The Astrophysical Journal, 809:61 (21pp), 2015 August 10 doi:10.1088/0004-637X/809/1/61
© 2015. The American Astronomical Society. All rights reserved.

PAPER-64 CONSTRAINTS ON REIONIZATION: THE 21 cm POWER SPECTRUM AT $z = 8.4$

Zaki S. Ali[1], Aaron R. Parsons[1,2], Haoxuan Zheng[3], Jonathan C. Pober[4], Adrian Liu[1,5], James E. Aguirre[6], Richard F. Bradley[7,8,9], Gianni Bernardi[10,11,12], Chris L. Carilli[13,14], Carina Cheng[1], David R. DeBoer[2], Matthew R. Dexter[2], Jasper Grobbelaar[10], Jasper Horrell[10], Daniel C. Jacobs[15], Pat Klima[8], David H. E. MacMahon[2], Matthys Maree[10], David F. Moore[6], Nima Razavi[14], Irina I. Stefan[14], William P. Walbrugh[10], and Andre Walker[10]
[1] Astronomy Dept., University of California, Berkeley, CA, USA
[2] Radio Astronomy Lab., University of California, Berkeley, CA, USA
[3] Dept. of Physics, Massachusetts Inst. of Tech., Cambridge, MA, USA
[4] Physics Dept. University of Washington, Seattle, WA, USA
[5] Berkeley Center for Cosmological Physics, Berkeley, CA, USA
[6] Dept. of Physics and Astronomy, University of Penn., Philadelphia, PA, USA
[7] Dept. of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA, USA
[8] National Radio Astronomy Obs., Charlottesville, VA, USA
[9] Dept. of Astronomy, University of Virginia, Charlottesville, VA, USA
[10] Square Kilometer Array, S. Africa, Cape Town, South Africa
[11] Dept. of Physics and Electronics, Rhodes University, South Africa
[12] Harvard-Smithsonian Cen. for Astrophysics, Cambridge, MA, USA
[13] National Radio Astronomy Obs., Socorro, NM, USA
[14] Cavendish Lab., Cambridge, UK
[15] School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA
Received 2015 February 23; accepted 2015 July 6; published 2015 August 11

ABSTRACT

In this paper, we report new limits on 21 cm emission from cosmic reionization based on a 135 day observing campaign with a 64-element deployment of the Donald C. Backer Precision Array for Probing the Epoch of Reionization in South Africa. This work extends the work presented in Parsons et al. with more collecting area, a longer observing period, improved redundancy-based calibration, improved fringe-rate filtering, and updated power-spectral analysis using optimal quadratic estimators. The result is a new 2σ upper limit on $\Delta^2(k)$ of $(22.4\,\mathrm{mK})^2$ in the range $0.15 < k < 0.5h\ \mathrm{Mpc}^{-1}$ at $z = 8.4$. This represents a three-fold improvement over the previous best upper limit. As we discuss in more depth in a forthcoming paper, this upper limit supports and extends previous evidence against extremely cold reionization scenarios. We conclude with a discussion of implications for future 21 cm reionization experiments, including the newly funded Hydrogen Epoch of Reionization Array.

Key words: cosmology: observations – dark ages, reionization, first stars – early universe – instrumentation: interferometers – intergalactic medium

1. INTRODUCTION

The *cosmic dawn* of the universe, which begins with the birth of the first stars and ends approximately one billion years later with the full reionization of the intergalactic medium (IGM), represents one of the last unexplored phases in cosmic history. Studying the formation of the first galaxies and their influence on the primordial IGM during this period is among the highest priorities in modern astronomy. During our cosmic dawn, IGM characteristics depend on the matter density field, the mass and clustering of the first galaxies (Lidz et al. 2008), their ultraviolet luminosities (McQuinn et al. 2007), the abundance of X-ray sources and other sources of heating (Pritchard & Loeb 2008; Mesinger et al. 2013), and higher-order cosmological effects like the relative velocities of baryons and dark matter (McQuinn & O'Leary 2012; Visbal et al. 2012).

Recent measurements have pinned down the bright end of the galaxy luminosity function at $z \lesssim 8$ (Bouwens et al. 2010; Schenker et al. 2013) and have detected a few sources at even greater distances (Ellis et al. 2013; Oesch et al. 2013). In parallel, a number of indirect techniques have constrained the evolution of the neutral fraction with redshift. Examples include integral constraints on reionization from the optical depth of Thomson scattering to the CMB (Planck Collaboration et al. 2014, 2015), large-scale CMB polarization anisotropies (Page et al. 2007), and secondary temperature fluctuations generated by the kinetic Sunyaev–Zel'dovich effect (Mesinger et al. 2012; Zahn et al. 2012; Battaglia et al. 2013; Park et al. 2013; George et al. 2014). Other probes of the tail end of reionization include observations of resonant scattering of Lyα by the neutral IGM toward distant quasars (the "Gunn–Peterson" effect; Fan et al. 2006), the demographics of Lyα emitting galaxies (Schenker et al. 2013; Treu et al. 2013; Faisst et al. 2014), and the Lyα absorption profile toward very distant quasars (Bolton et al. 2011; Bosman & Becker 2015). As it stands, the known population of galaxies falls well short of the requirements for reionizing the universe at redshifts compatible with CMB optical depth measurements (Robertson et al. 2013, 2015), driving us to deeper observations with, e.g., *James Webb Space Telescope* and ALMA, to reveal the fainter end of the luminosity function.

Complementing these probes of our cosmic dawn are experiments targeting the 21 cm "spin-flip" transition of neutral hydrogen at high redshifts. This signal has been recognized as a potentially powerful probe of the cosmic dawn (Furlanetto et al. 2006; Morales & Wyithe 2010; Pritchard & Loeb 2012) that can reveal large-scale fluctuations in the ionization state and temperature of the IGM, opening a unique window into the

complex astrophysical interplay between the first luminous structures and their surroundings. Cosmological redshifting maps each observed frequency with a particular emission time (or distance), enabling 21 cm experiments to eventually reconstruct three-dimensional pictures of the time-evolution of large scale structure in the universe. While such maps can potentially probe nearly the entire observable universe (Mao et al. 2008), in the near term, 21 cm cosmology experiments are focusing on statistical measures of the signal.

There are two complementary experimental approaches to accessing 21 cm emission from our cosmic dawn. So-called "global" experiments such as EDGES (Bowman & Rogers 2010), the LWA (Ellingson et al. 2013), LEDA (Greenhill & Bernardi 2012; Bernardi et al. 2015), DARE (Burns et al. 2012), SciHi (Voytek et al. 2014), BigHorns (Sokolowski et al. 2015), and SARAS (Patra et al. 2015) seek to measure the mean brightness temperature of 21 cm relative to the CMB background. These experiments typically rely on auto-correlations from a small number of dipole elements to access the sky-averaged 21 cm signal, although recent work is showing that interferometric cross-correlations may also be used to access the signal (Vedantham et al. 2015; Presley et al. 2015). In contrast, experiments targeting statistical power-spectral measurements of the 21 cm signal employ larger interferometers. Examples of such interferometers targeting the reionization signal include the GMRT (Paciga et al. 2013), LOFAR (van Haarlem et al. 2013), the MWA (Tingay et al. 2013), the 21CMA (Peterson et al. 2004; Wu 2009), and the Donald C. Backer Precision Array for Probe the Epoch of Reionization (PAPER; Parsons et al. 2010).

PAPER is unique for being a dedicated instrument with the flexibility to explore non-traditional experimental approaches, and is converging on a self-consistent approach to achieving both the level of foreground removal and the sensitivity that are required to detect the 21 cm reionization signal. This approach focuses on spectral smoothness as the primary discriminant between foreground emission and the 21 cm reionization signal and applies an understanding of interferometric responses in the delay domain to identify bounds on instrumental chromaticity (Parsons et al. 2012b, hereafter P12b). This type of "delay-spectrum" analysis permits data from each interferometric baseline to be analyzed separately without requiring synthesis imaging for foreground removal. As a result, PAPER has been able to adopt new antenna configurations that are densely packed and highly redundant. These configurations are poorly suited for synthesis imaging but deliver a substantial sensitivity boost for power-spectral measurements that are not yet limited by cosmic variance (Parsons et al. 2012a, hereafter P12a). Moreover, they are particularly suited for redundancy-based calibration (Wieringa 1992; Liu et al. 2010; Zheng et al. 2014), on which PAPER now relies to solve for the majority of the internal instrumental degrees of freedom (dof). The efficacy of this approach was demonstrated with data from a 32-antenna deployment of PAPER, which achieved an upper limit on the 21 cm power spectrum $\Delta^2(k) \leqslant (41\,\mathrm{mK})^2$ at $k = 0.27h\ \mathrm{Mpc}^{-1}$ (Parsons et al. 2014, hereafter P14). That upper limit improved over previous limits by orders of magnitude, showing that the early universe was heated from adiabatic cooling, presumably by emission from high-mass X-ray binaries or mini-quasars.

In this paper, we improve on this previous result using a larger 64-element deployment of PAPER and a longer observing period, along with better redundant calibration, an improved fringe-rate filtering technique, and an updated power-spectrum estimation pipeline. The result is an upper limit on $\Delta^2(k)$ of $(22.4\,\mathrm{mK})^2$ in the range $0.15 < k < 0.5h\ \mathrm{Mpc}^{-1}$ at $z = 8.4$. This result places constraints on the spin temperature of the IGM, and as is shown in a companion paper, Pober et al. (2015), this supports and extends previous evidence against extremely cold reionization scenarios. In Section 2 we describe the observations used in this analysis. In Sections 3 and 4, we discuss the calibration and the stability of the PAPER instrument. We then move on to a discussion of our power-spectrum analysis pipeline in Section 5. We present our results in Section 6 along with new constraints on the 21 cm power spectrum. We discuss these results in Section 7 and conclude in Section 8.

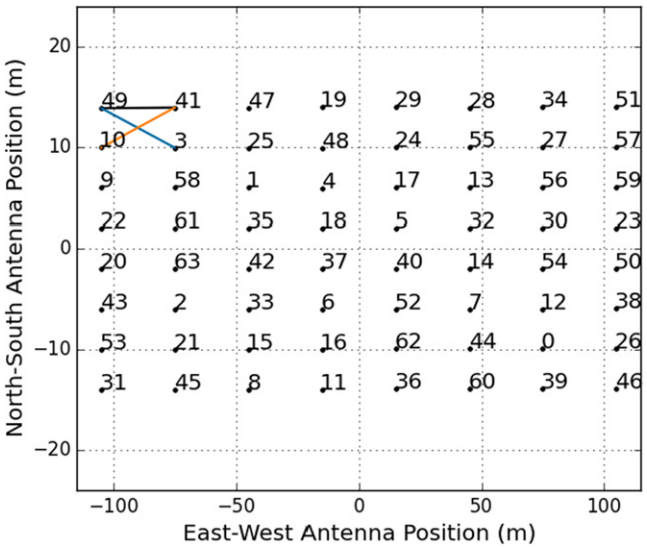

Figure 1. Antenna position within the PAPER-64 array. This analysis only makes use of east–west baselines between adjacent columns that have row separations of zero (black; e.g., 49–41, 41–47, 10–3,...) one in the northward direction (orange; e.g., 10–41, 3–47, 9–3,...) or one in the southward direction (blue; e.g., 49–3, 41–25, 10–58,...). Because of their high levels of redundancy, these baselines constitute the bulk of the array's sensitivity for power spectrum analysis.

2. OBSERVATIONS

We base our analysis on drift-scan observations with 64 dual-polarization PAPER antennas (hereafter, "PAPER-64") deployed at the Square Kilometre Array South Africa (SKA-SA) reserve in the Karoo desert in South Africa (30:43:17°5 S, 21:25:41°8 E). Each PAPER element features a crossed-dipole design measuring two linear (X, Y) polarizations. The design of the PAPER element, which features spectrally and spatially smooth responses down to the horizon with a FWHM of 60°, is summarized in Parsons et al. (2010) and Pober et al. (2012). For this analysis, we use only the XX and YY polarization cross-products.

As shown in Figure 1, PAPER-64 employs a highly redundant antenna layout where multiple baselines measure the same Fourier mode on the sky (P12a; P14). We rely on all 2016 baselines for calibration, but only use a subset of the baselines for the power spectrum analysis. This subset consists of three types of baselines: the 30 m strictly east–west baselines between adjacent columns (e.g., 49–41, black in Figure 1; hereafter referred to as *fiducial baselines*), 30 m east–west baselines whose eastern element is staggered one row up (e.g.,

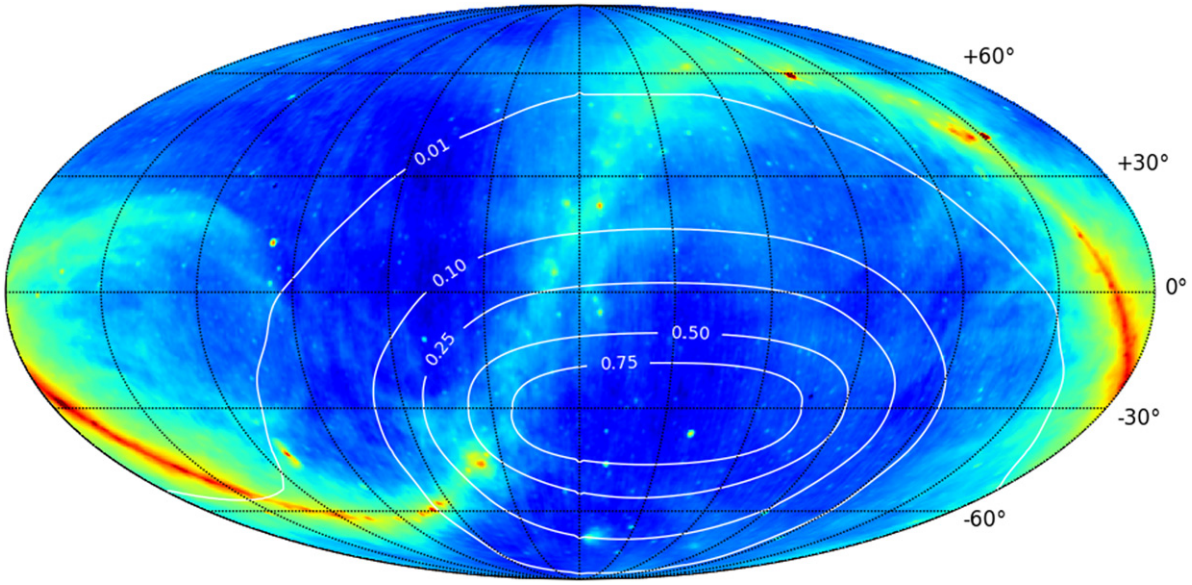

Figure 2. Global Sky Model (de Oliveira-Costa et al. 2008), illustrating foregrounds to the 21 cm cosmological signal, with contours indicating beam-weighted observing time (relative to peak) for the PAPER observations described in Section 2. The map is centered at 6:00 hr in R.A.

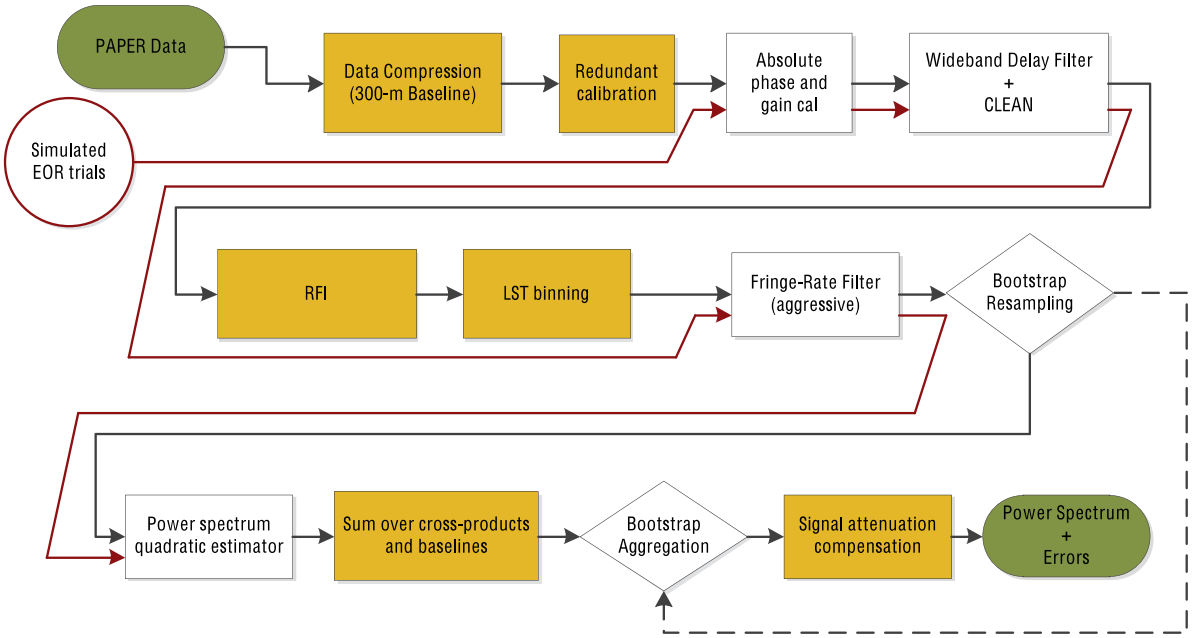

Figure 3. Stages of power-spectrum analysis. Black lines indicate data flow; red lines indicate Monte Carlo simulations used to measure signal loss. Yellow boxes indicate stages that by construction have negligible signal loss. Signal loss in other stages is tabluted in Table 1.

