same, the differences would be identically zero. Since the goal of these experiments is to detect the 21 cm signal from the EOR, the typical amplitude of the EOR signal—approximately $10^6 \, \mathrm{mK}^2 \, h^{-3} \, \mathrm{Mpc}^3$ at $k \sim 0.1 \, h \, \mathrm{Mpc}^{-1}$ (Furlanetto et al. 2006; Mesinger et al. 2011)—provides a rough scale for assessing the significance of our improvements. Using redundant calibration in addition to FHD (through either FOcal or OFcal) removes foreground contamination at or above the level of the EOR signal. The differences between FOcal and OFcal are much smaller and are thus unlikely to be significant for EOR experiments.

7. Discussion

Section 6.2 has shown that our hybrid approaches (OF/FOcal) can improve the PS in the EOR window. The intuition for this improvement is that tiles in the two hexagons were calibrated based on redundant baseline assumption; thus, all nondegenerate parameters are then free from the sky model error described in Barry et al. (2016). The four degenerate parameters per frequency channel per time per polarization still are sky model dependent, but overall we expect to have mitigated the error introduced by the imperfect sky model. Additionally, FHD only uses long baselines for calibration, which could potentially overfit gain parameters to noises in long-baseline data; however, we are more interested in short-baseline data in EOR observations, and we expect OF/FOcal to mitigate this effect by including short baselines in calibration. Although we are ignoring cross talk and ionospheric effects in this work, all calibration methods work on the same data and therefore have the same challenges, and nonetheless a noise power reduction was achieved by our hybrid approach.

So far we have not considered systematic errors that can affect `OMNICAL`. Redundant calibration is based on two assumptions: that redundant baselines have the exact same length and orientation and that all tile beams are identical. These two assumptions are not exactly true in practice. According to MWA Phase II baseline coordinates, position deviations from perfect redundancy are relatively small (at a level of 5 cm). We have performed noiseless foreground simulations to study the effect of systematic errors introduced through redundant calibration using the imperfect tile positions of Phase II. As we see in the simulated data, the so-called "redundant" visibilities are not identical, but we assume that they are when we do calibration. We found that the errors introduced to the PS by the "wrong" redundancy assumption are unbiased, as well as below the typical EOR level by 2 orders of magnitude. Beam variation can also be significant. It is also a possible cause for the large systematic disagreement for baseline type (1001, 1002) (east–west 14 m baselines) that we saw in these data. In future work, we will explore the error introduced by both effects through detailed calibration simulations, similar to Barry et al. (2016).

In our analysis, we performed `OMNICAL` on each data file after averaging it over the time axis (2 minutes), which gives a better S/N in calibration and allows us to conveniently avoid flagged samples. However, there is a concern of washing signals out for

the channels near sub-band gaps seem to show the most tile-to-tile variation. `OMNICAL` is capable of capturing this variation, while FHD only fits a smooth polynomial function in frequency (after dividing by a cable-averaged bandpass) to capture tile-to-tile variation (see Section 3.2). Because these variations appear on the sub-band scale of 1.28 MHz, FHD cannot calibrate them out as well as `OMNICAL`.

To further investigate the PS differences in the EOR window, we pick regions of k space that are free from foregrounds and sub-band contaminations in the 3D PS.[33] We illustrate these cuts in Figure 8, where the contamination we are excluding is evident. To more clearly demonstrate improvements from the combined calibration technique, we apply this k space cut to the PS differences shown in Figure 9 and average in $k_\perp$ to make a 1D power difference versus $k_\parallel$, which we show in Figure 10. We have excluded low-$k_\parallel$ modes that are foreground contaminated (dark gray), as well as sub-band harmonic modes (light gray). Figure 10 shows that OF/FOcal both show less contamination than FHD in general (i.e., the differences in Figure 10 are mostly positive). Both hybrid approaches show better performance at 150 m reflection modes (the expected $k_\parallel$ value for a 150 m cable reflection at redshift 7 is marked by the vertical dot-dashed line; Ewall-Wice et al. 2016). We can also see some improvements near the 230 m cable reflection mode marked by the vertical dashed line.

