

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Observing the Epoch of Reionization: Power Spectrum Limits and Commissioning Next Generation 21 cm Experiments

Permalink

<https://escholarship.org/uc/item/3046m4mb>

Author

Ali, Zaki Shiraz

Publication Date

2018

Peer reviewed|Thesis/dissertation

Observing the Epoch of Reionization: Power Spectrum Limits and Commissioning Next
Generation 21 cm Experiments

By

Zaki Shiraz Ali

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Astrophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Aaron Parsons, Chair

Professor Carl Heiles

Professor Adrian Lee

Professor Chung-Pei Ma

Summer 2018

Observing the Epoch of Reionization: Power Spectrum Limits and Commissioning Next
Generation 21 cm Experiments

Copyright 2018
by
Zaki Shiraz Ali

Abstract

Observing the Epoch of Reionization: Power Spectrum Limits and Commissioning Next Generation 21 cm Experiments

by

Zaki Shiraz Ali

Doctor of Philosophy in Astrophysics

University of California, Berkeley

Professor Aaron Parsons, Chair

As one of the last unobserved frontiers in the Universe, the Epoch of Reionization (EoR) marks the period when the Universe transitioned from a neutral state to an ionized state, marking the last global phase change. Understanding how the EoR evolved over time and when it occurred will provide evidence for the nature of the first luminous sources of the Universe. Specifically, we'll be able to answer questions such as what were the first galaxies like in terms of their luminosities, masses, and spectral energy distributions? Were they similar to today's galaxies? How did they form? How were they spatially distributed? These questions and many others begin to probe the early times of galaxy formation, a field in need of observations of the earliest galaxies.

In the first chapter, I provide an introduction to 21 cm cosmology, which uses the 21 cm line transition from neutral hydrogen to study the evolution of the Universe. I briefly review 21 cm physics and the evolution of the cosmological 21 cm signal. The challenges of experiments measuring highly redshifted neutral hydrogen are discussed before describing the instruments used to make the measurements of the cosmological 21 cm signal.

The second chapter in this thesis begins by presenting results from Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER) 64 antenna configuration. I start by describing the dataset used in the chapter, followed by the novel redundant calibration technique used to calibrate the grid layout of the PAPER array. I then discuss the application of delay and fringe-rate filtering to suppress foregrounds and optimally combine time samples, respectively. Before presenting results, I layout a quadratic estimator formalism for measuring the 21 cm power spectrum. I finally present the results from the PAPER-64 dataset.

The third chapter of this thesis shifts focus to the new Hydrogen Epoch of Reionization Array (HERA), a close packed, redundant, low frequency array dedicated to detecting and characterizing the EoR and Dark Ages. As a new experiment with terabytes of data being collected per day, the need for a real time processing system is necessary to avoid a backlog

of data and to avoid data storage issues. I present the real time system (RTS) developed for analyzing HERA data, which automatically detects and flags bad antennas, redundantly calibrates the array, and excises radio frequency interference (RFI) in real time. This work builds off the lessons learned from PAPER and the Murchison Wide-Field Array (MWA) based in western Australia. I then present some initial results from the RTS system.

In the final chapter, I conclude the thesis by summarizing the work presented and discuss future directions for 21 cm cosmology.

For my parents.

Contents

List of Figures	iv
List of Tables	xii
Acknowledgments	xiii
1 Introduction	1
1.1 Introduction	1
1.2 A Brief History	3
1.2.1 From the Big Bang to Decoupling	3
1.2.2 Post-Recombination	4
1.3 21 cm Cosmology	5
1.3.1 The Hyperfine Transition of Hydrogen	5
1.3.2 The Evolution of the 21 cm signal	7
1.3.3 Challenges of 21 cm EoR Experiments	11
1.3.4 Experiments to Detect Cosmic Hydrogen	14
1.3.5 PAPER	16
1.3.6 HERA	19
1.4 Outline of Thesis	21
2 Results from the Donald C. Backer Precision Array for Probing the Epoch of Reionization	22
2.1 Introduction	22
2.2 Observations	24
2.3 Calibration	28
2.3.1 Relative Calibration	29
2.3.2 Absolute Calibration	32
2.3.3 Wideband Delay Filtering	33
2.3.4 Binning in LST	38
2.3.5 Fringe-rate Filtering	38
2.4 Instrumental Performance	40
2.4.1 Instrument Stability	40

2.4.2	System Temperature	40
2.5	Power Spectrum Analysis	42
2.5.1	Review of OQEs	45
2.5.2	Application of OQE	47
2.5.3	Covariance Matrix and Signal Loss	52
2.5.4	Bootstrapped Averaging and Errors	56
2.6	An Updated Understanding of Signal Loss and Power Spectrum Estimation	57
2.7	Power Spectrum Results	61
2.7.1	Power Spectrum Constraints	62
2.7.2	Spin Temperature Constraints	65
2.8	Discussion	69
2.9	Conclusions	69
3	Real-Time Systems in Astronomy: An Application to the Hydrogen Epoch of Reionization Array (HERA)	71
3.1	Introduction	71
3.2	Observations	77
3.3	A Real-Time System	77
3.3.1	Data Ingestion and Tracking	77
3.3.2	Bad Antenna Detection	79
3.3.3	Redundant Calibration	85
3.3.4	RFI Excision	93
3.3.5	Absolute Calibration	96
3.4	Conclusions	99
4	Conclusions	101
4.1	Summary of Thesis	101
	Bibliography	103