10–41, orange in Figure 1), and those whose eastern element is one row down (e.g., 49–3, blue in Figure 1). These baseline groups consist of 56, 49, and 49 baselines, respectively. We define a redundant group of baselines as being the set of baselines that have the same grid spacing; baselines in each of the three redundant groups described above are instantaneously redundant and therefore measure the same Fourier modes on the sky. Thus, within a redundant group, measurements from baselines may be coherently added to build power-spectrum sensitivity as N rather than $\sqrt{N}$, where N is the number of baselines added.

PAPER-64 conducted nighttime observations over a 135 day period from 2012 November 8 (JD 2456240) to 2013 March 23 (JD 2456375). Since solar time drifts with respect to local sidereal time (LST), this observing campaign yielded more samples of certain LSTs (and hence, sky positions) than others. For the power spectrum analysis, we use observations between 0:00 and 8:30 hr LST. This range corresponds to a "cold patch" of sky away from the galactic center where galactic synchrotron power is minimal, but also accounts for the weighting of coverage in LST. Figure 2 shows our observing field with the contours labeling the beam weighted observing time relative to the peak, directly over head the array.

The PAPER-64 correlator processes a 100–200 MHz bandwidth, first channelizing the band into 1024 channels of width 97.6 kHz, and then cross multiplying every antenna and

polarization with one another for a total of 8256 cross products, including auto correlations. Following the architecture in Parsons et al. (2008), this correlator is based on CASPER[16] open-source hardware and signal processing libraries (Parsons et al. 2006). Sixteen ROACH boards each hosting eight 8 bit analog-to-digital converters digitize and channelize antenna inputs. New to this correlator relative to previous PAPER correlators (Parsons et al. 2010), the cross multiplication engine is implemented on eight servers each receiving channelized data over two 10 Gb Ethernet links. Each server hosts two NVIDIA GeForce 580 GPUs running the open-source cross-correlation code developed by Clark et al. (2013). Visibilities are integrated for 10.7 s on the GPUs before being written to disk. All polarization cross-products are saved, although the work presented here only made use of the XX and YY polarization products.

3. CALIBRATION

Foreground contamination and signal sensitivity represent the two major concerns for 21 cm experiments targeting power spectrum measurements. Sources of foregrounds include galactic synchrotron radiation, supernova remnants, and extragalactic radio sources. In the low-frequency radio band (50–200 MHz) where 21 cm reionization experiments operate, emission from these foregrounds is brighter than the predicted reionization signal by several orders of magnitude (Santos et al. 2005; Ali et al. 2008; de Oliveira-Costa et al. 2008; Jelić et al. 2008; Bernardi et al. 2009, 2010; Ghosh et al. 2011). However, the brightest foregrounds are spectrally smooth, and this provides an important hook for their isolation and removal (Liu et al. 2009a; Petrovic & Oh 2011; Liu & Tegmark 2012). Unfortunately, interferometers, which are inherently chromatic instruments, interact with spectrally smooth foregrounds to produce unsmooth features that imitate line of sight Fourier modes over cosmological volumes (P12b; Morales et al. 2006; Bowman et al. 2009a). One approach to solving this problem involves an ambitious calibration and modeling approach to spatially localize and remove foreground contaminants (Liu et al. 2009b; Bowman et al. 2009b; Harker et al. 2009; Sullivan et al. 2012; Chapman et al. 2013). Perhaps the most impressive example of this approach is being undertaken by LOFAR, where dynamic ranges of 4.7 orders of magnitude have been achieved in synthesis images (Yatawatta et al. 2013), although it is expected that additional suppression of smooth-spectrum foreground emission will be necessary (Chapman et al. 2013).

The analysis for this paper employs a contrasting approach based on the fact that the chromaticity of an interferometer is fundamentally related to the length of an interferometric baseline. This relationship, known colloquially as "the wedge," was derived analytically (P12b; Vedantham et al. 2012; Thyagarajan et al. 2013; Liu et al. 2014a, 2014b), and has been confirmed in simulations (Datta et al. 2010; Hazelton et al. 2013) and observationally (Pober et al. 2013; Dillon et al. 2014). As described in P12b, the wedge is the result of the delay between when a wavefront originating from foreground emission arrives at the two antennas in a baseline. The fact that this delay is bounded by the light-crossing time between two antennas (which we call the "horizon limit" since such a wavefront would have to originate from the horizon) places a fundamental bound on the chromaticity of an interferometric baseline. So far, PAPER has had the most success in exploiting this bound (P14; Jacobs et al. 2015). In this analysis, we continue to use the properties of the wedge in order to isolate and remove smooth spectrum foregrounds.

As illustrated in Figure 2, our analysis pipeline begins by running a compression algorithm to reduce the volume of our raw data by a factor of 70. As described in Appendix A of P14, this is achieved by first performing statistical flagging to remove radio frequency interference (RFI) at the 6σ level, applying low-pass delay and fringe-rate filters that limit signal variation to delay scales of $|\tau| \lesssim 1$ μs and fringe-rate scales of $f \lesssim 23$ mHz, and then decimating to critical Nyquist sampling rates of 493 kHz along the frequency axis and 42.9 s along the time axis. We remind the reader that while information is lost in this compression, these sampling scales preserve emission between $-0.5 \leqslant k_{\parallel} \leqslant 0.5h$ Mpc^{-1} that rotates with the sky, making this an essentially lossless compression for measurements of the 21 cm reionization signal in these ranges.

After compression, we calibrate in two stages, as described in more detail below. The first stage (Section 3.1) uses instantaneous redundancy to solve for the majority of the per-antenna internal dof in the array. In the second stage (Section 3.2), standard self-calibration is used to solve for a smaller number of absolute phase and gain parameters that cannot be solved by redundancy alone. After suppressing foregrounds with a wide-band delay filter (Section 3.3) and additional RFI flagging and crosstalk removal, we average the data in LST (Section 3.4) and apply a fringe-rate filter (Section 3.5) to combine time-domain data. Finally, we use an optimal quadratic estimator (OQE) (Section 5) to make our estimate of the 21 cm power spectrum.

3.1. Relative Calibration

Redundant calibration has gained attention recently as a particularly powerful way to solve for internal dof in radio interferometric measurements without simultaneously having to solve for the distribution of sky brightness (Wieringa 1992; Liu et al. 2010; Noorishad et al. 2012; Marthi & Chengalur 2014; Zheng et al. 2014; P14). The grid-based configuration of PAPER antennas allows a large number of antenna calibration parameters to be solved for on the basis of redundancy (P12a; P14; Zheng et al. 2014). Multiple baselines of the same length and orientation measure the same sky signal. Differences between redundant baselines result from differences in the signal chain, including amplitude and phase effects attributable to antennas, cables, and receivers. Redundant calibration only constrains the relative complex gains between antennas and is independent of the sky. Since redundant calibration preserves signals common to all redundant baselines, this type of calibration does not result in signal loss.

In practice, redundant calibration often takes on two flavors: log calibration (LOGCAL) and linear calibration (LINCAL) (Liu et al. 2010; Zheng et al. 2014). LOGCAL uses logarithms applied to visibilities,

$$v_{ij} = g_i^* g_j y_{i-j} + n_{ij}^{\rm res}, \quad (1)$$

where g denotes the complex gain of antennas indexed by i and j, and y represents the "true" visibility measured by the baseline, to give a linearized system of equations

$$\log v_{ij} = \log g_i^* + \log g_j + \log y_{i-j}. \quad (2)$$

[16] http://casper.berkeley.edu

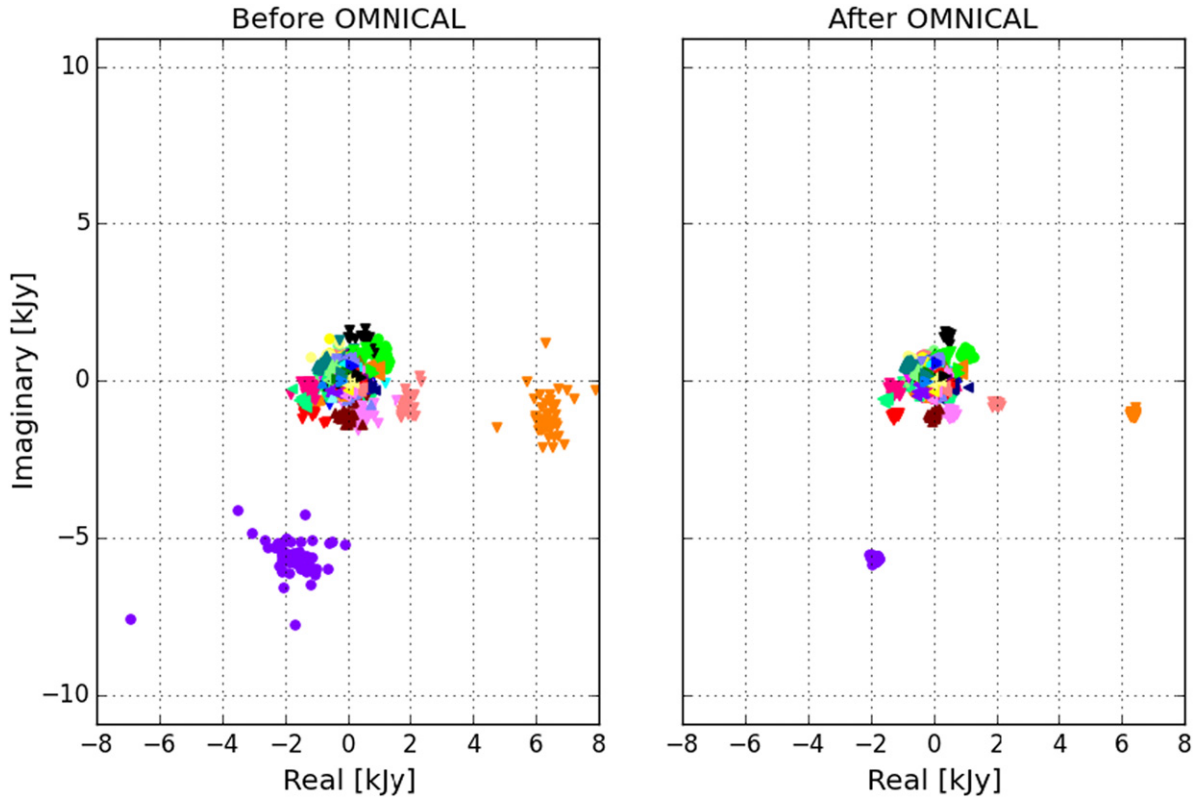

Figure 4. PAPER visibilities plotted in the complex plane before (left) and after (right) the application of the improved redundancy-based calibration with OMNICAL (Zheng et al. 2014). All baselines in the array measured at 159 MHz for a single time integration are plotted. Instantaneously redundant baselines are assigned the same symbol/color. The tighter clustering of redundant measurements with OMNICAL indicates improved calibration.

In solving for per-antenna gain parameters with a number of measurements that scales quadratically with antenna number, redundancy gives an over-constrained system of equations that can be solved using traditional linear algebra techniques. While LOGCAL is useful for arriving at a coarse solution from initial estimates that are far from the true value, LOGCAL has the shortcoming of being a biased by the asymmetric behavior of additive noise in the logarithm (Liu et al. 2010).

LINCAL, on the other hand, uses a Taylor expansion of the visibility around initial estimates of the gains and visibilities,

$$v_{ij} = g_i^{0*} g_j^0 y_{i-j}^0 + g_i^{1*} g_j^0 y_{i-j}^0 + g_i^{0*} g_j^1 y_{i-j}^0 + g_i^{0*} g_j^0 y_{i-j}^1, \quad (3)$$

where 0 denotes initial estimates and 1 represents the perturbation to the original estimate and is the solutions we fit for. Using initial estimates taken from LOGCAL, LINCAL constructs an unbiased estimator.

Redundant calibration was performed using OMNICAL[17]—an open-source redundant calibration package that is relatively instrument agnostic (Zheng et al. 2014). This package implements both LOGCAL and LINCAL, solving for a complex gain solution per antenna, frequency, and integration. The solutions are then applied to visibilities and the results are shown in Figure 4.

In addition to solving for gain solutions, OMNICAL also characterizes the quality of the calibration parameters by calculating the χ^2 for every integration. As defined in Zheng et al. (2014),

$$\chi^2 = \sum_{ij} \frac{|v_{ij} - y_{i-j} g_i^* g_j|^2}{\sigma_{ij}^2}, \quad (4)$$

where σ^2 is the noise in the visibilities. The χ^2 measures sum of the deviation of measured visibilities to that of the best fit model derived from the LINCAL relative to a noise model, and gives us a tool to use in order to check the quality of our data. The number of dof, as defined in Zheng et al. (2014), is given by

$$\begin{aligned} \text{dof} &= N_{\text{measurements}} - N_{\text{parameters}} \\ &= 2N_{\text{baselines}} - 2\left(N_{\text{antennas}} + N_{\text{unique baselines}}\right), \end{aligned} \quad (5)$$

and is effectively the number of visibilities for which χ^2 is calculated. If the data are noise-dominated, χ^2/dof is drawn from a χ^2 distribution with $\mu = 1$ and $\sigma^2 = 2$/dof. The calculated χ^2/dof for every frequency and integration of a fiducial day of observation in this season and for the fiducial power spectrum baselines is shown in Figure 5, demonstrating the stability of the PAPER instrument.

We measure a mean χ^2/dof of 1.9. This indicates that the redundant calibration solutions, while a substantial improvement over the previous PAPER-32 calibration (P14), do not quite result in residuals that are thermal noise dominated. Possible sources of this excess include instrumental crosstalk and poorly performing signal chains. While the latter will be down-weighted by the inverse of the estimated signal covariance described in Section 5, crosstalk is a defect in the

[17] https://github.com/jeffzhen/omnical

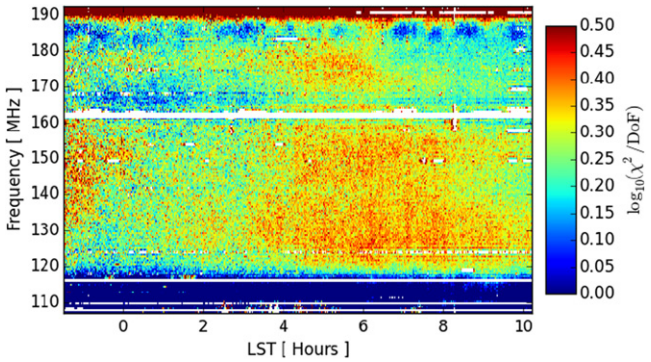

Figure 5. Log of χ^2 per degree of freedom of all baseline residuals after the application of OMNICAL. The plot comprises a observations over one day, with a frequency resolution of 493 kHz and a time resolution of 42.9 s.

data that must be addressed. Crosstalk caused by the cross-coupling of signals between antennas reveals itself as a static complex bias to a visibility that varies on timescales much longer than typical fringe rates. This effect skews the distribution of the χ^2 of the residuals away from 1. To minimize crosstalk, we first use OMNICAL to solve for antenna-dependent gains, and then average the residual deviations from redundancy over 10 minute windows before subtracting the average from the original visibilities. This crosstalk removal preserves signals common to redundant baseline groups (such as the 21 cm signal). Unfortunately, it also preserves a term that is the average of the crosstalk of all baselines in the redundant group. This residual crosstalk is removed by a fringe-rate filter later in the analysis.

3.2. Absolute Calibration

After solving for the relative complex gains of the antennas using redundant calibration, an overall phase and gain calibration remains unknown. We use the standard self calibration method for radio interferometers to solve for the absolute phase calibration. We used Pictor A, Fornax A, and the Crab Nebula to fit for the overall phase solutions. Figure 6 shows an image of the field with Pictor A (5:19:49.70, −45:46:45.0) and Fornax A (3:22:41.70, −37:12:30.0).

We then set our over all flux scale by using Pictor A as our calibrator source with source spectra derived in Jacobs et al. (2013),

$$S_\nu = S_{150} \times \left(\frac{\nu}{150\ \mathrm{MHz}}\right)^\alpha, \quad (6)$$

where $S_{150} = 381.88$ Jy $\pm$ 5.36 and $\alpha = -0.76 \pm 0.01$, with 1σ error bars.