We also see that both OFcal and FOcal have strikingly similar differences with FHD (i.e., the two difference PSs in the right column of Figure 7 are very similar), although they are not identical. A difference PS plot of FOcal minus OFcal is shown in Figure 9 on the right, and the 1D version of this PS

---

[33] Similar selections were used in Beardsley et al. (2016), although we exclude select values of $k_\perp$ where the Phase II baseline sampling is poorer than in Phase I.
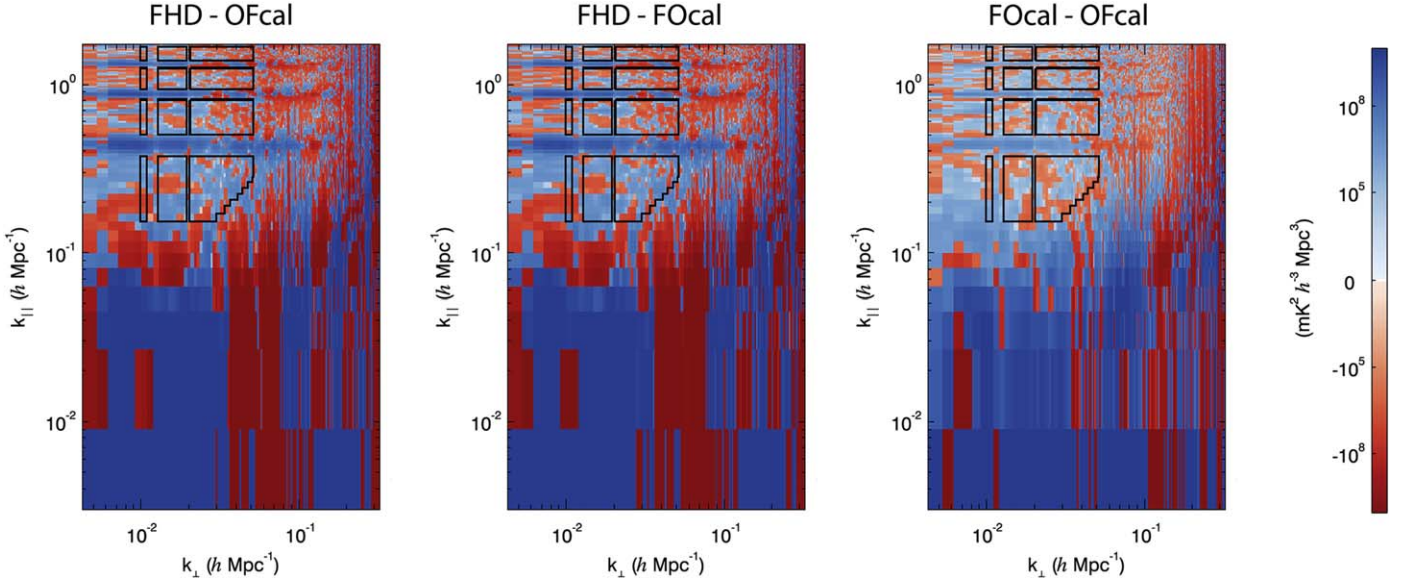
Figure 9. PS difference plots of FHD minus OFcal (left; see Figure 7, top right), FHD minus FOcal (middle; see Figure 7, bottom right), and FOcal minus OFcal (right) with the same contours as in Figure 8.
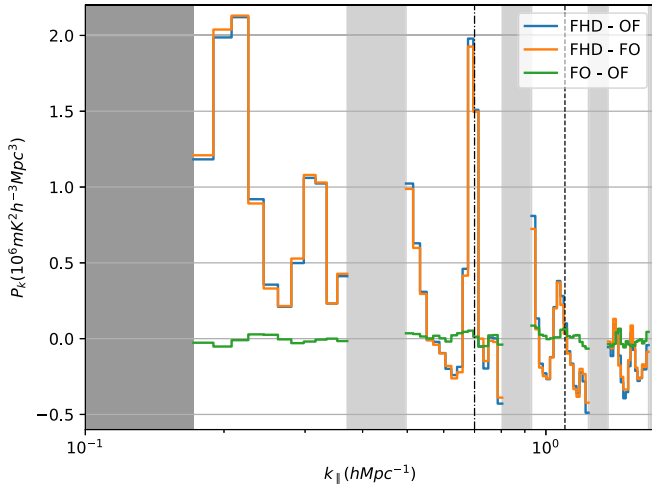