List of Figures

- 1.1 A schematic diagram of the observable Universe. The CMB is located at $z = 1088$. Between the surface of last scattering and $z \sim 200$ is a region where no observable radiation is emitted and hence remains unobservable. However, during these Dark Ages, the first signals after Recombination are theoretically observable between $z \sim 200$ and $z \sim 50$. The center of the diagram consists of the SDSS galaxy sample, which has systematically surveyed over 3 million objects (with spectra) over a large fraction of the sky up to $z \sim 1$. This is the extent of the locally observed Universe. A number of objects have been detected at redshifts greater than 1, but these are one off detections and does not provide a picture of the Universe as a whole. However, the blue region is a still yet unobserved volume of the Universe that will be able to be mapped out by highly red-shifted neutral hydrogen using the 21 cm spin flip transition (Mao et al. 2008). 2
- 1.2 The evolution of the 21 cm hydrogen signal from a 21 cm simulation. *top*: The evolution of the spatial fluctuations in the differential brightness temperature as defined in Equation 1.5. The color indicates whether the signal is in absorption (blue), emission (red), or no signal (black). As a function of redshift, fluctuations in different fields are dominating the spatial fluctuations, such as the spin temperature or the neutral fraction. These spatial variations encode information on the nature of the first sources such as their spectral energy distributions and what type of sources they were (black holes, population III stars, etc...). *bottom*: The global evolution of the signal which is defined as the spatial average of the differential brightness temperature at a given redshift (black solid curve). The widely varying features in the global signal are due to different physical effects such as collisional coupling, the Wouthuysen-Field effect, heating from the first stars, and finally ionization of neutral hydrogen. The evolution is further described in Section 1.3.2. Figure used with permission from (Pritchard & Loeb 2012). 8
- 1.3 The Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER) in its 128 antenna redundant configuration in September 2013. Located in the Karoo Desert in South Africa on the SKA Radio Reserve, PAPER arranged it's antennas in a grid configuration to boost sensitivity in the Power Spectrum due to its limited collecting area. 18

1.4	The Hydrogen Epoch of Reionization Array (HERA) in the Summer of 2018. Over 100 dishes have been built with old original feeds.	20
2.1	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.	25
2.2	The Global Sky Model (de Oliveira-Costa et al. 2008a), illustrating foregrounds to the 21cm cosmological signal, with contours indicating beam-weighted observing time (relative to peak) for the PAPER observations described in Section 2.2. The map is centered at 6:00 hours in RA.	26
2.3	The 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 tabulated in Table 2.1.	27
2.4	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.	30
2.5	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.	31
2.6	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.	34

2.7	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 minutes 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.	35
2.8	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.	37
2.9	The 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.	41
2.10	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 2.11.	42
2.11	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.	43
2.12	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).	44

2.13	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.	49
2.14	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.	50
2.15	The window function matrix $\mathbf{W}$, as defined in Equation (2.15). The i^{th} row corresponds to the window function used in the estimate of the power spectrum for the i^{th} k -mode. Color scale indicates $\log_{10} \mathbf{W}$. The inset plot illustrates the window function along the dashed line in the upper panel. As described in Section 2.5.2, $\mathbf{M}$ in Equation (2.16) 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.	53
2.16	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 2.18. Error bars indicate 95% confidence intervals as determined from the Monte Carlo simulations described in Section 2.5.3. 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 $P_{\text{in}}/P_{\text{out}}$ 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.	55
2.17	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 $0.44h \text{ Mpc}^{-1}$ (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 2.18, along with a correction factor compensating for the difference between the mean and median in estimating variance.	56

2.18	Measured lossy 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 $0.06h \text{ Mpc}^{-1}$ show the bounds of the delay filter described in Section 2.3.3. The predicted 2σ upper limit in the absence of a celestial signal is shown in dashed cyan, assuming $T_{\text{sys}} = 500K$. The triangles indicate 2σ 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).	63
2.19	Diagnostic power spectra in the style of Figure 2.18 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 2.18.	64
2.20	Posterior distribution of power spectrum amplitude for a flat $\Delta^2(k)$ power spectrum over $0.15 < k < 0.5h \text{ Mpc}^{-1}$ (solid black), assuming Gaussian error bars. The blue and orange vertical lines correspond to the 1σ and 2σ bounds, respectively. This posterior was calculated using the result in Figure 2.18 and is therefore biased low, but represents the type of information aggregation that can be used in a power spectrum estimate.	66
2.21	Constraints on the 21cm spin temperature at $z = 8.4$, assuming the patchy reionization model in Equations (2.39) and (2.41), which hold in the limit that $T_s < T_{\text{CMB}}$.	67
3.1	Antenna configuration of HERA-47 consisting of 40 nominally good antennas (blue) and 7 bad antennas that have been flagged (red) from the dataset. These 7 bad antennas were found as part of HERA's RTS using the mean visibility and correlation metrics discussed in Section 3.3.2. This was the configuration of HERA on October 2, 2017 and for the data used in this Chapter.	74