To derive the source spectrum from our measurements, we use data that have been LST-averaged prior to the wide-band delay filter described in Section 3.3, for the hour before and after the transit of Pictor A. We image a 30° × 30° field of view for every frequency channel for each 10 minute snapshot and apply uniform weights to the gridded visibilities. We account for the required three-dimensional Fourier transform in wide field imaging by using the w-stacking algorithm implemented in WSclean (Offringa et al. 2014) although we note that the standard w-projection algorithm implemented in CASA[18] gives similar performances as the PAPER array is essentially

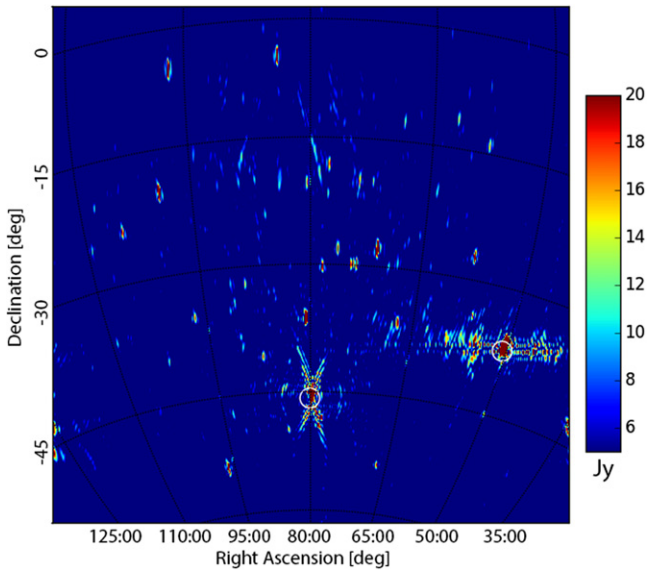

Figure 6. PAPER-64 image of a field including Pictor A and Fornax A, with white circles indicating catalog positions (Jacobs et al. 2011). Image was synthesized with two hours of visibilities while Pictor A was in transit and 53 MHz of instantaneous bandwidth from 120 to 173 MHz. Image quality is limited by the redundant configuration of the array (e.g., grating lobes as a result of periodic antenna spacing, elongated lobes arising from poor uv-coverage in the north–south direction). Nonetheless, this image demonstrates accurate phase calibration over a wide field of view.

instantaneously coplanar. A source spectrum is derived for each snapshot by fitting a two-dimensional Gaussian to Pictor A by using the PyBDSM[19] source extractor. Spectra are optimally averaged together by weighting them with the primary beam model evaluated in the direction of Pictor A. To fit our bandpass, we divide the model spectrum by the measured one and fit a 9th order polynomial over the 120–170 MHz frequency range. Figure 7 shows the calibrated Pictor A spectrum and the model spectrum from Jacobs et al. (2013). Also plotted are the 1σ error bars derived from the PyBDSM source extractor and averaged over the multiple snapshots used after being weighted by the beam-squared.

Fitting a polynomial to the bandpass has the potential for signal loss which would include suppressing modes that may contain the cosmological signal. In order to quantify the signal loss associated with fitting a ninth degree polynomial to the bandpass, we run a Monte Carlo simulation of the effect the bandpass has on a model 21 cm reionization signal. We construct a model baseline visibility as a Gaussian random signal multiplied by the derived bandpass for every independent mode measured. We calculate the total number of independent modes by counting the number of independent uv-modes sampled for the different baseline types over the two hour time interval used to measure the bandpass. We average each mode together and fit a 9th degree polynomial. Using this as our measured bandpass for this simulated signal, we finally compare the power spectrum from the output of the simulated signal to the input power spectrum as a function fo k-mode. We find that between $-0.06 < k < 0.06$, the width of our wideband delay filter described below, the signal loss is less than 3% and at the mode right outside the above limit is

[18] http://casa.nrao.edu

[19] http://www.lofar.org/wiki/doku.php?id=public:user_software:pybdsm

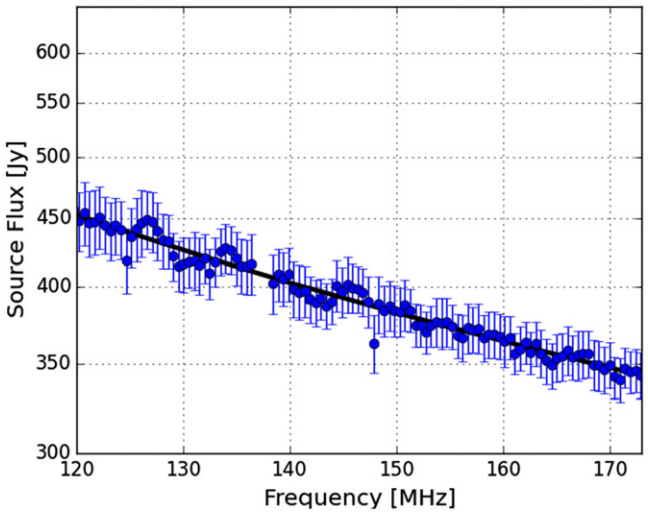

Figure 7. Measured spectrum of Pictor A in Stokes I (blue) relative to its catalog value (black; Jacobs et al. 2013). Flux density measurements are extracted from images of Pictor A, made independently for each frequency channel in 10 minute snapshots as Pictor transits between hour angles of −1:49 and 1:10. Each measurement is then divided by the PAPER beam model and averaged to obtain the measured spectrum, which serves to characterize the flux scale of the PAPER-64 observations. Error bars indicate 68% confidence intervals, derived from the Gaussian fits in the source extractor used to measure the flux density in PyBDSM, combined from all snapshots.

2×10^{-7}%. We apply the latter correction factor for all modes outside the width of the delay filter to the final power spectrum.

3.3. Wideband Delay Filtering

Before implementing our foreground removal techniques, we combine the two linear polarizations for an estimate of Stokes I as per Moore et al. (2013). Namely, Stokes I can be estimated as

$$V_{\rm I} = \frac{1}{2}\left(V_{\rm XX} + V_{\rm YY}\right), \tag{7}$$

where $V_{\rm XX}$ and $V_{\rm YY}$ are the visibilities of the two linear polarizations measured by the interferometer. There are some important caveats to the estimate of Stokes I provided by Equation (7). One important caveat is that it neglects the beam asymmetry between the two linear polarization states. This mismatch can cause polarization leakage from Stokes Q into Stokes I, thus contaminating our measurement of the power spectrum with any polarized emission from the sky. This effect for PAPER, as shown in Moore et al. (2013), leaks 4% of Q in to I in amplitude (2.2×10^{-3} in the respective power spectra). We take the conservative approach and do not correct for this effect, noting that the leakage of Q in to I will result in positive power, increasing our limits.

Foreground removal techniques discussed in the literature include spectral polynomial fitting (Wang et al. 2006; Bowman et al. 2009a; Liu et al. 2009a), principal component analysis (Liu & Tegmark 2011; Paciga et al. 2011; Masui et al. 2013; Paciga et al. 2013), non-parametric subtractions (Harker et al. 2009; Chapman et al. 2013), and inverse covariance weighting (Liu & Tegmark 2011; Dillon et al. 2013, 2014; Liu et al. 2014a, 2014b), Fourier-mode filtering Petrovic & Oh (2011), and per-baseline delay filtering described in P12b. This delay-spectrum filtering technique is well-suited to the maximum redundancy PAPER configuration, which is not optimized for the other approaches where high fidelity imaging is a prerequisite. The delay-spectrum foreground filtering method is described in detail by P14; its application is unchanged here. In summary; we Fourier transform each baseline spectrum into the delay domain

$$\begin{aligned}\tilde{V}_\tau &= \int W_\nu A_\nu I_\nu e^{-2\pi i \tau_g} \cdot e^{2\pi i \tau \nu} d\nu \\ &= \tilde{W}_\tau * \tilde{A}_\tau * \tilde{I}_\tau * \delta(\tau_g - \tau),\end{aligned} \tag{8}$$

where A_ν is the frequency dependent antenna response, W_ν is a sampling function that includes RFI flagging and a Blackman-Harris tapering function that minimizes delay-domain scattering from RFI flagging, and I_ν is the source spectrum. In the delay domain, a point source appears as a δ-function at delay τ_g, convolved by the Fourier transforms of the source spectrum, the antenna response, and the sampling function. We note that the antenna response effectively determines a finite bandpass, which imposes a lower bound of $1/B \approx 10$ ns on the width of any delay-domain convolving kernel. As per Parsons & Backer (2009) and P14, we deconvolve the kernel resulting from $W(\tau)$ using an iterative CLEAN-like procedure (Högbom 1974) restricting CLEAN components to fall within the horizon plus a 15 ns buffer that includes the bulk of the kernels convolving the δ-function in Equation (8). To remove the smooth spectrum foreground emission we subtract the CLEAN components from the original visibility.

Applying the delay filter to fiducial baselines used in the power spectrum analysis, foregrounds are suppressed by ∼4 orders of magnitude in power, or −40 dB of foreground suppression, as seen in Figure 8. As discussed in P14, there is a small amount of signal loss associated with this filter. For the baselines and filter parameters used, the loss was found to be 4.8% for the first mode outside of the horizon, 1.3% for the next mode out, and less than 0.0015% for the higher modes.

3.4. Binning in LST

After the wideband delay filter, we remove a second layer of RFI that was overshadowed by the foreground signal. RFI are excised with a filter which flags values 3σ above the median using a variance calculated in a localized time and frequency window.

We then average the entire season in LST with 43 s bin widths, matching the cadence of the compressed data. The full season was 135 days long; of these, 124 days were included in the average. We make two separate LST-binned data sets, averaging every other Julian day together to obtain an "even" and "odd" data set. The use of these two data sets allows us to construct an unbiased power spectrum estimate.

Sporadic RFI events result in measurements that, in any individual LST bin, deviate from the Gaussian distribution characteristic of thermal noise. To catch these events, we compute the median of a LST bin for each frequency and flag values 3σ above the median, before averaging. Since we are narrowing the distribution of visibilities about the median, the measured thermal noise variance is not preserved under this filter. However, since the central value is preserved, the expectation value of the measured visibility in each LST bin is unchanged, and there is no associated signal loss for power spectrum measurements. Moreover, because errors are

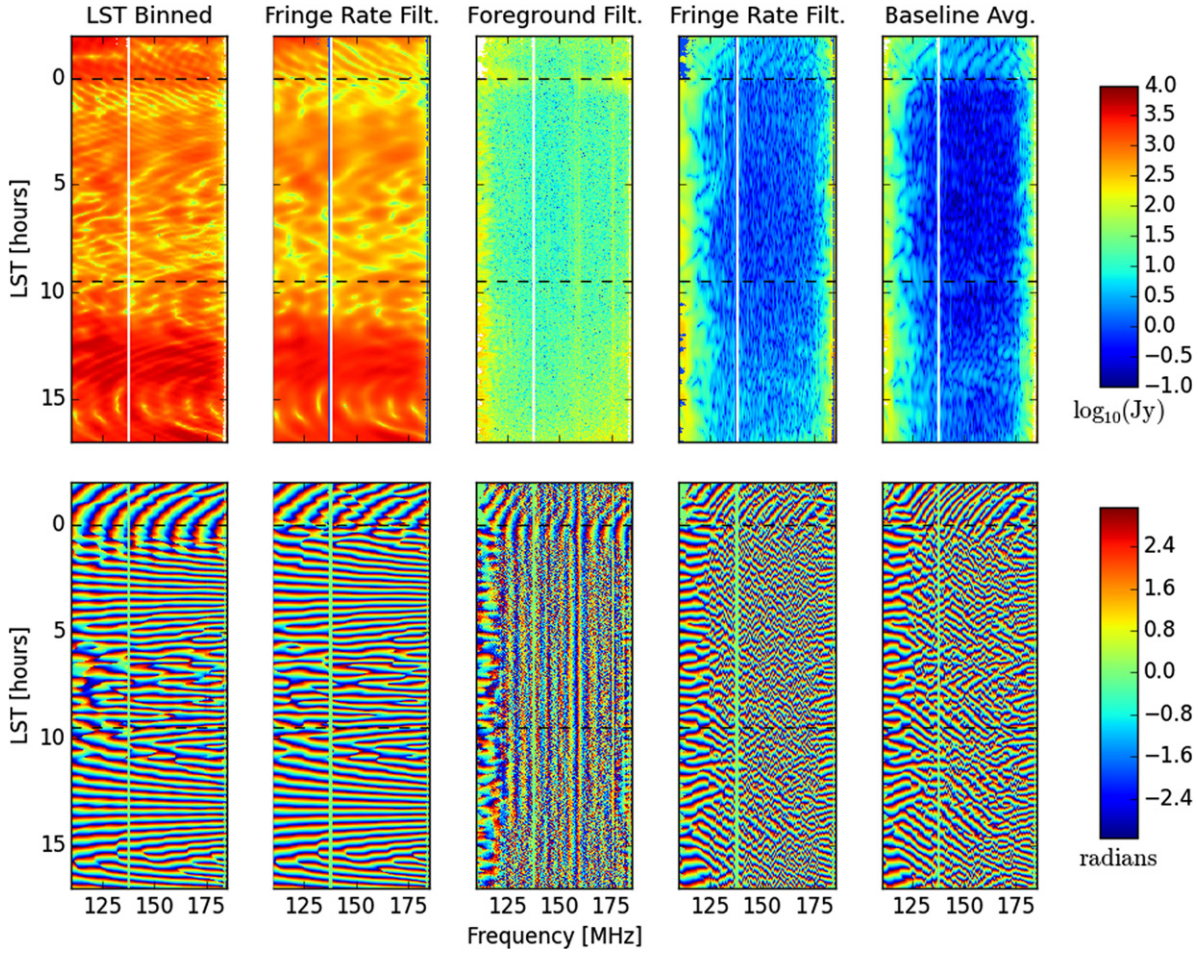

Figure 8. Visibilities measured by a fiducial baseline in the PAPER-64 array, averaged over 135 days of observation. From left to right, columns represent data that: (1) contain foregrounds prior to the application of a wideband delay filter or fringe-rate filtering, (2) are fringe-rate filtered but not delay filtered, (3) are delay filtered at 15 ns beyond the horizon limit but are not fringe-rate filtered, (4) are both delay and fringe-rate filtered, and (5) are delay and fringe-rate filtered and have been averaged over all redundant measurements of this visibility. The top row shows signal amplitude on a logarithmic scale; the bottom row illustrates signal phase. Dashed lines indicate the 0:00–8:30 range in LST used for power spectrum analysis. The putative crosstalk is evident in the center panel as constant phase features which do not fringe as the sky. The two right panels show some residual signal in the phase structure which is present at low delay. Away from the edges of the observing band, over four orders of magnitude of foreground suppression is evident.

estimated empirically through bootstrapping (see Section 5.4), the slight increase in measurement error associated with truncating the tails of the Gaussian distribution are naturally accounted for.

3.5. Fringe-rate Filtering

By averaging visibilities in time, we aim to maximize sensitivity by coherently combining repeated measurements of k-modes before squaring these measurements and averaging over independent k-modes to estimate the power spectrum amplitude. This is mathematically similar to the more traditional process of gridding in the uv plane, but applied to a single baseline. However, rather than applying a traditional box-car average, we can apply a kernel—a so-called "fringe-rate" filter—that weights different temporal rates by the antenna beam corresponding to the parts of the sky moving at the same rate.

For a given baseline and frequency, different parts of the sky exhibit different fringe-rates. Maximum fringe rates are found along the equatorial plane, where the rotation rate of the sky is highest, and zero fringe rates are found at the poles, where the sky does not rotate and hence sources do not move through the fringes of a baseline (Parsons & Backer 2009). Fringe rates are not constant as a function of latitude. Bins of constant fringe rate correspond to rings in R.A. and decl., where the east–west projection of a baseline projected toward a patch of the sky is constant. We use this fact in conjunction with the rms beam response for each contour of constant fringe rate to construct a time average kernel or "fringe-rate filter."

As examined in Parsons et al. (2015), it is possible to tailor fringe-rate filters to optimally combine time-ordered data for power-spectrum analysis. Fringe-rate filters can be chosen that up-weight points of the sky where our instrument is more sensitive and down-weight those points farther down in the primary beam, which are less sensitive. For white noise, all fringe-rate bins will contain the same amount of noise, but the amount of signal in each bin is determined by the primary beam response on the sky. By weighting fringe-rate bins by the rms of the beam response, we can get a net increase in sensitivity.

Applying this filter effectively weights the data by another factor of the beam area, changing the effective primary beam response,[20] $A(l, m)$ (Parsons et al. 2015). By utilizing prior

[20] The angular area in Equation (24) will reflect the new angular area corresponding to the change in beam area.

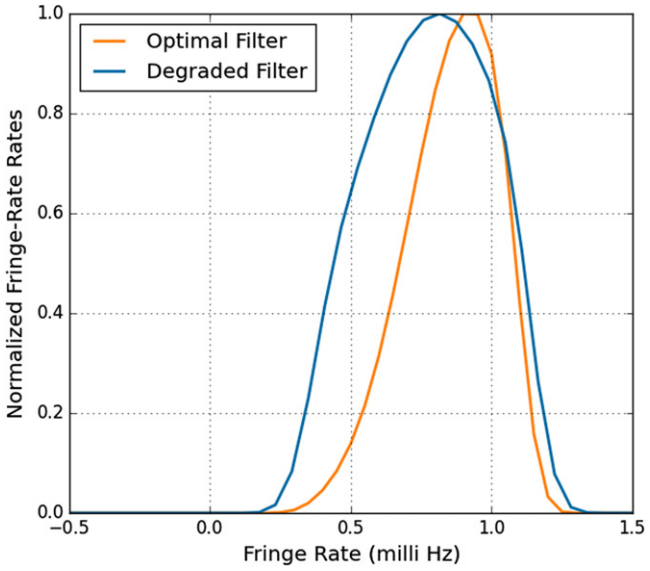

Figure 9. Optimal fringe-rate filter (orange) that and the degraded fringe-rate filter (blue) actually used in the analysis at 151 MHz, normalized to peak at unity.

knowledge about the beam area, we are selectively down-weighting areas on the sky contributing little signal. This will result in a net improvement in sensitivity depending on the shape of the beam and the decl. of the array. For PAPER, this filter roughly doubles the sensitivity of our measurements.