Figure 10. 1D difference PS vs. $k_{\parallel}$ made from the subset of modes illustrated in Figure 9. Blue: FHD only minus OFcal; orange: FHD only minus FOcal; green: FOcal minus OFcal. The dark-gray shaded region indicates low-k modes that are foreground contaminated; the light-gray shaded regions are sub-band harmonic modes we cut out. The vertical dot-dashed and dashed lines highlight the 150 and 230 m cable reflection modes, respectively.

relatively long baselines. We investigated three averaging scenarios through noiseless foreground simulations using real baseline coordinates where there are no flagged samples: calibrating visibilities each 2 s interval and then directly applying these solutions to the data; averaging calibration solutions derived for each 2 s interval over 2 minutes and then applying the time-averaged solutions to the data; and, most similar to the analysis performed here, averaging 2 minutes of data and then calibrating using the averaged data and applying these to the unaveraged data. By evaluating the PS as we did in Section 6.2, we conclude that none of the three scenarios show bias relative to others, and the amplitudes of differences among them are 3 orders of magnitudes lower than the typical EOR level. This validates our averaging strategy used in redundant calibration in the real data.

8. Conclusion

We have explored the application of both sky-based calibration and redundant calibration to data from Phase II of the MWA and investigated their respective trade-offs and possible complementarity. Sky calibration is model dependent, and a reasonable calibration requires a fairly good model of the radio sky. The sky model, as well as the beam model, cannot be perfect to a certain level. Errors such as wrong source positions, brightness errors, or missing sources can potentially introduce calibration error to the PS (Barry et al. 2016). In addition, since the sky model in FHD is a point-source catalog and it is difficult to model diffuse sources, the short baselines are omitted (Bowman et al. 2009; Sullivan et al. 2012; Patil et al. 2016), which leads to a loss of information of those baselines in calibration. Redundant calibration provides an opportunity to remedy these shortcomings: it is sky model independent; thus, it is not restricted by baseline length. However, redundant calibration leaves four intrinsic degeneracy parameters unsolved. In addition, redundant calibration may also be contaminated by tile position error and beam variation (Liu et al. 2010). Section 6.2 shows that using redundant calibration and sky-based calibration together can alleviate the potential error introduced by assumptions these two approaches made. We aim to make use of the advantages of both calibration approaches and combine them together to improve our calibration.

In this paper, we have shown the success of OMNICAL on ORBComm observations from MWA Phase II and compared OMNICAL and FHD on EOR0 data, showing consistent results from these two approaches. This is the first time these two independent methods have been confirmed to agree in real data calibration. We further attempted to combine FHD with OMNICAL in two ways: OMNICAL first, FHD second (OFcal), and FHD first, OMNICAL second (FOcal). By comparing them with FHD in the PS scheme, we conclude that both OFcal and FOcal show improved behavior in the k modes with the most EOR sensitivity in the PS, especially in modes contaminated by 150 and 230 m cable reflections.

This result substantially improves on similar comparisons in the literature. Noorishad et al. (2012) use redundancy between

individual dipole elements within a LOFAR phased-array tile, but the array has little to no redundancy between tiles. Nikolic et al. (2017) use a point-source model for the Galactic center to calibrate the 19-element, highly redundant HERA commissioning array, but they present no comparisons with redundant calibration methods. When the complete 350-element HERA is finished, however, it will be a valuable tool for performing studies similar to the one presented here (Dillon & Parsons 2016).

Appendix A
Degeneracy Projection

Section 3.3.2 describes the four intrinsic degeneracy parameters per polarization per frequency per time in redundant calibration. In this section, we describe details about how we treat these degeneracy parameters in redundant calibration for MWA Phase II data.

The DP technique introduced in our work is a process where we look for the best-fit four degeneracy parameters for input solutions (e.g., OMNICAL solutions) that make them comparable to the target solutions (e.g., FHD solutions).

We perform DP in two cases. First, redundant calibration cannot provide a correct answer for these degeneracy parameters, necessitating an absolute calibration after OMNICAL. We do this by projecting OMNICAL solutions in the degenerate space to FHD solutions because FHD is our best guess about the sky information. The other case is that when we are averaging OMNICAL solutions from a set of adjacent observations (here an observation refers to a single 112 s file), the degeneracy parameters may vary slightly from observation to observation. Simply averaging solutions with inconsistent degeneracy parameters can bias the average in an unknown direction. We pick one observation's solutions to be the target solutions and project the degeneracy of the other observation solutions to this target before averaging them.