Generally, a fringe-rate filter integrates visibilities in time. For a fringe-rate filter, $f_{\rm fr}$, the effective integration time can be calculated by comparing the variance statistic before and after filtering:

$$t_{\rm int,after} = t_{\rm int,before} \frac{\int \sigma_f^2 df}{\int \sigma_f^2 f_{\rm fr}^2 df}, \quad (9)$$

where $t_{\rm int,before}$ is the integration time before filtering, σ_f denotes the noise variance in fringe rate space and the integral is taken over all possible fringe rates for a given baseline and frequency. As discussed in Parsons et al. (2015), the signal re-weighting associated with this fringe-rate filter can be interpreted as a modification to the shape of the primary beam.

For the fiducial baseline at 151 MHz, the integration time, as given in Equation (9), associated with an optimal fringe rate filter is 3430 s. The number of statistically independent samples on the sky decreases from 83 to 1 sample per hour. As discussed in Section 5.3, empirically estimating a covariance matrix with a small number of independent samples can lead to signal loss in the OQE. In order to counteract the signal loss, we degrade the optimal fringe-rate filter, as shown in Figure 9, to have an effective integration time of 1886 s, increasing the number of independent modes to 2 per hour. The fringe rate filter is now sub-optimal, but is still an improvement on the boxcar weighting as used in P14. As documented in Table 1, the correction factor for the associated signal loss of the filter we have chosen is 1.39.

We implement the modified filter on a per baseline basis by weighting the fringe-rate bins on the sky by the rms of the beam at that same location. In order to obtain a smooth filter in the fringe-rate domain, we fit a Gaussian with a hyperbolic tangent tail to this filter. In addition, we multiply this response with another hyperbolic tangent function that effectively zeros out fringe rates below 0.2 mHz. This removes the slowly varying signals that we model as crosstalk. We convolve the time-domain visibilities with the Fourier transform of the resulting fringe-rate filter, shown in Figure 9, to produce an averaged visibility. The effect on the data can be seen in Figure 8.

Table 1
Signal Loss vs. Analysis Stage

Analysis Stage	Typical Loss	Maximum Loss
Bandpass Calibration	$<2 \times 10^{-7}$%	3.0%
Delay Filtering	1.5×10^{-3}%	4.8%
Fringe-rate Filtering	28.1%	28.1%
Quadratic Estimator	<2.0%	89.0%
Median of Modes	30.7%	30.7%

4. INSTRUMENTAL PERFORMANCE

4.1. Instrument Stability

In order to build sensitivity to the 21 cm reionization signal, it is critical that PAPER be able to integrate coherently measurements made with different baselines on different days. Figure 10 shows the visibility repeatability between baselines and nights as a function of LST. Specifically, we histogram the real part of the visibilities for all redundant fiducial baselines in a given LST bin for foreground contained data. We see that for a given LST bin, the spread in values over all the baselines is ~50 Jy which corresponds with our observed $T_{\rm sys} \sim 500$ K. We get more samples per LST bin in the range of 2–10 hr due to our observing season, therefore the density of points in this LST region is greater, as shown by the color scale. This density plot shows that redundant baselines agree very well with one another; OMNICAL has leveled the antenna gains to within the noise.

Delving in a little deeper, we also examine the stability in time for measurements in a particular LST bin. In order to quantify the stability in time we extract one channel for a given baseline for every observation day and LST bin. We then Fourier transform along the time direction for every LST bin and compute the power spectrum. As shown in Figure 11, for timescales greater than one day, we see that signal variance drops by almost four orders of magnitude, with the exception of an excess on two-day timescales caused by the changing alignment of the 42.9 s integration timescale relative to a sidereal day. The implication of this measurement is that, after calibration, PAPER measurements are sufficiently stable to be integrated coherently over the entire length of a 135 day observation. This implies day-to-day stability of better than 1%, contributing negligibly to the uncertainties in the data.

4.2. System Temperature

During the LST binning step, the variance of the visibilities that are averaged together for a given frequency and LST bin are recorded. Using these variances, we calculate the system temperature as a function of LST, averaging over each LST hour.

$$T_{\rm rms} = T_{\rm sys} / \sqrt{2\Delta\nu t}, \quad (10)$$

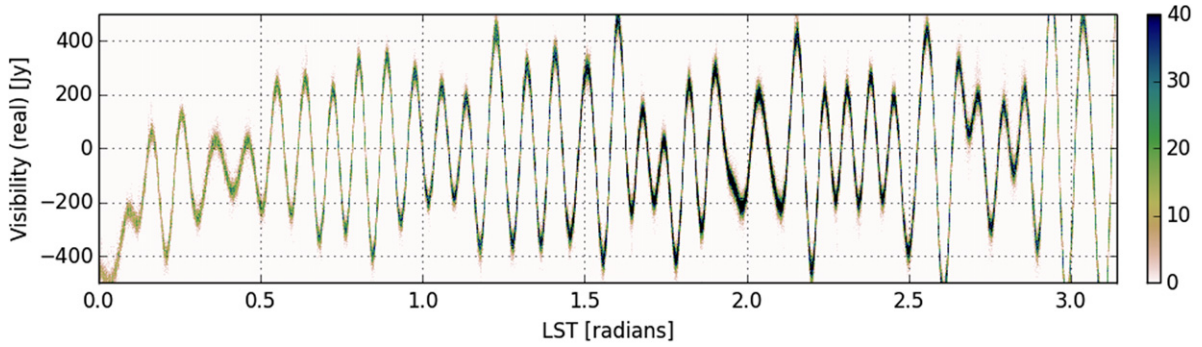

Figure 10. Histogram of the real component of all calibrated visibilities measured over 135 days with every redundant instance of the fiducial baseline at 150 MHz. Color scale indicates the number of samples falling in an LST/flux-density bin. This plot serves to illustrate the stability of the PAPER instrument and the precision of calibration. The temporal stability of a single LST bin over multiple days is shown in Figure 11.

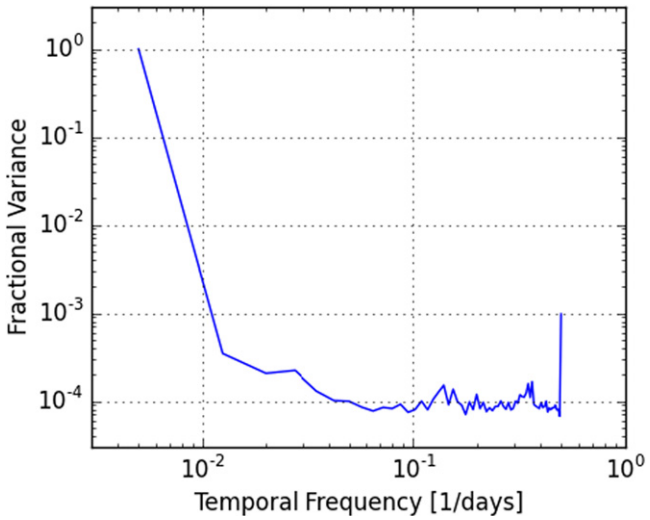

Figure 11. Power spectrum of 135 days of time-series data contributing to a single LST bin, illustrating the stability of measurements over the observing campaign. Relative to the average value, variation in the measured value across days (quantified by variance as a function of time period) is orders of magnitude lower. The excess at two-day timescales is a beat frequency associated with the changing alignment of integration windows in the correlator with respect to sidereal time.

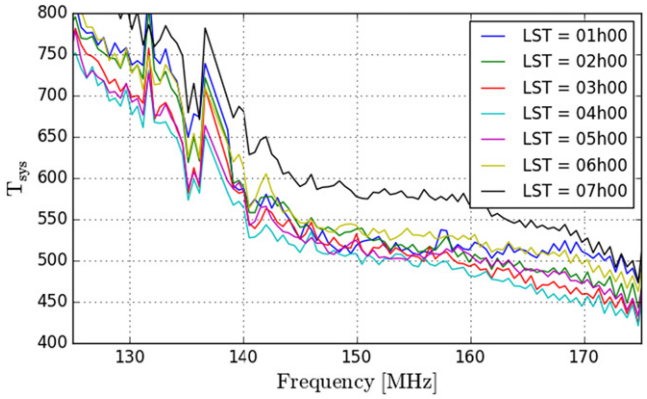

Figure 12. System temperature, inferred from the variance of samples falling in an LST bin, averaged over one-hour intervals in LST. The measured value in the 150–160 MHz range is consistent with previous determinations of system temperature (Jacobs et al. 2015; P14).

where $\Delta\nu$ is the bandwidth, t is the integration time, and $T_{\rm rms}$ is the rms temperature, or the variance statistic described above. Figure 12 shows the results of this calculation. In this observing season, the system temperature drops just below previous estimates as in P14 and Jacobs et al. (2015) of $T_{\rm sys} = 560$ K, at $T_{\rm sys} = 500$ K at 160 MHz. However, this estimate is more consistent with the results derived in (Moore et al. 2015), where $T_{\rm sys} = 505$ K at 164 MHz. The change in the system temperature can be attributed to the reduced range of LST used in the calculation. We note that at 7:00 LST, there is an increase in the system temperature due to the rising of the galactic plane as seen in Figure 3.

When calculating the system temperature using the variance in the visibilities for a given LST and frequency, we take into account the fact that we flag 3σ outliers from the median. To calculate an effective correction factor to account for the filtering, we assume the visibilities follow a Gaussian distribution which would require a correction factor of 1.34 for the removal of data points that are 3σ above the median. In other words, we are accounting for the wings of the Gaussian that would contribute to the variance in the visibility.

Previous estimates of the system temperature (P14; Jacobs et al. 2015) relied on differencing and averaging baselines, time samples, and/or frequency channels. The relative agreement between these various methods of estimating the system temperature provides a robust measure of the system temperature of the PAPER instrument. Agreement between the instantaneous measurements of the system temperature, the LST repetition variance, and the predicted power spectrum noise level (see below) indicates a robustly stable system with no significant long term instability contributing appreciable noise.

5. POWER SPECTRUM ANALYSIS

In this section we first review the OQE formalism, followed by a walk-through of our particular applications of the OQE method to our data. Finally, we discuss the effects of using an empirically estimated covariance matrix in our analysis.

5.1. Review of OQEs

We use the OQE method to estimate our power spectrum as done in Liu & Tegmark (2011), Dillon et al. (2013), Liu et al. (2014a), Liu et al. (2014b), and Trott et al. (2012). Here we briefly review the OQE formalism with an emphasis on our application to data, which draws strongly from the aforementioned works, but also relies on empirical techniques similar to

those used in P14. The end goal of this analysis is to estimate the 21 cm power spectrum, $P_{21}(\boldsymbol{k})$, defined such that

$$\langle \widetilde{T}_b(\boldsymbol{k})\widetilde{T}_b^*(\boldsymbol{k}')\rangle = (2\pi)^3\delta^{\rm D}(\boldsymbol{k}-\boldsymbol{k}')P_{21}(\boldsymbol{k}), \quad (11)$$

where $\widetilde{T}_b(\boldsymbol{k})$ is the spatial Fourier transform of the brightness temperature distribution on the sky, $\langle\rangle$ denotes an ensemble average, and $\delta^{\rm D}$ is the Dirac delta function.

In order to make an estimate of the power spectrum in the OQE formalism, one begins with a data vector $\boldsymbol{x}$. This vector could, for example, consist of a list of brightness temperatures on the sky for an imaging-based data analysis, or (in our case) a list of measured visibilities. We form the intermediate quantity,

$$\hat{q}_\alpha = \frac{1}{2}\boldsymbol{x}^\dagger \boldsymbol{C}^{-1}\boldsymbol{Q}_\alpha \boldsymbol{C}^{-1}\boldsymbol{x} - b_\alpha, \quad (12)$$

which will be needed to form the OQE of our power spectrum. Here, $\boldsymbol{C} \equiv \langle \boldsymbol{x}\boldsymbol{x}^\dagger\rangle$ is the true covariance matrix of the data vector $\boldsymbol{x}$, $\boldsymbol{Q}_\alpha$ is the operator that takes visibilities into power spectrum k-space and bins into the αth bin, and b_α is the bias to the estimate that needs to be subtracted off. In general, $\boldsymbol{Q}_\alpha$ represents a family of matrices, one for each k bin indexed by α. Each matrix is defined as $\boldsymbol{Q}_\alpha \equiv \frac{\partial \boldsymbol{C}}{\partial p_\alpha}$, i.e., the derivative of the covariance matrix with respect to the band power p_α. The bandpower p_α can be intuitively thought of as the value of the power spectrum in the αth k bin. Therefore, $\boldsymbol{Q}_\alpha$ encodes the response of the data covariance matrix to the αth bin of the power spectrum.

The bias term b_α in Equation (12) will include contributions from both instrumental noise and residual foregrounds. Their presence in the data is simply due to the fact that both contributions have positive *power*. One approach to dealing with these biases is to model them and to subtract them off, as is suggested by Equation (12). An alternate approach is to compute a cross-power spectrum between two data sets that are known to have the same sky signal but independent instrumental noise realizations. Labeling these two data sets as $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ and computing

$$\hat{q}_\alpha = \frac{1}{2}\boldsymbol{x}_1^\dagger \boldsymbol{C}^{-1}\boldsymbol{Q}_\alpha \boldsymbol{C}^{-1}\boldsymbol{x}_2, \quad (13)$$

one arrives at a cross-power spectrum that by construction has no noise bias. There is thus no need to explicitly model and subtract any noise bias, although any residual foreground bias will remain, since it is a contribution that is sourced by signals on the sky, and therefore must exist in all our data sets.

The set of $\hat{q}_\alpha$s do not yet constitute a properly normalized estimate of the power spectrum (as evidenced, for example, by the extra factors of $\boldsymbol{C}^{-1}$). To normalize our results, we group the unnormalized bandpowers into a vector $\hat{\boldsymbol{q}}$ and apply a matrix $\boldsymbol{M}$ (whose exact form we specify later), so that

$$\hat{\boldsymbol{p}} = \boldsymbol{M}\hat{\boldsymbol{q}} \quad (14)$$

is a normalized estimate $\hat{\boldsymbol{p}}$ of the true power spectrum $\boldsymbol{p}$. We emphasize that the vector space that contains $\hat{\boldsymbol{q}}$ and $\hat{\boldsymbol{p}}$ is an "output" vector space over different k-bins, which is separate from the "input" vector space of the measurements, in which $\boldsymbol{x}$ and $\boldsymbol{C}$ reside.

To select an $\boldsymbol{M}$ matrix that properly normalizes the power spectrum, we must compute the window function matrix $\boldsymbol{W}$ for our estimator. The window matrix is defined such that the true bandpowers $\boldsymbol{p}$ and our estimates $\hat{\boldsymbol{p}}$ of them are related by

$$\hat{\boldsymbol{p}} = \boldsymbol{W}\boldsymbol{p}, \quad (15)$$

so that each row gives the linear combination of the true power that is probed by our estimate. With a little algebra, one can show that

$$\boldsymbol{W} = \boldsymbol{M}\boldsymbol{F}, \quad (16)$$

where

$$\boldsymbol{F}_{\alpha\beta} = \frac{1}{2}\,{\rm tr}\left(\boldsymbol{C}^{-1}\boldsymbol{Q}_\alpha \boldsymbol{C}^{-1}\boldsymbol{Q}_\beta\right), \quad (17)$$

which we have suggestively denoted with the symbol $\boldsymbol{F}$ to highlight the fact that this turns out to be the Fisher information matrix of the bandpowers. In order to interpret each bandpower as the weighted average of the true bandpowers, we require each row of the window function matrix to sum to unity. As long as $\boldsymbol{M}$ is chosen in such a way that $\boldsymbol{W}$ satisfies this criterion, the resulting bandpower estimates $\hat{\boldsymbol{p}}$ will be properly normalized.

Beyond the normalization criterion, a data analyst has some freedom over the precise form of $\boldsymbol{M}$, which effectively also rebins the bandpower estimates. One popular choice is $\boldsymbol{M} = \boldsymbol{F}^{-1}$, which implies that $\boldsymbol{W} = \boldsymbol{I}$. Each window function is then a delta function, such that bandpowers do not contain leakage from other bins, and contain power from only that bin. However, the disadvantage of this becomes apparent if one also computes the error bars on the bandpower estimates. The error bars are obtained by taking the square root of the diagonal of the covariance matrix, which is defined as

$$\boldsymbol{\Sigma} = {\rm Cov}(\hat{\boldsymbol{p}}) = \langle \hat{\boldsymbol{p}}\hat{\boldsymbol{p}}^\dagger\rangle - \langle\hat{\boldsymbol{p}}\rangle\langle\hat{\boldsymbol{p}}\rangle^\dagger. \quad (18)$$

Since $\hat{\boldsymbol{p}} = \boldsymbol{M}\hat{\boldsymbol{q}}$, it is easily shown that

$$\boldsymbol{\Sigma} = \boldsymbol{M}\boldsymbol{F}\boldsymbol{M}^\dagger. \quad (19)$$

The choice of $\boldsymbol{M} = \boldsymbol{F}^{-1}$ tends to give rather large error bars. At the other extreme, picking $\boldsymbol{M}_{\alpha\beta} \propto \delta_{\alpha\beta}/\boldsymbol{F}_{\alpha\alpha}$ (with the proportionality constant fixed by our normalization criterion) leads to the smallest possible error bars (Tegmark 1997), at the expense of broader window functions. In our application of OQEs in the following sections, we will pick an intermediate choice for $\boldsymbol{M}$, one that is carefully tailored to avoid the leakage of foreground power from low k modes to high k modes.

5.2. Application of OQE

Here we describe the specifics of our application of the OQE formalism to measure the power spectrum. Doing so requires defining various quantities such as $\boldsymbol{x}$, $\boldsymbol{C}$, $\boldsymbol{Q}_\alpha$ for our analysis pipeline.

First, we consider $\boldsymbol{x}$, which represents the data in our experiment. Our data set consists of visibilities as a function of frequency and time for each baseline in the array. In our analysis, we group the baselines into three groups of redundant baselines (described in Section 2), in the sense that within each group there are multiple copies of the same baseline. In the description that follows, we first estimate the power spectrum separately for each group. Power spectrum estimates obtained

from the different redundant groups are then combined in a set of averaging and bootstrapping steps described in Section 5.4. Note that because our data have been fringe-rate filtered in the manner described in Section 3.5, we may reap all the benefits of coherently integrating in time simply by estimating the power spectrum for every instant in the LST-binned data before averaging over the time-steps within the LST-binned day (Parsons et al. 2015).

For the next portion of our discussion, consider only the data within a single redundant group. Within each group there are not only multiple identical copies of the same baseline, but in addition (as discussed in Section 3.3), our pipeline also constructs two LST-binned data sets, one from binning all even-numbered days in our observations, and the other from all odd-numbered days. Thus, we have not a single data vector, but a whole family of them, indexed by baseline (i) and odd versus even days (r). Separating the data out into independent subgroups allows one to estimate cross-power spectra rather than auto-power spectra in order to avoid the noise bias, as discussed in the previous section. The data vectors take the form

$$\boldsymbol{x}^{ri}(t) = \begin{pmatrix} V^{ri}(\nu_1, t) \\ V^{ri}(\nu_2, t) \\ \vdots \end{pmatrix}, \tag{20}$$

where $V^{ri}(\nu, t)$ is the visibility at frequency ν at time t. Each data vector is 20 elements long, being comprised of 20 channels of a visibility spectrum spanning 10 MHz of bandwidth centered on 151.5 MHz.

Having formed the data vectors, the next step in Equation (12) is to weight the data by their inverse covariance. To do so, we of course require the covariance matrix $\boldsymbol{C}$, which by definition, is the ensemble average of $\boldsymbol{x}\boldsymbol{x}^\dagger$, namely $\boldsymbol{C} = \langle \boldsymbol{x}\boldsymbol{x}^\dagger \rangle$. Unfortunately, in our case the covariance is difficult to model from first principles, and we must resort to an empirically estimated $\boldsymbol{C}$. We make this estimation by taking the time average of the quantity $\boldsymbol{x}\boldsymbol{x}^\dagger$ over 8.5 hr of LST, estimating a different covariance matrix for each baseline and for odd versus even days. While an empirical determination of the covariance is advantageous in that it captures features that are difficult to model from first principles, it carries the risk of cosmological signal loss (Switzer & Liu 2014). We will discuss and quantify this signal loss in Section 5.3.

To gain some intuition for the action of $\boldsymbol{C}^{-1}$ on our data, let us examine the combination

$$\boldsymbol{z}^{ri} = \left(\boldsymbol{C}^{ri}\right)^{-1} \boldsymbol{x}^{ri} \tag{21}$$

for select baselines. This is a crucial step in the analysis since it suppresses coherent frequency structures (such as those that might arise from residual foregrounds). Note that the inverse covariance weighting employed here differs from that in P14, in that P14 modeled and included covariances between different baselines, whereas in our current treatment we only consider covariances between different frequency channels. Figure 13 compares the effects of applying the inverse covariance matrix to a data vector that contains foregrounds (and thus contains highly correlated frequency structures) to one in which foregrounds have been suppressed by the wideband delay filter described in Section 3.3. In the figure, the top row corresponds to the data vector $\boldsymbol{x}^{ri}$ for three selected baselines in the form of a waterfall plot of visibilities, with frequency on the horizontal axis and time on the vertical axis. The middle section shows the empirical estimate of the covariance by taking the outer product of $\boldsymbol{x}$ with itself and averaging over the time axis. Finally, the last row shows the results of inverse covariance weighting the data, namely $\boldsymbol{z}^{ri}$. In every row, the foreground-dominated data are shown in the left half of the figure, while the foreground-suppressed data are shown in the right half.

Consider the foreground-dominated $\boldsymbol{x}^{ri}$ in Figure 13, and their corresponding covariance matrices. The strongest modes that are present in the data are the eigenmodes of the covariance matrix with the largest eigenvalues. Figure 14 shows the full eigenvalue spectrum and the four strongest eigenmodes. For the foreground-dominated data, one sees that the eigenvalue spectrum is dominated by the first few modes, and the corresponding eigenmodes are rather smooth, highly suggestive of smooth spectrum foreground sources. The application of the inverse covariance weighting down-weights these eigenmodes, revealing waterfall plots in the bottom row of Figure 13 that look more noise-dominated. With the foreground-suppressed portion (right half) of Figure 13, the initial $\boldsymbol{x}^{ri}$ vectors already appear noise dominated (which is corroborated by the relatively noisy form of the eigenvalue spectra in Figure 14). The final $\boldsymbol{z}^{ri}$ vectors remain noise-like, although some smooth structure (perhaps from residual foregrounds) has still been removed, and finer scale noise has been up-weighted.

With intuition established for the behavior of $\boldsymbol{C}^{-1}$, we may group our identical baselines into five different sets and average together $\boldsymbol{z}^{ri}$ vectors for baselines within the same set. That is, we form

$$\boldsymbol{z}_A^r = \sum_{i \in A} \left(\boldsymbol{C}^{ri}\right)^{-1} \boldsymbol{x}^{ri}, \tag{22}$$

where A ranges from 1 to 5 and indexes the baseline set. At this point, we have 10 weighted data vectors $\boldsymbol{z}$ (5 baseline sets, each of which has an even day and odd day version) for every LST-binned time-step. As discussed in the previous section, instrumental noise bias may be avoided by forming cross-power spectra rather than auto-power spectra. Generalizing Equation (13) to our present case where we have 10 different data vectors, we have

$$\hat{\boldsymbol{q}}_\alpha = \sum_{\substack{A,B,r,s \\ r \neq s, A \neq B}} \boldsymbol{z}_A^{r\dagger} \boldsymbol{Q}_\alpha \boldsymbol{z}_B^s, \tag{23}$$

so that auto-power contributions from identical baseline groups or identical even/odd indices never appear. Residual foreground bias will remain in Equation (23), but in order to avoid possible signal loss from an overly aggressive foreground bias removal scheme, we conservatively allow the foreground bias to remain. Since foreground power will necessarily be positive, residual foregrounds will only serve to raise our final upper limits.

In order to implement Equation (23), it is necessary to derive a form for $\boldsymbol{Q}_\alpha \equiv \partial \boldsymbol{C} / \partial p_\alpha$. To do so, we follow the delay

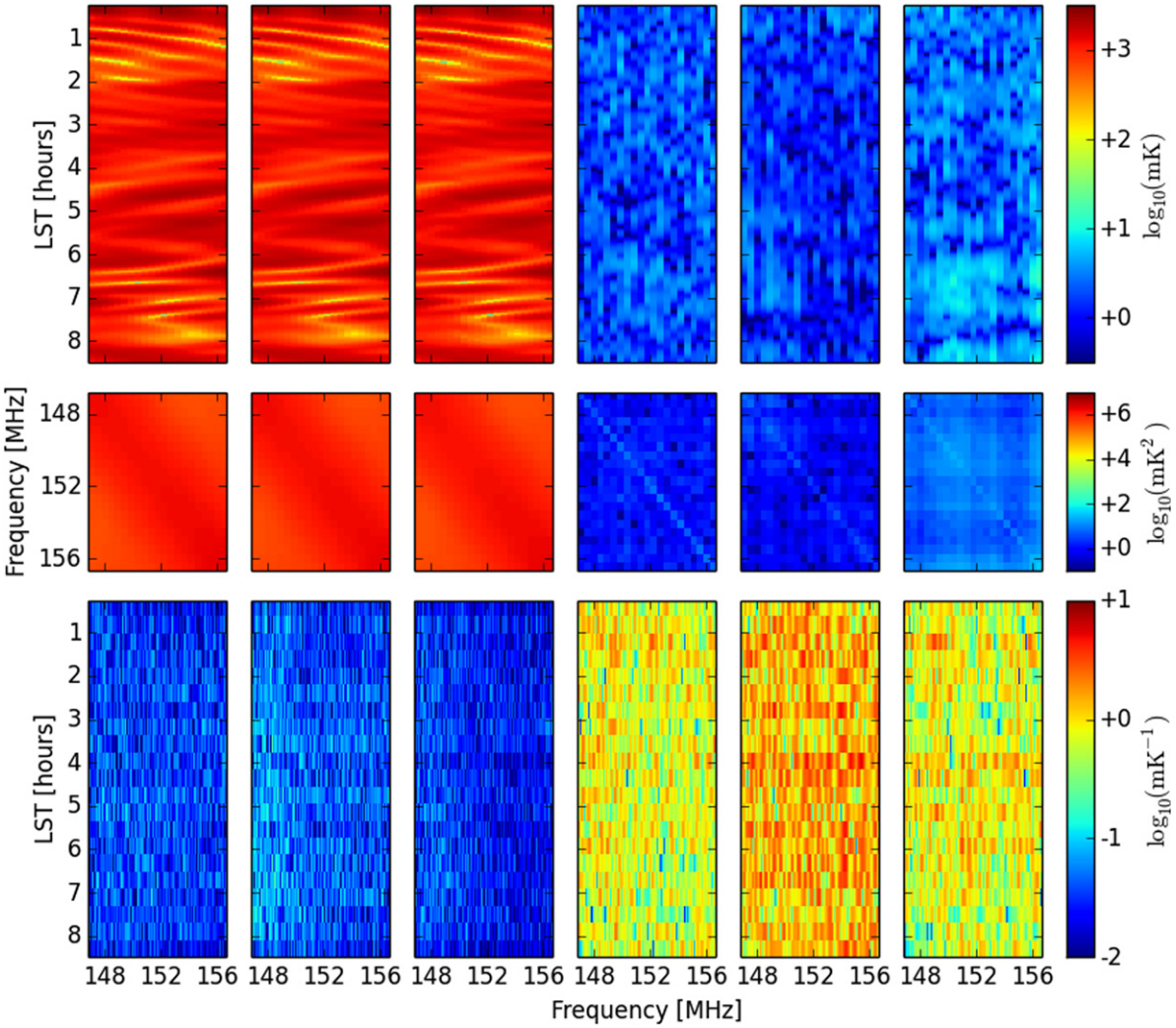

Figure 13. Visibilities before (top row) and after (bottom row) inverse covariance weighting. Signal covariance (middle row) is estimated empirically, averaging over LST. The three left/right columns show visibilities from three different baselines in a redundant group before/after delay filtering, respectively.

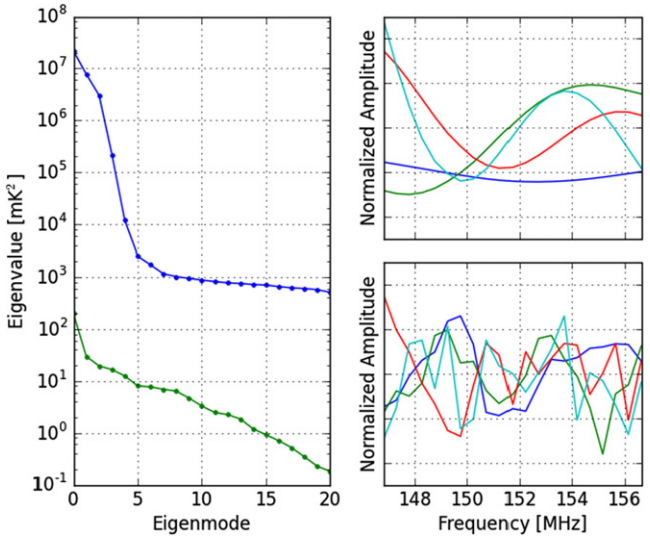

Figure 14. Eigenvalue spectrum of covariance matrices (left) empirically estimated from visibilities before (blue) and after (green) delay filtering. The four strongest eigenmodes of the filtered/unfiltered data are plotted on the top/bottom panels on the right, respectively.

spectrum technique of P12a, where it was shown that

$$P(\boldsymbol{k}_{t\tau}) \approx \left(\frac{\lambda^2}{2k_{\rm B}}\right)^2 \frac{X^2 Y}{\Omega B} \langle \tilde{V}_i(t,\tau)\tilde{V}_j^*(t,\tau)\rangle, \tag{24}$$

where $V_i(t,\tau)$ is the delay transform of baseline visibilities given by Equation (8), X and Y are the constants that convert from angles and frequency to the co-moving coordinate, respectively, Ω is the power squared beam (see Appendix B of P14), B is the bandwidth, λ is the spectral wavelength, and $k_{\rm B}$ is Boltzmann's constant. This suggests that in order to estimate the power spectrum from visibilities, one only needs to Fourier transform along the frequency axis (converting the spectrum into a delay spectrum) before squaring and multiplying by a scalar. Thus, the role of $\boldsymbol{Q}_\alpha$ in Equation (23) is to perform a frequency Fourier transform on each copy of $\boldsymbol{z}$. It is therefore a separable matrix of the form $\boldsymbol{Q}_\alpha = \boldsymbol{m}_\alpha \boldsymbol{m}_\alpha^\dagger$, where $\boldsymbol{m}_\alpha$ is a complex sinusoid of a specific frequency corresponding to delay mode α. We may thus write

$$\hat{\boldsymbol{q}}_\alpha = \sum_{\substack{A,B,r,s \\ r\neq s, A\neq B}} \boldsymbol{z}_A^{r\dagger} \boldsymbol{m}_\alpha \boldsymbol{m}_\alpha^\dagger \boldsymbol{z}_B^s. \tag{25}$$

With an explicit form for $\boldsymbol{Q}_\alpha$, one now also has the necessary ingredients to compute the Fisher matrix using Equation (17).

Having computed the $\hat{\boldsymbol{q}}_\alpha$s, we group our results into a vector $\hat{\boldsymbol{q}}$. This vector of unnormalized bandpowers is then normalized to form our final estimates of the power spectrum $\boldsymbol{p}$. As noted above, the normalization occurs by the $\boldsymbol{M}$ matrix in Equation (14), and can be any matrix of our desire. Even though the choices of the normalization matrices described above have certain good properties, e.g., small error bars or no leakage, we opt for a different choice of window function, as follows. We first reorder the elements in $\hat{\boldsymbol{q}}$ (and therefore in $\boldsymbol{F}$, $\boldsymbol{M}$, and $\hat{\boldsymbol{p}}$ for consistency) so that the k-modes are listed in ascending order, from low k to high k, with the exception that we place the highest k bin third after the lowest two k bins. (The reason for this exception will be made apparent shortly). We then take the Cholesky decomposition of the Fisher matrix, such that $\boldsymbol{F} = \boldsymbol{L}\boldsymbol{L}^\dagger$, where $\boldsymbol{L}$ is a lower triangular matrix. Following that, we pick $\boldsymbol{M} = \boldsymbol{D}\boldsymbol{L}^{-1}$, where $\boldsymbol{D}$ is a diagonal matrix chosen to adhere to the normalization constraint that $\boldsymbol{W} = \boldsymbol{M}\boldsymbol{F}$ has rows that sum to unity. In this case, the window function matrix becomes, $\boldsymbol{W} = \boldsymbol{D}\boldsymbol{L}^\dagger$. This means that $\boldsymbol{W}$ is upper triangular, and with our ordering scheme, has the consequence of allowing power to leak from high to low k, but not vice versa. Since our k axis is (to a good approximation) proportional to the delay axis, foregrounds preferentially appear at low k because their spectra are smooth. Reducing leakage from low k to high k thus mitigates leakage of foregrounds into the cleaner, more noise-dominated regions. Additionally, our placement of the highest k bin as the third element in our reordering of $\hat{\boldsymbol{p}}$ prevents leakage from this edge bin that will contain aliased power. Figure 15 shows the resulting window functions.

Our choice of normalization matrix also has the attractive property of eliminating error correlations between bandpower estimates. Using Equation (19), we have that

$$\boldsymbol{\Sigma} = \boldsymbol{D}\boldsymbol{L}^{-1}\boldsymbol{L}\boldsymbol{L}^\dagger\boldsymbol{L}^{-\dagger}\boldsymbol{D} = \boldsymbol{D}^2. \quad (26)$$

The error covariance matrix on the bandpowers is thus diagonal, which implies that our final data points are uncorrelated with one another. This stands in contrast to the power-spectrum estimator used in P14, where the Blackmann–Harris taper function induced correlated errors between neighboring data points.