A.1. Degeneracy Description

DP for gain amplitudes is straightforward. The η values of the input solutions are chosen to have the same average over all tiles as that of the target solutions per polarization per frequency.

As we mentioned in Section 5.1, there are three degeneracy parameters in phase. There is actually one extra phase offset degeneracy parameter for the MWA Phase II array. Since there is no interhexagon tile sharing the same baseline type as any intrahexagon baseline, adding a uniform phase offset to gains of all tiles in one of the hexagons does not break any visibility redundancy. Thus, the offset terms ψ are treated separately for the north hexagon (ψ_N) and south hexagon (ψ_S), while the phase slope Φ is the same for both hexagons.

To solve for these phase parameters, as mentioned in Section 5.1, we can fit a plane in (x, y, $\Delta\Psi$) space, where $\Delta\Psi$ is the phase difference between the input gain solutions and target gain solutions

$$\Delta\Psi_i = \Phi_x x_i + \Phi_y y_i + \psi. \qquad (16)$$

However, when we difference the phase between two complex numbers, the outcome can have a 2π ambiguity, i.e., a phase wrap. If the phase wrapping happens frequently, we are not able to directly fit a plane as Equation (16).

A.2. DP without Phase Wrapping

We first consider the case where there is no phase wrapping. Practically, this case occurs in two places in our analysis. First, when we do FOcal, we are running OMNICAL on FHD-calibrated data. If FHD were to calibrate it again, it should return ones as the solutions (i.e., no calibration needed). Thus, DP in FOcal is equivalent to projecting OMNICAL solutions to ones, or in other words, we do not want OMNICAL to add extra nonzero values to degeneracy parameters that have already been calibrated by FHD. Since in this case OMNICAL is looking for solutions around 1.0, the phase differences between OMNICAL solutions and 0.0 are small, so there is no phase wrapping.

The second place where phase wrapping is not an issue is when comparing OMNICAL solutions from adjacent observations. These solutions are very similar to each other, and when we do DP between close solutions, we do not have to worry about phase wrapping. It is safe to directly apply plane fitting.

To calculate the best fit for Equation (16), we minimize the quantity in Equation (17):

$$\chi^2 = \sum_{i_N} (\Delta\Psi_{i_N} - \Phi_x x_{i_N} - \Phi_y y_{i_N} - \psi_N)^2$$
$$+ \sum_{i_S} (\Delta\Psi_{i_S} - \Phi_x x_{i_S} - \Phi_y y_{i_S} - \psi_S)^2, \qquad (17)$$

where i_N is the tile index in the north hexagon and i_S is the tile index in the south hexagon. An example of the fitting result in FOcal at a single frequency and single polarization is shown in Figure 11.

A.3. DP with Phase Wrapping

Now we discuss the case where phase wrapping shows up frequently. This happens when we do OFcal. We project OMNICAL solutions on raw data to FHD solutions. The phase
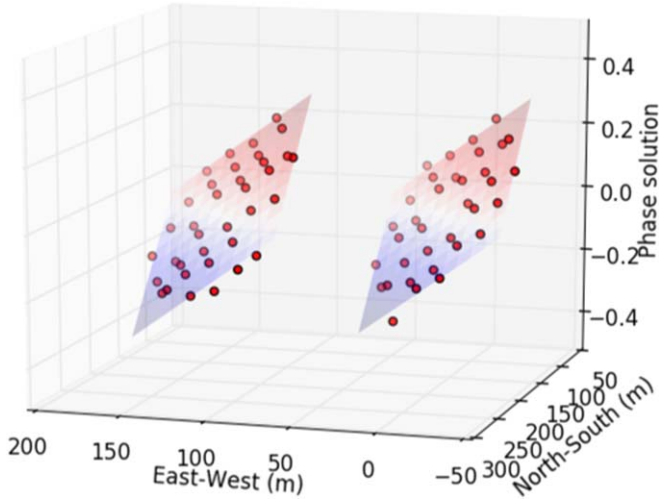
Figure 11. Example of plane fitting in FOcal. Red circles represent the OMNICAL phase solutions (or equivalently, its difference from 0) at a single frequency and single polarization vs. ideal tile positions. The x- and y-axes represent east–west positions and north–south positions, respectively, and z-axis represents the phase (in radians). The two planes are fitted results. These two planes have the same $\boldsymbol{\Phi}$ but different phase offsets.

difference between OMNICAL and FHD is normally large. We have to unwrap the phase in two dimensions before plane fitting, which is challenging. Instead of directly fitting a plane, we choose to calculate a rough value of these phase parameters and remove them so that we have a close answer to our target, and then we apply our method in Appendix A.2.