5.3. Covariance Matrix and Signal Loss

We now discuss some of the subtleties associated with empirically estimating the covariance matrix from the data. Again, the covariance matrix is defined as the ensemble average of the outer product of a vector with itself, i.e.,

$$\boldsymbol{C} = \langle \boldsymbol{x}\boldsymbol{x}^\dagger \rangle, \quad (27)$$

where $\boldsymbol{x}$ is the data (column) vector used in the analysis. In our analysis, we do not have a priori knowledge of the covariance matrix. and thus we must resort to empirical estimates (Dillon et al. 2015). As we have alluded to above, we replace the ensemble average with a time average that runs from 0 to 8:30 LST hours.

Since the OQE method for power spectrum estimation requires the inversion of $\boldsymbol{C}$, it is crucial that our empirically estimated covariance be a full rank matrix. With our data consisting of visibilities over 20 frequency channels, the covariance matrix is a 20 × 20 matrix. Thus, a necessary condition for our estimate to be full rank is for there to be at least 20 independent time samples in our average. As noted in Section 3.5 the fringe-rate filter used corresponds to averaging time samples for 31 minutes. Over the LST range used in this analysis, this corresponds to roughly 20 statistically independent modes in our data after fringe-rate filtering. We therefore have just enough samples for our empirical estimate, and in practice, our covariance matrices are invertible and allow OQE techniques to be implemented.

Another potential problem that occurs from empirically estimating covariances is that it leads to models of the covariance matrix that over-fit the noise. In this scenario, the covariance matrix tells us that there may be modes in the data that should be down-weighted, for example, but if the empirical covariance estimates are dominated by noise, these may just be random fluctuations that need not be down-weighted. Said differently, the weighting of the data by the inverse covariance is heavily influenced by the noise in the estimate of the covariance matrix and thus has the ability to down-weight valid high-variance samples. Over-fitting the noise in this manner carries with it the possibility of cosmological signal loss. This seems to contradict the conventionally recognized feature of OQEs as lossless estimators of the power spectrum (Tegmark 1997). However, the standard proofs of this property assume that statistics such as $\boldsymbol{C}$ are known a priori, which is an assumption that we are violating with our empirical estimates.

In order to deal with possible signal loss, we perform simulations of our analysis pipeline, deriving correction factors that must be applied to our final constraints. We simulate

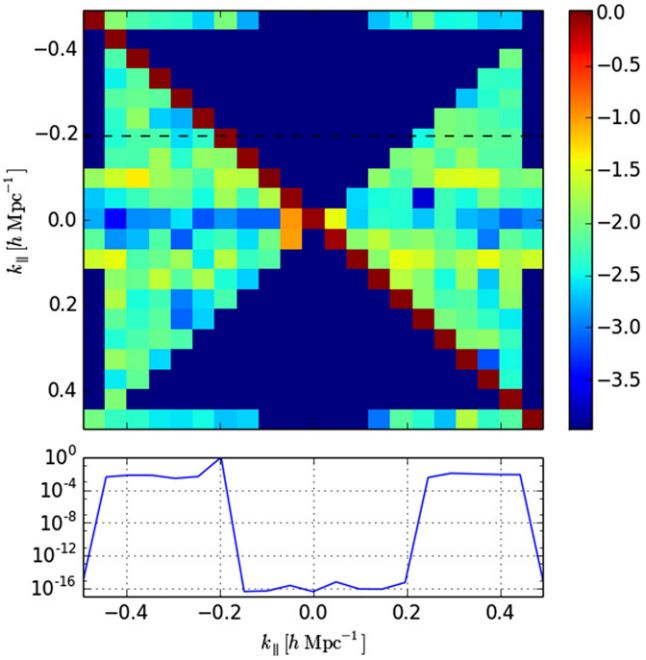

Figure 15. Window function matrix $\boldsymbol{W}$, as defined in Equation (15). The ith row corresponds to the window function used in the estimate of the power spectrum for the ith k-mode. Color scale indicates $\log_{10}\boldsymbol{W}$. The inset plot illustrates the window function along the dashed line in the upper panel. As described in Section 5.2, $\boldsymbol{M}$ in Equation (16) has been chosen so that each window function peaks at the waveband while achieving a high degree of isolation from at lower k-modes that are likely to be biased by foregrounds.

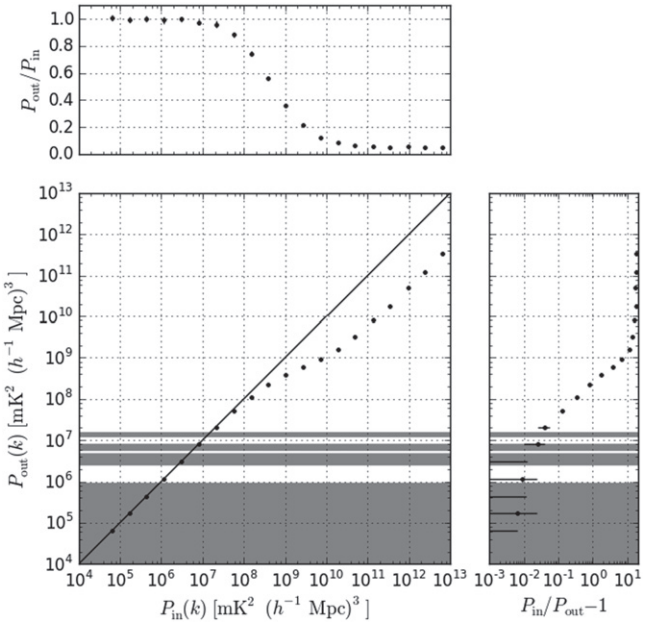

Figure 16. Recovered power spectrum signal as a function of injected signal amplitude. Shaded regions indicate the range in measured amplitude of power spectrum modes in Figure 18. Error bars indicate 95% confidence intervals as determined from the Monte Carlo simulations described in Section 5.3. Because the recovered signal amplitude is a monotonic function of the injected signal amplitude, it is possible to invert the effects of signal loss in the measured power spectrum values to infer the true signal amplitude on the sky. Over the range of powers measured, the maximum correction factor P_{in}/P_{out} is less than 1.02 at 97.5% confidence. The transition to significantly higher correction factors at larger signal amplitudes occurs as the injected signal dominates over the foreground modes present in estimates of the data covariance.

visibilities for Gaussian temperature field with a flat amplitude in $P(k)$ that rotates with the sky, which is fringe-rate filtered in the same way as the data for our fiducial baselines. This signal is processed through our pipeline, and the output power spectrum compared to the input power spectrum, for various levels of input signal amplitude. We repeat this for 40 sky realizations at each signal level. Figure 16 shows the resultant signal loss associated with estimating the covariance matrix from the data. Error bars were obtained through bootstrapping.

As a function of the increasing input amplitude of the simulated power spectra, we find that the ratio of output power to input power decreases, which we interpret as signal loss through the use of our empirical OQE of the power spectrum. However, since the transfer function through this analysis is an invertible function, we can correct for the transfer by using the output value to infer a signal loss that is then divided out to obtain the original input signal level. In Figure 16, we see that deviations from unity signal transfer begin at power spectrum amplitudes of 10^7 mK$^2(h^{-1}$ Mpc$)^3$. For the range of output power spectrum amplitudes in our final estimate of the 21 cm power spectrum (Figure 18), we show that signal loss is <2% at 95% confidence.

As shown in Table 1, the signal loss we characterize for quadratic estimation of the power spectrum band powers is tabulated along with the signal loss associated with each other potentially lossy analysis stage (see Figure 2). We correct for the signal loss in each stage by multiplying the final power spectrum results by the typical loss for each stage, except for modes within the horizon limit and immediately adjacent to the horizon limit, where we apply the maximum signal loss correction to be conservative.

5.4. Bootstrapped Averaging and Errors

When estimating our power spectra via OQEs, we generate multiple samples of the power spectrum in order to apply the bootstrap method to calculate our error bars. In detail, the power spectrum estimation scheme proposed above requires averaging at several points in the pipeline.

1. Visibilities are averaged into five baseline groups after inverse covariance weighting (see Equation (22)).
2. Power spectrum estimates from each of the three redundant baseline types (described in Section 2) are averaged together.
3. Power spectrum estimates from each LST are averaged together.

With the bootstrapping technique, we do not directly perform these averages. Instead, one draws random samples within the three-dimensional parameter space specified above, with replacement, until one has as many random samples as there are total number of parameter space points. These random samples are then propagated through the power spectrum pipeline and averaged together as though they were the original data. This forms a single estimate (a "bootstrap") of $P(\boldsymbol{k})$. Repeating random draws allows one to quantify the inherent scatter—and hence the error bars—in our estimate of $P(\boldsymbol{k})$. When plotting $\Delta^2(k) \equiv k^3P(k)/2\pi^2$ instead of $P(\boldsymbol{k})$, we bin power falling in $+k$ and $-$k, and so we additionally randomize the inclusion of positive and negative k bins.

We compute a total of 400 bootstraps. In combining independent samples for our final power spectrum estimate, we elect to use the median, rather than the mean, of the samples. One can see the behavior of both statistics in Figure 17, where we show how the absolute value of $\Delta^2(k)$ integrates down as more independent samples are included in the mean and median. In this plot, one can see modes integrating down consistent with a noise-dominated power spectrum until they bottom out on a signal. In the noise-dominated regime, the mean and the median behave similarly. However, we see that the median routinely continues to integrate down as noise for longer. This is an indication that the mean is skewed by outlier modes, suggesting variations beyond thermal noise. The magnitude of the difference is also not consistent with the Rayleigh distribution expected of a cosmological power spectrum limited by cosmic variance. For a Rayleigh distribution, the median is $\ln 2 \sim 0.69$ times the mean. Instead, we interpret the discrepancy as a sign of contributions from foregrounds, which are neither isotropic nor Gaussian distributed. Since median provides better rejection of outliers in the distribution that might arise from residual foreground power, we choose to use the median statistic to combine measurements across multiple modes. As listed in Table 1, we apply a $1/\ln 2$ correction factor to our power spectrum estimates to infer the mean from the median of a Rayleigh distribution.

6. RESULTS

6.1. Power Spectrum Constraints

To summarize the previous section, we follow the power spectrum analysis procedure outlined in Section 5.2, we

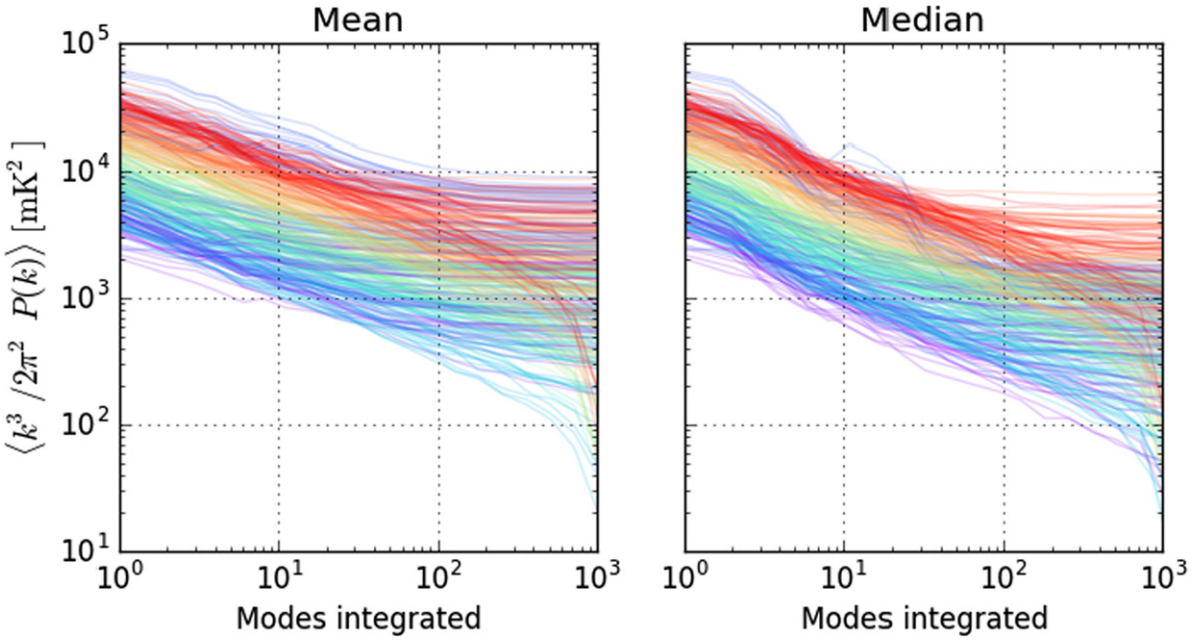

Figure 17. Absolute value of the cumulative mean (left) and median (right), as a function of number of modes of the power spectrum band power for $k_\parallel$ modes ranging from −0.49 (red) to $0.44h$ Mpc^{-1} (violet). Here, modes are defined as samples from different redundant baseline groups and LSTs. This Allen variance plot shows modes averaging down as the square root of number of modes combined until a signal floor is reached. The difference in behavior between the mean and median is an indication of outliers in the distribution of values, likely as a result of foreground contamination. We use the median in the estimation of the power spectrum in Figure 18, along with a correction factor compensating for the difference between the mean and median in estimating variance.

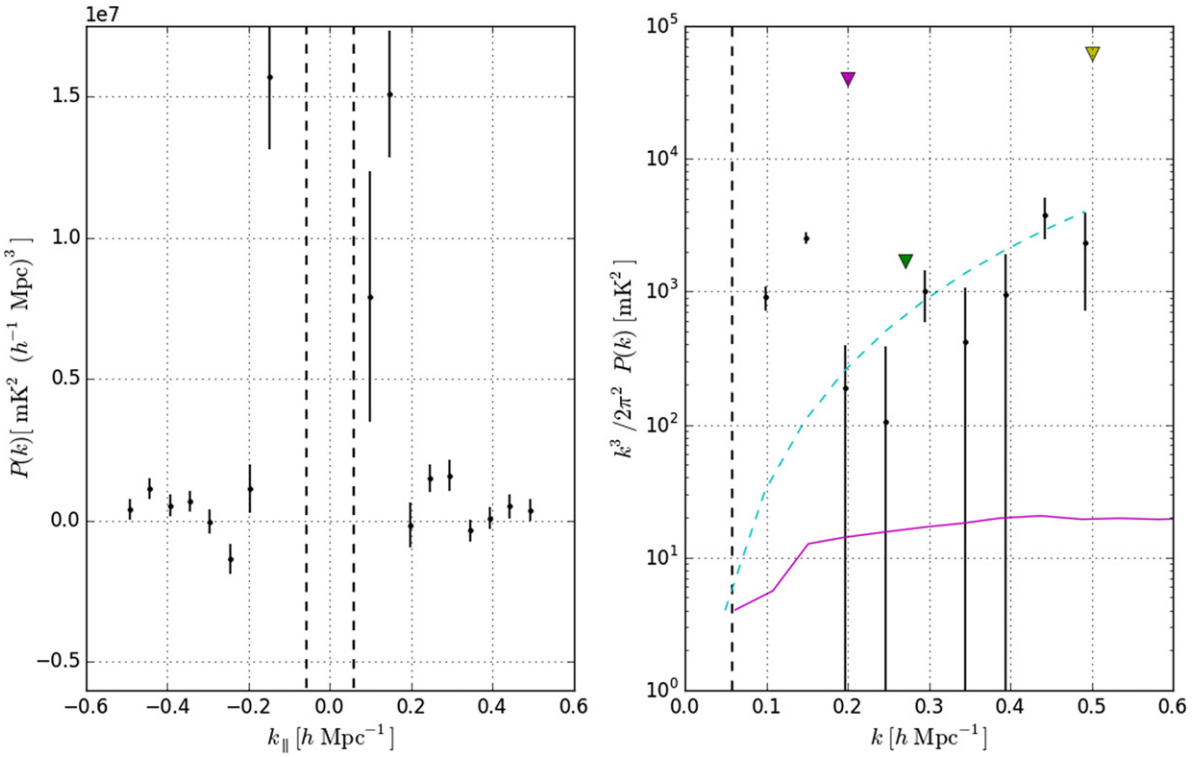

Figure 18. Measured power spectrum (black dots with 2σ error bars) at $z = 8.4$ resulting from a 135 day observation with PAPER-64. The dashed vertical lines at $0.6h$ Mpc^{-1} show the bounds of the delay filter described in Section 3.3. The predicted 2σ upper limit in the absence of the a celestial signal is shown in dashed cyan, assuming $T_{\rm sys} = 500\ K$. The triangles indicate 2 σ upper limits from GMRT (Paciga et al. 2011) (yellow) at $z = 8.6$, MWA (Dillon et al. 2014) at $z = 9.5$ (magenta), and the previous PAPER upper limit (P14) at $z = 7.7$ (green). The magenta curve shows a predicted model 21 cm power spectrum at 50% ionization (Lidz et al. 2008).

incoherently combine independent power spectrum measurements made at different times and with different baseline groups using the median statistic. As described in Section 5.4, we bootstrap over all of these independent measurements, as well as over the selection of baselines included in the power spectrum analysis for each baseline group, in order to estimate