The first step is to remove the ψ terms for both hexagons by setting tile 1001 as the reference tile for the north hexagon and setting tile 1072 as the reference tile for the south hexagon. For each observation, the input phase solution of the reference tile is shifted to the target phase solution of that tile in the reference observation, and simultaneously the phases of all other tiles in the corresponding hexagon subarray are shifted by the same amount. In addition, the reference tile functions as the origin for each hexagon subarray, i.e., any tile position vector in that subarray $\boldsymbol{r}_i$ originates from the reference tile, i.e., $\boldsymbol{r}_i = \boldsymbol{r}_i' - \boldsymbol{r}_{\mathrm{reference}}'$, and thus ψ_N and ψ_S vanish at this point. The phase degeneracy reduces to $\boldsymbol{\Phi} \cdot \boldsymbol{r}_i$.

By removing the phase offset terms for both hexagons, we are only left with two degeneracy parameters in $\boldsymbol{\Phi}$. To illustrate the fitting for $\boldsymbol{\Phi}$, we select two basis vectors to describe the tile positions:

$$\boldsymbol{a}_1 = 14\hat{x}$$
$$\boldsymbol{a}_2 = -7\hat{x} - 7\sqrt{3}\,\hat{y}, \tag{18}$$

where $\hat{x}$ represents a vector pointing east with a length of 1 m and $\hat{y}$ represents a vector pointing north with a length of 1 m. Any tile location can be represented as

$$\boldsymbol{r}_i = n_{1i}\boldsymbol{a}_1 + n_{2i}\boldsymbol{a}_2, \tag{19}$$

where n_{1i} and n_{2i} are integers (see Figure 2 for tile positions). Note here that $\boldsymbol{r}_i$ originates from tile 1001 for the north hexagon and from tile 1072 for the south hexagon. To illustrate the

phase slope $\boldsymbol{\Phi}$, we use the basis shown in Equation (20):

$$\boldsymbol{b}_1 = \frac{\hat{z} \times \boldsymbol{a}_2}{\hat{z} \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_1)}$$
$$\boldsymbol{b}_2 = \frac{\boldsymbol{a}_1 \times \hat{z}}{\hat{z} \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_1)}, \tag{20}$$

where $\hat{z}$ is a unit vector perpendicular to the plane of the array. Thus, we have

$$\boldsymbol{a}_i \cdot \boldsymbol{b}_j = \delta_{ij}. \tag{21}$$

$\boldsymbol{\Phi}$ is represented as

$$\boldsymbol{\Phi} = \alpha_1 \boldsymbol{b}_1 + \alpha_2 \boldsymbol{b}_2. \tag{22}$$

The two components of $\boldsymbol{\Phi}$ are parameterized as α_1 and α_2. There is no cross term in the dot product of $\boldsymbol{\Phi}$ and $\boldsymbol{r}_i$, as Equation (23) shows:

$$\boldsymbol{\Phi} \cdot \boldsymbol{r}_i = \alpha_1 n_{1i} + \alpha_2 n_{2i}. \tag{23}$$

As we know, unwrapping the phase in two dimensions is difficult, but unwrapping the phase in one dimension is easier to do. The reason we choose Equation (18) as position basis and Equation (20) as phase slope basis is that we want to pick two nonparallel directions, which are $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in tile position space, and do 1D phase unwrapping and fitting in these two directions separately. The corresponding phase slopes in $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are α_1 and α_2, respectively. Although there is also degeneracy for α_i ($\alpha_i + 2\pi N$, where N is an integer, is also a solution), by evaluating Equation (23), any 2π wrap should vanish because the wrapping term gets multiplied by an integer; thus, the final result is unique.

After a rough estimation of degeneracy parameters is solved in this fashion, we apply them to the input solution. At this point the input solutions and target solutions are close. To get a finer solution, we further do a plane fitting as in Appendix A.2.