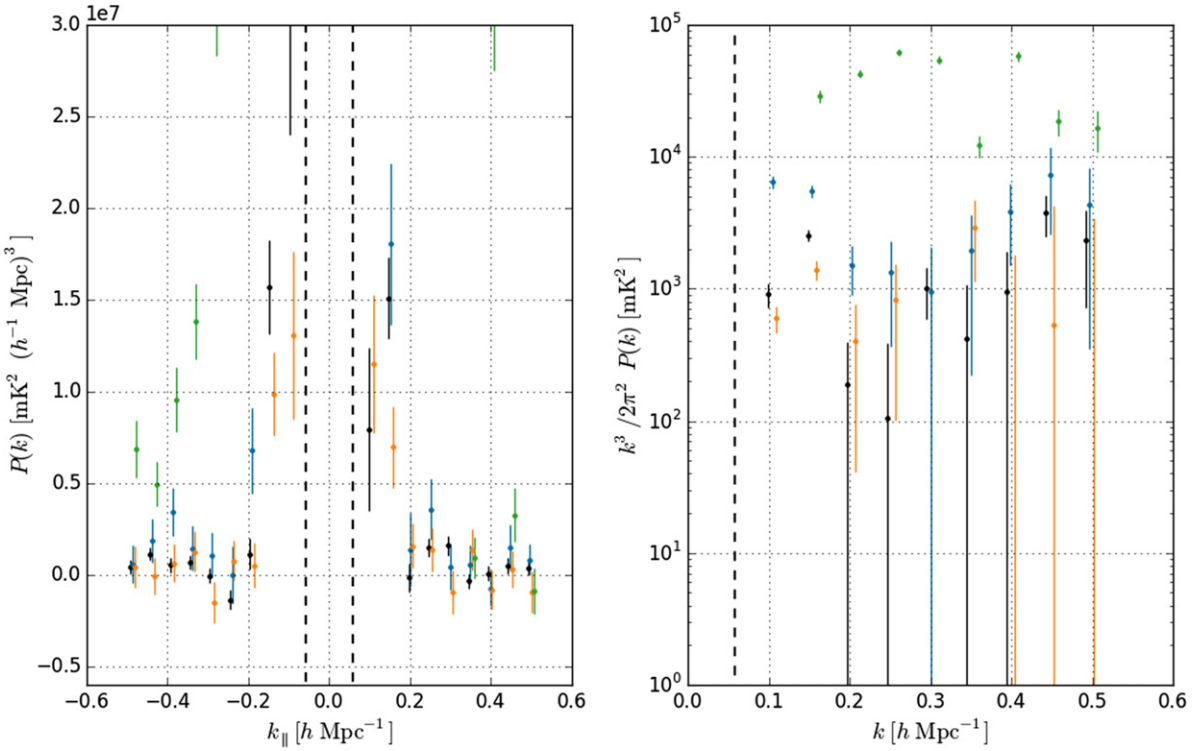

Figure 19. Diagnostic power spectra in the style of Figure 18 illustrating the impact of various analysis stages. The blue power spectrum uses the P14 fringe-rate filter combined with crosstalk removal. Green illustrates the result using the improved fringe-rate filter, but without crosstalk removal. A power spectrum derived without the application of OMNICAL is shown in orange. Black includes improved fringe-rate filtering, crosstalk removal, and OMNICAL calibration; it is the same power spectrum shown in Figure 18.

the error bars on the spherically averaged power spectrum $P(k)$, where positive and negative $k_\parallel$ measurements are kept separate for diagnostic purposes. In the estimation of the dimensionless power spectrum $\Delta^2(k) \equiv k^3 P(k)/2\pi^2$, the folding of $\pm k_\parallel$ is handled along with the rest of the bootstrapping over independent modes. Finally, the measured values for $P(k)$ and $\Delta^2(k)$ are corrected for signal loss through all stages of analysis, as summarized in Table 1.

The final results are plotted in Figure 18. For the first two modes outside of the horizon where $\Delta^2(k)$ is measured, we have clear detections. We attribute these to foreground leakage from inside the horizon related to the convolution kernels in Equation (8) (either from the chromaticity of the antenna response, or from the inherent spectrum of the foregrounds themselves). Somewhat more difficult to interpret are the 2.4σ excess at $k \approx 0.30h$ Mpc^{-1} and the 2.9σ excess at $k \approx 0.44h$ Mpc^{-1}. Having two such outliers is unlikely to be chance.

In examining the effects on the power spectrum of omitting various stages of analysis (see Figure 19), we see a pronounced excess in the green curve corresponding to the omission of crosstalk removal in fringe-rate filtering. While the signal is heavily attenuated in the filtering step, it remains a possibility that the remaining detections are associated with instrumental crosstalk. We do note, however, that the qualitative shape of the excess in the crosstalk-removed data does not appear to match that of the crosstalk-containing data.

Another likely possibility is that the signal might be associated with foregrounds. Foregrounds, which are not generally isotropically distributed on the sky, are likely to be affected by the spatial filtering associated with fringe-rate filtering, whereas a statistically isotropic signal is not. Indeed, we see that excesses in many modes measured using the P14 stype time-domain filtering (blue in Figure 19) decrease significantly using the improved fringe-rate filter. As discussed in Parsons et al. (2015), the normalization applied to $\Omega_{\rm eff}$ for fringe-rate filtering correctly compensates for the effect of this filtering on power-spectral measurements of a statistically isotropic Gaussian sky signal. We can surmise from any significant change in amplitude of the excess under fringe-rate filtering that it arises from emission that violates these assumptions. We conclude, therefore, that this excess is unlikely to be cosmic reionization, and is more likely the result of non-Gaussian foregrounds. As discussed earlier, one possible culprit is polarization leakage (Jelić et al. 2010, 2014; Moore et al. 2013), although further work will be necessary to confirm this. The interpretation of the signal as polarization leakage is, however, rather high to be consistent with recent measurements in Stokes Q presented in Moore et al. (2015), where the leakage is constrained to be <100 mK^2 for all k.

That the excesses at $k \approx 0.30$ and $0.44h$ Mpc^{-1} are relatively unaffected by the filtering could be an indication that they are more isotropically distributed, but more likely, it may mean that the simply arise closer to the center of the primary beam where they are down-weighted less. Both excesses appear to be significantly affected by omitting OMNICAL calibration (orange in Figure 19). This could be interpreted as indicating the excess is a modulation induced by frequency structure in the calibration solution. However, OMNICAL is constrained to prohibit structure common to all baselines, so a more likely interpretation is that this faint feature decorrelates without the

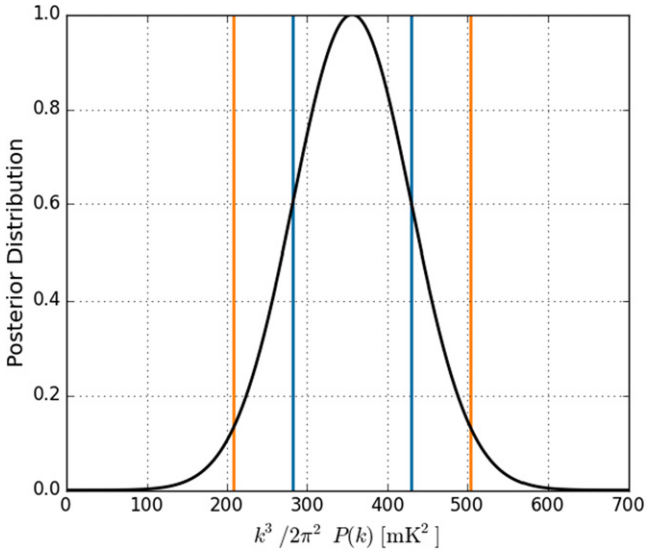

Figure 20. Posterior distribution of power spectrum amplitude for a flat $\Delta^2(k)$ power spectrum over $0.15 < k < 0.5h$ Mpc^{-1} (solid black), assuming Gaussian error bars. The blue and orange vertical lines correspond to the 1σ and 2σ bounds, respectively.

precision of redundant calibration. To determine the nature of these particular excesses, further work will be necessary.

In order to aggregate the information presented in the power spectrum into a single upper limit, we fit a flat $\Delta^2(k)$ model to measurements in the range $0.15 < k < 0.5h$ Mpc^{-1}. We use a uniform prior of amplitudes between −5000 and 5000 mK^2, and assume measurement errors are Gaussian. Figure 20 shows the posterior distribution of the fit. From this distribution, we determine a mean of $(18.9\ \mathrm{mK})^2$ and a 2σ upper limit of $(22.4\ \mathrm{mK})^2$. The measured mean is inconsistent with zero at the 4.7σ level, indicating that we are detecting a clear power spectrum excess at $k > 0.15h$ Mpc^{-1}.

We suspect that the excess in our measured power spectrum is likely caused by crosstalk and foregrounds. We therefore suggest ignoring the lower bound on the power spectrum amplitude as not being of relevance for the cosmological signal. On the other hand, since foreground power is necessarily positive, the 2σ upper limit of $(22.4\ \mathrm{mK})^2$ at $z = 8.4$, continues to serve as a conservative upper limit. This significantly improves over the previous best upper limit of $(41\ \mathrm{mK})^2$ at $z = 7.7$ reported in P14. As we show below and in greater detail in Pober et al. (2015), this limit begins to have implications for the heating of the IGM prior to the completion of reionization.

6.2. Spin Temperature Constraints

In this section, we examine the implication of the measured upper limits on 21 cm emission in Figure 18 on the spin temperature of the 21 cm line at $z = 8.4$. In a forthcoming paper (Pober et al. 2015), we conduct a thorough analysis of the constraints that can be put on the IGM using a simulation-based framework. As a complement to that more thorough analysis, we focus here on a simpler parameterization of the shape of the 21 cm power spectrum signal.

The brightness temperature of the 21 cm signal, δT_b, arising from the contrast between the cosmic microwave background, T_γ, and the spin temperature, T_s, is given by

$$\delta T_b = \frac{T_s - T_\gamma}{1+z}\left(1 - e^{-\tau}\right) \approx \frac{T_s - T_\gamma}{1+z}\tau, \tag{28}$$

where temperatures are implicitly a function of redshift z, and the approximation holds for low optical depth, τ. The optical depth is given by (Zaldarriaga et al. 2004)

$$\tau = \frac{3c^3\hbar A_{10} n_{\mathrm{H\,I}}}{16k\nu_0^2 T_s H(z)} \tag{29}$$

where A_{10} is the Einstein A coefficient for the 21 cm transition, $n_{\mathrm{H\,I}}$ is the density of the neutral hydrogen, $H(z)$ is the Hubble constant, $x_{\mathrm{H\,I}}$ is the neutral fraction of hydrogen, δ is the local baryon overdensity, ν_0 is the rest frequency of the 21 cm transition, and the remainder are the usual constants. Plugging in the cosmological parameters from Planck Collaboration et al. (2015), we get

$$\delta T_b \approx T_0\, x_{\mathrm{H\,I}}(1+\delta)\xi, \tag{30}$$

where $\xi \equiv 1 - T_\gamma/T_s$ and $T_0 \equiv 26.7\ \mathrm{mK}\sqrt{(1+z)/10}$.

If the spin temperature is larger than T_γ, we get the 21 cm signal in emission with respect to the CMB, and $\xi \sim 1$. However, if T_s is less than T_γ, δT_b is negative and ξ can potentially become large.

As in P14, we consider a "weak heating" scenario in which T_s is coupled to the gas temperature via the Wouthuysen-Field effect (Wouthuysen 1952; Field 1958; Hirata 2006), but little heating has taken place prior to reionization, so that $T_s < T_\gamma$. In this scenario, because we have assumed little heating, we can approximate ξ as having negligible spatial dependence, and therefore $T_0^2\xi^2$ becomes a simple multiplicative scalar to the 21 cm power spectrum:

$$\Delta^2_{21}(k) = T_0^2\xi^2(z)\Delta_i^2(k), \tag{31}$$

where $\Delta_i^2(k)$ is the dimensionless H I power spectrum.

As shown in P14, the maximum value of the prefactor in Equation (31) is given by a no-heating scenario where the spin temperature follows the kinetic gas temperature, which is held in equilibrium with the CMB via Compton scattering until $z_{\mathrm{dec}} \approx 150$ (Furlanetto et al. 2006) and then cools adiabatically as $(1+z)^2$. In this case, ξ is given by

$$\xi = 1 - \frac{1+z_{\mathrm{dec}}}{1+z} \approx -\frac{150}{1+z}. \tag{32}$$

At $z = 8.4$, this corresponds to a minimum bound on the spin temperature of $T_s > 1.5$ K.

We can now flip this argument around and, for a measured upper bound on $\Delta^2_{21}(k)$, we can use models for $\Delta_i^2(k)$ in Equation (31) to place a bound on T_s. We consider a class of "patchy" reionization models (P12a; P14) which approximates the ionization power spectrum as flat between minimum and maximum bubble sizes, k_{min} and k_{max}, respectively:

$$\Delta_i^2(k) = \left(x_{\mathrm{H\,I}} - x_{\mathrm{H\,I}}^2\right)\Big/\ln\left(k_{\mathrm{max}}/k_{\mathrm{min}}\right). \tag{33}$$

For combinations of k_{min} and k_{max}, we determine the minimum spin temperature implied by the 2σ 21 cm power spectrum upper limits shown in Figure 18. Figure 21 shows the results of

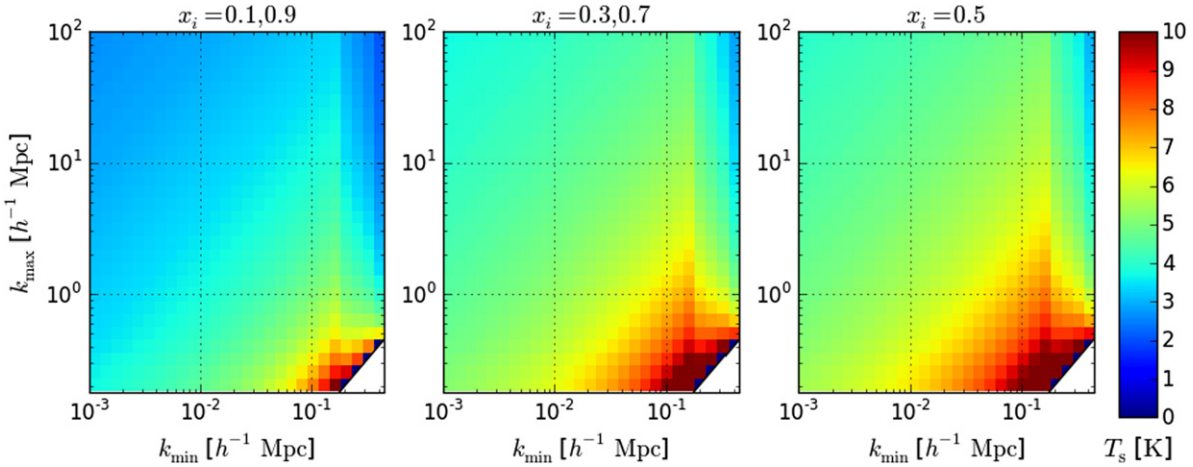

Figure 21. Constraints on the 21 cm spin temperature at $z = 8.4$, assuming the patchy reionization model in Equations (31) and (33), which hold in the limit that $T_{\rm s} < T_{\rm CMB}$.

these bounds for neutral fractions of $x_{\rm H\,I} = 0.1$, 0.3, 0.5, 0.7, and 0.9. In almost all cases (excepting $x_{\rm H\,I} = 0.1$, 0.9 for $k_{\rm min} < 0.1h$ Mpc^{-1}), we find that $T_{\rm s} \gtrsim 3$ K, indicating that our measurements are inconsistent with the spin temperature being coupled to a kinetic temperature governed strictly by adiabatic expansion.

Our results become more interesting in the range of $k_{\rm min} \sim 0.1$ and $k_{\rm max} \sim 30$ representative of fiducial simulations (Zahn et al. 2007; Lidz et al. 2008). For neutral fractions of 0.3, 0.5, and 0.7, we find that $T_{\rm s} \gtrsim 4$ K. Pober et al. (2015) improves on these results by using a simulation-based framework, rather than relying on coarse parametrizations of the power spectrum shape. They compare the limits they find to the amount of heating possible given the currently observed star formation rates in high-redshift galaxy populations (Bouwens et al. 2014; McLeod et al. 2015) and assumptions about the relationship between star formation rates and X-ray luminosities (Furlanetto et al. 2006; Pritchard & Loeb 2008; Fialkov et al. 2014). Assuming the midpoint of reionization lies close to $z = 8.4$ (a reasonable assumption given that Planck Collaboration et al. 2015 suggests a midpoint of $z = 8.8$), both the bounds found in this paper and Pober et al. (2015) show evidence for heating that places constraints on the possible values for the star formation rate/X-ray luminosity correlation given certain models of the star formation rate density redshift evolution. We refer the reader to Pober et al. (2015) for a detailed examination of these results.

7. DISCUSSION

The improvement in our results over those in P14 are the result of four major advances:

1. the expansion of PAPER to 64 antennas doubled our instrument's power spectrum sensitivity,
2. using OMNICAL for redundant calibration significantly improved the clustering of measurements over the previous implementation of LOGCAL used in P14,
3. fringe-rate filtering further improved power spectrum sensitivity by ∼50% and suppressed systematics associated with foregrounds low in the primary beam, and
4. moving from a lossless quadratic estimator targeting difference modes in redundant measurements to an OQE (with carefully calibrated signal loss) significantly reduced contamination from residual foregrounds.