Appendix B
OMNICAL Convergence

The OMNICAL package has shown good computational efficiency in redundant calibration. However, we have discovered a convergence issue of lincal in OMNICAL. In our work, we have done tests on convergence by using different starting points for calibration. Ideally, the solution to the least-squares problem in lincal should converge to the same answer regardless of what starting points we give it. However, OMNICAL only converges to a level of 0.1% in our data set. This level of uncertainty is above the EOR signal.

In this work, we have solved this issue and have solutions converged to machine precision. All our results presented in this paper do not have this problem.

As an example of different starting points for OMNICAL, we can use different baseline groups in firstcal, which yields the same phase slopes but different phase offset results. This in turn leads to different results from logcal; thus, we have different starting points for lincal. Not only can we use firstcal to get an initial guess for the phase solutions, but also we can implement a rough calibration method introduced by Zheng et al. (2016), which we call roughcal. The relation between the phase of true visibilities, data, and gains is given by Equation (12). If we know the phases of true visibilities for

baseline types (1001, 1005) ($\theta_{1001,1005}$) and (1001, 1006) ($\theta_{1001,1006}$) (see Figure 2 for information of baseline types) and ϕ_{1001}, we are able to solve for ϕ_{1005} and ϕ_{1006}. With ϕ_{1005} and

$2N_{\text{redundant\_baselines}} \times (2N_{\text{unique\_baselines}} + 2N_{\text{tiles}})$-dimensional matrix:

$$
\underbrace{\begin{pmatrix} \Re(v_{ij} - g_i^0 g_j^{0*} y_{ij}^0) \\ \Im(v_{ij} - g_i^0 g_j^{0*} y_{ij}^0) \\ \vdots \end{pmatrix}}_{\equiv d} = \underbrace{\begin{pmatrix} \Re(g_j^{0*} y_{ij}^0) & -\Im(g_j^{0*} y_{ij}^0) & \Re(g_i^0 y_{ij}^0) & \Im(g_i^0 y_{ij}^0) & \cdots & \Re(g_i^0 g_j^{0*}) & -\Im(g_i^0 g_j^{0*}) & \cdots \\ \Im(g_j^{0*} y_{ij}^0) & \Re(g_j^{0*} y_{ij}^0) & \Im(g_i^0 y_{ij}^0) & -\Re(g_i^0 y_{ij}^0) & \cdots & \Im(g_i^0 g_j^{0*}) & \Re(g_i^0 g_j^{0*}) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \cdots \end{pmatrix}}_{\equiv A} \underbrace{\begin{pmatrix} \Re(\Delta g_i) \\ \Im(\Delta g_i) \\ \Re(\Delta g_j) \\ \Im(\Delta g_j) \\ \vdots \\ \Re(\Delta y_{ij}) \\ \Im(\Delta y_{ij}) \\ \vdots \end{pmatrix}}_{\equiv x}. \tag{25}
$$

ϕ_{1006} solved, we can move forward to solve for ϕ_{1002}, ϕ_{1010}, ϕ_{1011}, ϕ_{1012}, and so on and so forth. This guarantees that we will cover all tiles across this subarray. With this incomplete information, i.e., only the data from two types of baselines, we are able to get a rough guess for phase solutions per tile. The starting point for this process is the knowledge of $\theta_{1001,1005}$, $\theta_{1001,1006}$, and ϕ_{1001}. We are actually free to choose these three parameters because of the phase degeneracy. The most straightforward choice is $\theta_{1001,1005} = \gamma_{1001,1005}$, $\theta_{1001,1006} = \gamma_{1001,1006}$, and $\phi_{1001} = 0$, where γ_{ij} is as defined in Equation (12).

Either using different baseline subsets for firstcal or using roughcal instead of firstcal has the same effect: different starting points for OMNICAL. After calibration using any two sets of different starting points, we use the DP approach to force these two sets of solutions to have the same degeneracy parameters. By comparing these two sets of results, we see a level of 0.1% of difference in real data calibration. This level of calibration uncertainty is significant because a fraction of 10^{-3} of foregrounds is still brighter than the EOR signal.

In the algorithmic implementation of OMNICAL available at https://github.com/jeffzhen/omnical, the lincal iteration solves the least-squares problem by taking partial derivatives of χ^2 given by Equation (24) with respect to each individual parameter, forcing each partial derivative to be 0 to solve for the corresponding parameter, updating the solutions by a weighted average between the solutions from previous iteration and the new solutions

$$
\chi^2 = \sum_{ij} |v_{ij} - g_i g_j^* y_{ij}|^2. \tag{24}
$$

This algorithm is equivalent to iterating along the parameter axes, which is not as robust as approaching the local minimum along the real gradient in the parameter space.

To solve for the local minimum of χ^2, we add an extra step to obtain a finer convergence. We write Equation (8) into a matrix form as Equation (25) (Liu et al. 2010), where d is a $2N_{\text{redundant\_baselines}}$-dimensional vector, x is a $2N_{\text{unique\_baselines}} + 2N_{\text{tiles}}$-dimensional vector, and A is a

The solution to x is given by Equation (26) (Liu et al. 2010):

$$
x = (A^T A)^+ A^T d, \tag{26}
$$

where the superscript "+" denotes Moore–Penrose pseudo-inverse. This step is more computationally expensive not only because the matrix A is large but also because it needs to be updated and to find the pseudo-inverse per iteration for each time and frequency sample. In our calibration, we run OMNICAL first, followed by this new step to get solutions to converge. The difference in solutions from different starting points is reduced to machine precision. Although we have made this approach work within a reasonable amount of time for MWA Phase II data, this method is still not time efficient enough. For future experiments with a redundant array such as HERA (DeBoer et al. 2017), this matrix will be much larger than that of MWA Phase II; thus, higher efficiency in redundant calibration is desired. We will further explore a better approach in future work.

ORCID iDs

J. C. Pober https://orcid.org/0000-0002-3492-0433
B. J. Hazelton https://orcid.org/0000-0001-7532-645X
M. F. Morales https://orcid.org/0000-0001-7694-4030
J. S. Dillon https://orcid.org/0000-0003-3336-9958
A. P. Beardsley https://orcid.org/0000-0001-9428-8233
J. D. Bowman https://orcid.org/0000-0002-8475-2036
A. Ewall-Wice https://orcid.org/0000-0002-0086-7363
D. C. Jacobs https://orcid.org/0000-0002-0917-2269
M. Johnston-Hollitt https://orcid.org/0000-0003-2756-8301
C. Jordan https://orcid.org/0000-0002-1220-2940
D. L. Kaplan https://orcid.org/0000-0001-6295-2881
H. Kim https://orcid.org/0000-0001-8263-1006
P. Kittiwisit https://orcid.org/0000-0003-0953-313X
D. A. Mitchell https://orcid.org/0000-0002-1828-1969
S. Murray https://orcid.org/0000-0003-3059-3823
A. Neben https://orcid.org/0000-0001-7776-7240
A. R. Offringa https://orcid.org/0000-0002-5809-2783
S. Paul https://orcid.org/0000-0002-8671-2177
N. Udaya Shankar https://orcid.org/0000-0001-8974-3369

M. Tegmark ⦿ https://orcid.org/0000-0001-7670-7190
N. Thyagarajan ⦿ https://orcid.org/0000-0003-1602-7868
S. J. Tingay ⦿ https://orcid.org/0000-0002-8195-7562
C. Trott ⦿ https://orcid.org/0000-0001-6324-1766
R. B. Wayth ⦿ https://orcid.org/0000-0002-6995-4131
R. L. Webster ⦿ https://orcid.org/0000-0002-5325-2709
S. Wyithe ⦿ https://orcid.org/0000-0001-7956-9758

References

Barry, N., Hazelton, B., Sullivan, I., Morales, M., & Pober, J. 2016, MNRAS, 461, 3135