Figure 19 illustrates the effect of some of these advances on the final power spectrum. Other important advances include the use of the median statistic to reduce the impact of non-Gaussian outliers in power-spectral measurements, and the use of a Cholesky decomposition of the Fisher information matrix to help reduce leakage from highly contaminated modes within the wedge.

These new techniques and improvements to calibration have reduced the measured bias in nearly all wavebands by an order of magnitude or more. The use of OMNICAL to accurately calibrate the relative complex gains of the antennas has shown to be a major improvement to the data-reduction pipeline. The accuracy and improvement of this calibration brings redundant baselines into impressive agreement with one another (see Figures 4 and 10), and provides important diagnostic information for monitoring the health of the array, flagging RFI events, and otherwise assessing data quality. Fringe-rate filtering, which is described in greater depth in Parsons et al. (2015), is also proving to be a flexible and powerful tool for controlling direction-dependent gains and improving sensitivity.

As sensitivity improves, it will be possible to determine more accurately than Moore et al. (2015) what the actual level of polarized emission, and thus leakage, may be. Independent fringe-rate filtering of the XX and YY polarizations prior to summation has the potential to better match these polarization beams and further suppress the leakage signal if the polarized signal turns out to be significant.

The end result is a major step forward, both for PAPER and for the field of 21 cm cosmology. While we have not yet made a detection of the 21 cm cosmological signal, our limits are now within the range of some of the brighter models. As discussed in Pober et al. (2015), another order-of-magnitude improvement in sensitivity will make 21 cm measurements highly constraining.

8. CONCLUSIONS

We present new upper limits on the 21 cm reionization power spectrum at $z = 8.4$, showing a factor of ∼4 improvement over the previous best result (P14). We find a

2σ upper limit of $(22.4\,\mathrm{mK})^2$ by fitting a flat power spectrum in a k range from $0.15 < k < 0.5\,h\,\mathrm{Mpc}^{-1}$ to the dimensionless power spectrum, $\Delta^2(k)$, measured by the PAPER instrument. We coarsely show that these upper limits imply a minimum spin temperature for hydrogen in the IGM. Although these limits are dependent on the model chosen for the power spectrum, we use a patchy reionization model to show that limits of $T_s > 4$ K are fairly generic for models with ionization fractions between 0.3 and 0.7. A more detailed analysis of the implied constraints on spin temperature using semi-analytic reionization/heating simulations is presented in a forthcoming paper (Pober et al. 2015).

The power spectrum results that we present continue to be based on the delay-spectrum approach to foreground avoidance presented in P12b and first applied in P14. The application of a delay filter over a wide bandwidth continues to be one of the most powerful techniques yet demonstrated for managing bright smooth-spectrum foregrounds. In this paper, we extend the analysis in P14 with improved fringe-rate filtering, improved redundant calibration with OMNICAL, and with an OQE that, while not perfectly lossless, is more adept at down-weighting residual foregrounds. The combined effect of these improvements leaves a power-spectral measurement that is not consistent with zero at the 4.7σ-level, which we expect is a result of contamination from crosstalk and foregrounds. With the expansion of PAPER to 64 antennas, the extended 135 day observing campaign, and the added sensitivity benefits of fringe-rate filtering, combined with the optimization of antenna positions in PAPER for highly redundant measurements, this thermal noise limit is beginning to enter the realm of constraining realistic models of reionization.

Forthcoming from PAPER will be two seasons of observation with a 128-element array. Following the same analysis as presented here, that data set is expected to improve over the PAPER-64 sensitivity by a factor of ~4 (in mK^2), with the potential for another boost to sensitivity should the new 16 m baselines provided in the PAPER-128 array configuration prove to be usable. There also remains the potential for further improvements to sensitivity through the use of longer baselines, if foregrounds can be managed effectively. As has been done recently for PAPER-32 (Jacobs et al. 2015; Moore et al. 2015), future work will also extend PAPER-64 analysis to a range of redshifts and examine the power spectrum of polarized emission.

With recent breakthroughs in foreground management, the sensitivity limitations of current experiments are becoming clear. Although collecting area is vital, as discussed in Pober et al. (2014), the impact of collecting area depends critically on the interplay of array configuration with foregrounds. Despite a large spread in collecting areas between PAPER, the MWA, and LOFAR, in the limit that foreground avoidance is the only viable strategy, these arrays all deliver, at best, comparable low-significance detections of fiducial models of reionization. To move beyond simple detection, next-generation instruments must deliver much more collecting area with very compact arrays.

The HERA and the low frequency Square Kilometre Array (SKA-Low) are next generation experiments that aim to make significant detections of the 21 cm power spectrum and begin characterizing it. SKA-Low has secured pre-construction funding for a facility in western Australia. HERA was recently granted funding for its first phase under the National Science Foundation's *Mid-Scale Innovations Program*. HERA uses a close packing of 14 m diameter dishes designed to minimize the width of the delay-space kernel $\tilde{A}_\tau$ in Equation (8). Sensitivity forecasts for a 331-element HERA array and SKA-Low show that they can deliver detections of the 21 cm reionization signal at a significance of 39σ and 21σ, respectively, using the same the conservative foreground avoidance strategy employed in this paper (Pober et al. 2014). HERA is the natural successor to PAPER, combining a proven experimental strategy with the sensitivity to deliver results that will be truly transformative for understanding of our cosmic dawn.

PAPER is supported by grants from the National Science Foundation (NSF; awards 0804508, 1129258, and 1125558). A.R.P., J.C.P., and D.C.J. would like to acknowledge NSF support (awards 1352519, 1302774, and 1401708, respectively). J.E.A. would like to acknowledge a generous grant from the Mount Cuba Astronomical Association for computing resources. We graciously thank SKA-SA for site infrastructure and observing support. We also thank interns Monde Manzini and Ruvano Casper from Durban University of Technology, who helped expand the array from 32 to 64 antennas. Thanks also to Josh Dillon for helpful discussions on optimal quadratic estimators.

REFERENCES

Ali, S. S., Bharadwaj, S., & Chengalur, J. N. 2008, MNRAS, 385, 2166
Battaglia, N., Natarajan, A., Trac, H., Cen, R., & Loeb, A. 2013, ApJ, 776, 83
Bernardi, G., McQuinn, M., & Greenhill, L. J. 2009, A&A, 500, 965
Bernardi, G., McQuinn, M., & Greenhill, L. J. 2015, ApJ, 799, 90
Bernardi, G., de Bruyn, A. G., Brentjens, M. A., et al. 2010, A&A, 522, A67
Bolton, J. S., Haehnelt, M. G., Warren, S. J., et al. 2011, MNRAS, 416, L70
Bosman, S. E. I., & Becker, G. D. 2015, arXiv:1505.06880
Bouwens, R. J., Bradley, L., Zitrin, A., et al. 2010, ApJL, 709, L133
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2014, ApJ, 795, 126
Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2009a, ApJ, 695, 183
Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2009b, ApJ, 695, 183
Bowman, J. D., & Rogers, A. E. E. 2010, Natur, 468, 796
Burns, J. O., Lazio, J., Bale, S., et al. 2012, AdSpR, 49, 433
Chapman, E., et al. 2013, MNRAS, 429, 165
Clark, M., Plante, P. L., & Greenhill, L. 2013, International Journal of High Performance Computing Applications, 27, 178
Datta, A., Bowman, J. D., & Carilli, C. L. 2010, ApJ, 724, 526
de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., et al. 2008, MNRAS, 388, 247
Dillon, J. S., Liu, A., & Tegmark, M. 2013, PhRvD, 87, 043005
Dillon, J. S., Liu, A., Williams, C. L., et al. 2014, PhRvD, 89, 023002
Dillon, J. S., Neben, A. R., Hewitt, J. N., et al. 2015, arXiv:1506.01026
Ellingson, S. W., Taylor, G. B., Craig, J., et al. 2013, ITAP, 61, 2540
Ellis, R. S., McLure, R. J., Dunlop, J. S., et al. 2013, ApJL, 763, L7
Faisst, A. L., Capak, P., Carollo, C. M., Scarlata, C., & Scoville, N. 2014, ApJ, 788, 87
Fan, X., Carilli, C. L., & Keating, B. 2006, ARA&A, 44, 415
Fialkov, A., Barkana, R., & Visbal, E. 2014, Natur, 506, 197
Field, G. B. 1958, PIRE, 46, 240
Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, PhR, 433, 181
George, E. M., Reichardt, C. L., Aird, K. A., et al. 2014, arXiv:1408.3161
Ghosh, A., Bharadwaj, S., Ali, S. S., & Chengalur, J. N. 2011, MNRAS, 418, 2584
Greenhill, L. J., & Bernardi, G. 2012, arXiv:1201.1700
Harker, G., Zaroubi, S., Bernardi, G., et al. 2009, MNRAS, 397, 1138
Hazelton, B. J., Morales, M. F., & Sullivan, I. S. 2013, ApJ, 770, 156
Hirata, C. M. 2006, MNRAS, 367, 259
Högbom, J. A. 1974, A&AS, 15, 417
Jacobs, D. C., Aguirre, J. E., Parsons, A. R., et al. 2011, ApJL, 734, L34
Jacobs, D. C., Parsons, A. R., Aguirre, J. E., et al. 2013, ApJ, 776, 108
Jacobs, D. C., Pober, J. C., Parsons, A. R., et al. 2015, ApJ, 801, 51
Jelić, V., de Bruyn, A. G., Mevius, M., et al. 2014, A&A, 568, A101

Jelić, V., Zaroubi, S., Labropoulos, P., et al. 2008, MNRAS, 389, 1319
Jelić, V., Zaroubi, S., Labropoulos, P., et al. 2010, MNRAS, 409, 1647
Lidz, A., Zahn, O., McQuinn, M., Zaldarriaga, M., & Hernquist, L. 2008, ApJ, 680, 962
Liu, A., Parsons, A. R., & Trott, C. M. 2014a, PhRvD, 90, 023018
Liu, A., Parsons, A. R., & Trott, C. M. 2014b, PhRvD, 90, 023019
Liu, A., & Tegmark, M. 2011, PhRvD, 83, 103006
Liu, A., & Tegmark, M. 2012, MNRAS, 419, 3491
Liu, A., Tegmark, M., Bowman, J., Hewitt, J., & Zaldarriaga, M. 2009a, MNRAS, 398, 401
Liu, A., Tegmark, M., & Zaldarriaga, M. 2008, arXiv
Liu, A., Tegmark, M., & Zaldarriaga, M. 2009b, MNRAS, 394, 1575
Liu, A., Tegmark, M., Morrison, S., Lutomirski, A., & Zaldarriaga, M. 2010, MNRAS, 408, 1029
Mao, Y., Tegmark, M., McQuinn, M., Zaldarriaga, M., & Zahn, O. 2008, PhRvD, 78, 023529
Marthi, V. R., & Chengalur, J. 2014, MNRAS, 437, 524
Masui, K. W., Switzer, E. R., Banavar, N., et al. 2013, ApJL, 763, L20
McLeod, D. J., McLure, R. J., Dunlop, J. S., et al. 2015, MNRAS, 450, 3032
McQuinn, M., Lidz, A., Zahn, O., et al. 2007, MNRAS, 377, 1043
McQuinn, M., & O'Leary, R. M. 2012, ApJ, 760, 3
Mesinger, A., Ferrara, A., & Spiegel, D. S. 2013, MNRAS, 431, 621
Mesinger, A., McQuinn, M., & Spergel, D. N. 2012, MNRAS, 422, 1403
Moore, D. C., et al. 2015, arXiv:1502.05072
Moore, D. F., Aguirre, J. E., Parsons, A. R., Jacobs, D. C., & Pober, J. C. 2013, ApJ, 769, 154
Morales, M. F., Bowman, J. D., & Hewitt, J. N. 2006, ApJ, 648, 767
Morales, M. F., & Wyithe, J. S. B. 2010, ARA&A, 48, 127
Noorishad, P., Wijnholds, S. J., van Ardenne, A., & van der Hulst, J. M. 2012, A&A, 545, A108
Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2013, ApJ, 773, 75
Offringa, A. R., McKinley, B., Hurley-Walker, N., et al. 2014, MNRAS, 444, 606
Paciga, G., Albert, J. G., Bandura, K., et al. 2013, MNRAS, 433, 639
Paciga, G., Chang, T.-C., Gupta, Y., et al. 2011, MNRAS, 413, 1174
Page, L., Hinshaw, G., Komatsu, E., et al. 2007, ApJS, 170, 335
Park, H., Shapiro, P. R., Komatsu, E., et al. 2013, ApJ, 769, 93
Parsons, A. R., & Backer, D. C. 2009, AJ, 138, 219
Parsons, A., Backer, D., Chang, C., et al. 2006, in 40th Asilomar Conf. on Signals and Systems (Pacific Grove, CA), 2031
Parsons, A., Backer, D., Foster, G. S., et al. 2010, AJ, 139, 1468
Parsons, A. R., Liu, A., Aguirre, J. E., et al. 2014, ApJ, 788, 106
Parsons, A. R., Liu, A., Ali, Z. S., & Cheng, C. 2015, arXiv:1503.05564
Parsons, A., Pober, J., McQuinn, M., Jacobs, D., & Aguirre, J. 2012a, ApJ, 753, 81
Parsons, A. R., Pober, J. E., Aguirre, J. E., et al. 2012b, ApJ, 756, 165
Patra, N., Subrahmanyan, R., Sethi, S., Udaya Shankar, N., & Raghunathan, A. 2015, ApJ, 801, 138
Peterson, J. B., Pen, U.-L., & Wu, X.-P. 2004, MPLA, 19, 1001
Petrovic, N., & Oh, S. P. 2011, MNRAS, 413, 2103
Planck Collaboration et al. 2014, A&A, 571, A16
Planck Collaboration et al. 2015, arXiv:1502.01589
Pober, J. C., Ali, Z. S., Parsons, A. R., et al. 2015, ApJ, 809, 62
Pober, J. C., Parsons, A. R., Aguirre, J. E., et al. 2013, ApJL, 768, L36
Pober, J. C., Parsons, A. R., Jacobs, D. C., et al. 2012, AJ, 143, 53
Pober, J. C., Liu, A., Dillon, J. S., et al. 2014, ApJ, 782, 66
Presley, M., Liu, A., & Parsons, A. 2015, arXiv:1501.01633
Pritchard, J. R., & Loeb, A. 2008, PhRvD, 78, 103511
Pritchard, J. R., & Loeb, A. 2012, RPPh, 75, 086901
Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, ApJL, 802, L19
Robertson, B. E., Furlanetto, S. R., Schneider, E., et al. 2013, ApJ, 768, 71
Santos, M. G., Cooray, A., & Knox, L. 2005, ApJ, 625, 575
Schenker, M. A., Cooray, A., & Knox, L. 2013, ApJ, 768, 196
Sokolowski, M., Tremblay, S. E., Wayth, R. B., et al. 2015, PASA, 32, 4
Sullivan, I. S., Morales, M. F., Hazelton, B. J., et al. 2012, ApJ, 759, 17
Switzer, E. R., & Liu, A. 2014, ApJ, 793, 102
Tegmark, M. 1997, PhRvD, 55, 5895
Thyagarajan, N., Udaya Shankar, N., Subrahmanyan, R., et al. 2013, ApJ, 776, 6
Tingay, S. J., Goeke, R., Bowman, J. D., et al. 2013, PASA, 30, 7
Treu, T., Schmidt, K. B., Trenti, M., Bradley, L. D., & Stiavelli, M. 2013, ApJL, 775, L29
Trott, C. M., Wayth, R. B., & Tingay, S. J. 2012, ApJ, 757, 101
van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2
Vedantham, H., Udaya Shankar, N., & Subrahmanyan, R. 2012, ApJ, 745, 176
Vedantham, H. K., Koopsman, L. V. E., de Bruyn, A. G., et al. 2015, MNRAS, 450, 2291
Visbal, E., Barkana, R., Fialkov, A., Tseliakhovich, D., & Hirata, C. M. 2012, Natur, 487, 70
Voytek, T. C., Natarajan, A., Jáuregui García, J. M., Peterson, J. B., & López-Cruz, O. 2014, ApJL, 782, L9
Wang, X., Tegmark, M., Santos, M. G., & Knox, L. 2006, ApJ, 650, 529
Wieringa, M. H. 1992, ExA, 2, 203
Wouthuysen, S. A. 1952, AJ, 57, 31
Wu, X. 2009, in Bulletin of the American Astronomical Society 41, American Astronomical Society Meeting Abstracts, 213, 226.05
Yatawatta, S., de Bruyn, A. G., Brentjens, M. A., et al. 2013, A&A, 550, A136
Zahn, O., Lidz, A., McQuinn, M., et al. 2007, ApJ, 654, 12
Zahn, O., Reichardt, C. L., Shaw, L., et al. 2012, ApJ, 756, 65
Zaldarriaga, M., Furlanetto, S. R., & Hernquist, L. 2004, ApJ, 608, 622
Zheng, H., Tegmark, M., Buza, V., et al. 2014, MNRAS, 445, 1084