Barry, N. A. 2018, PhD thesis, Univ. Washington

Beardsley, A., Hazelton, B., Morales, M. F., et al. 2012, MNRAS, 425, 1781

Beardsley, A., Hazelton, B., Sullivan, I., et al. 2016, ApJ, 833, 102

Bowman, J. D., Cairns, I., Kaplan, D. L., et al. 2013, PASA, 30, e031

Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2009, ApJ, 695, 183

Carroll, P., Line, J., Morales, M., et al. 2016, MNRAS, 461, 4151

Datta, A., Bowman, J., & Carilli, C. 2010, ApJ, 724, 526

DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017, PASP, 129, 045001

Dillon, J. S., Kohn, S. A., Parsons, A. R., et al. 2017, arXiv:1712.07212

Dillon, J. S., Neben, A. R., Hewitt, J. N., et al. 2015, PhRvD, 91, 123011

Dillon, J. S., & Parsons, A. R. 2016, ApJ, 826, 181

Ewall-Wice, A., Dillon, J. S., Hewitt, J., et al. 2016, MNRAS, 460, 4320

Ewall-Wice, A., Dillon, J. S., Liu, A., & Hewitt, J. 2017, MNRAS, 470, 1849

Furlanetto, S. R. 2016, Understanding the Epoch of Cosmic Reionization (Berlin: Springer), 247

Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, PhR, 433, 181

Hazelton, B. J., Jacobs, D. C., Pober, J. C., & Beardsley, A. P. 2017, JOSS, 2, 140

Hazelton, B. J., Morales, M. F., & Sullivan, I. S. 2013, ApJ, 770, 156

Hurley-Walker, N., Callingham, J. R., Hancock, P. J., et al. 2016, MNRAS, 464, 1146

Intema, H., Jagannathan, P., Mooley, K., & Frail, D. 2017, A&A, 598, A78

Jacobs, D. C., Hazelton, B., Trott, C., et al. 2016, ApJ, 825, 114

Liu, A., Parsons, A. R., & Trott, C. M. 2014, PhRvD, 90, 023018

Liu, A., Tegmark, M., Morrison, S., Lutomirski, A., & Zaldarriaga, M. 2010, MNRAS, 408, 1029

Mesinger, A., Furlanetto, S., & Cen, R. 2011, MNRAS, 411, 955

Morales, M. F., Hazelton, B., Sullivan, I., & Beardsley, A. 2012, ApJ, 752, 137

Morales, M. F., & Wyithe, J. S. B. 2010, ARA&A, 48, 127

Neben, A., Bradley, R., Hewitt, J. N., et al. 2015, RaSc, 50, 614

Neben, A. R., Bradley, R. F., Hewitt, J. N., et al. 2016, ApJ, 826, 199

Nikolic, B., Carilli, C. & HERA Collaboration 2017, arXiv:1709.05245

Noorishad, P., Wijnholds, S. J., van Ardenne, A., & Van Der Hulst, J. 2012, A&A, 545, A108

Offringa, A., Wayth, R., Hurley-Walker, N., et al. 2015, PASA, 32, E008

Parsons, A., Pober, J., McQuinn, M., Jacobs, D., & Aguirre, J. 2012a, ApJ, 753, 81

Parsons, A. R., & Backer, D. C. 2009, AJ, 138, 219

Parsons, A. R., Pober, J. C., Aguirre, J. E., et al. 2012b, ApJ, 756, 165

Patil, A. H., Yatawatta, S., Zaroubi, S., et al. 2016, MNRAS, 463, 4317

Pober, J. C., Parsons, A. R., Aguirre, J. E., et al. 2013, ApJL, 768, L36

Procopio, P., Wayth, R., Line, J., et al. 2017, PASA, 34, E033

Salvini, S., & Wijnholds, S. J. 2014, A&A, 571, A97

Sullivan, I., Morales, M. F., Hazelton, B., et al. 2012, ApJ, 759, 17

Thyagarajan, N., Jacobs, D. C., Bowman, J. D., et al. 2015, ApJ, 804, 14

Thyagarajan, N., Shankar, N. U., Subrahmanyan, R., et al. 2013, ApJ, 776, 6

Tingay, S., Goeke, R., Bowman, J. D., et al. 2013, PASA, 30, E007

Trott, C. M., Pindor, B., Procopio, P., et al. 2016, ApJ, 818, 139

Trott, C. M., & Wayth, R. B. 2016, PASA, 33, E019

Trott, C. M., Wayth, R. B., & Tingay, S. J. 2012, ApJ, 757, 101

Vedantham, H., Shankar, N. U., & Subrahmanyan, R. 2012, ApJ, 745, 176

Wieringa, M. H. 1992, ExA, 2, 203

Zheng, H., Tegmark, M., Buza, V., et al. 2014, MNRAS, 445, 1084

Zheng, H., Tegmark, M., Dillon, J., et al. 2016, MNRAS, 465, 2901