3.2	A schematic diagram of a real time system (RTS) for HERA. The RTS starts a night of observation by first ingesting the data and loading meta data into the Librarian database (section 3.3.1). An initial observing report is emailed and posted to internal Slack channels for quick inspections of autocorrelations and input signal levels. As this information is shared with the HERA collaboration, bad antenna detection metrics (red, section 3.3.2) are then applied to the raw input data. Redundant calibration algorithms (magenta, section 3.3.3) are then used to derive the per antenna complex gain solutions. Each redundant calibration stage produces their own metrics to check the performance of the calibration algorithms. Finally, a first pass at RFI flagging is performed using the techniques discussed in section 3.3.4. Metadata from the observation is stored in the on-site Librarian database and data products are copied to NRAO for further processing. Daily Jupyter notebooks containing visualizations of the metrics from each stage in the RTS are stored on GitHub.	75
3.3	The modified Z-scores of each of the metrics described in section 3.3.2. The metrics are applied to a 10 minute window of data during observing (JD=2458140.35680). The mean visibility metric (top left) flags antennas 50, 68, and 104 since they are below our manually chosen threshold (dotted line) for flagging outliers. These antennas have very low amplitude compared to the others. However, the correlation metric (top right) additionally flags antennas 117 and 136, with antennas 98 and 68 on the border line for this integration. These antennas do not have amplitude correlations that are similar to others. In both cross polarization metrics in the bottom row indicate that the antenna 117 is actually cross polarized.	78
3.4	Antenna based metrics for a nights worth observation condensed into a single image. For a given observing night, HERA's antenna metrics are calculated on a 10 minute cadence (x-axis) and for each antenna-polarization (y-axis; for readability, only antenna numbers are shown with each tick representing the east polarization and the space between ticks represents the north polarization of the previous tick) it is determined whether the antenna is a good antenna (green), a dead antenna (red) according to the mean visibility amplitude metric (section 3.3.2) or the correlation metric (section 3.3.2), or a cross polarized antenna (blue) according to the cross polarization metrics (section 3.3.2). For this particular night (JD=2458140) we find that antennas 50, 68, and 104 are bad antennas. Antenna 117 is a cross polarized antenna (however it does get labelled as a dead antenna for part of the night). These are corroborated by the per file metrics shown in figure 3.3.	80
3.5	Per antenna delay solutions (top) and their fluctuation about the median delay (bottom) from a night of observation from HERA for calibration solutions derived from the FIRSTCAL algorithm. Only a quarter of the antennas and a single polarization are plotted here for clarity. The delays themselves appear to be fairly constant throughout the night with deviations below a nanosecond across the band. These delays redundantly calibrate the antennas to each other.	82

3.6	Complex visibilities for all baselines before (left) and after (right) redundant calibration when the galactic center is nearly at zenith. Each color and symbol pair denotes a different redundant baseline spacing and orientation. Before redundant calibration visibilities are uncalibrated and therefore have random phases with respect to one another and therefore their distribution is random in the real-imaginary plane. However, after calibration clusters of redundant baselines emerge and baselines that should be redundant begin to agree.	84
3.7	Histogram of the reduced χ^2 for the North (left) and East (right) polarizations for the OMNICAL run including all baseline separations. If OMNICAL perfectly calibrated the data we would end up with a χ^2 distribution (black) with mean of 1 and variance 2/DoF, where DoF is the degree of freedom for the fit which is equal to 647 for this example. Both histograms show evidence for a bimodal distribution indicating that there are two groups for the deviation from redundancy. The mean of the distribution is 1.49 (1.51) and the variance is .042 (.0405) for the East and North polarizations. The means greater than 1 indicate that there is considerable non-redundancy. This can be attributable to differences in baselines, antenna beam variations, and non-redundancies in the dish and feed structures.	88
3.8	Histogram of the reduced χ^2 for East (left) and North (right) polarizations from the OMNICAL run including all baseline separations except the shortest baselines (less than 15 meters). For this iteration of OMNICAL, DoF=567. We see that the bi-modality in the distributions seen in Figure 3.7 are no longer seen here indicating that the shortest baselines have more non-redundancy than the long baselines. However, the non long baselines (greater than 15 meters) still show evidence for non redundancy as indicated by the mean of the distributions being 1.31 (1.33) for the EE (NN) polarization. The variance for the EE (NN) histogram is .0017 (0.021).	91
3.9	Histogram of the reduced χ^2 for East polarization from the OMNICAL run including all baseline separations except the shortest baselines (less than 15 meters) for 8 different frequency ranges (10 MHz per range) across the 100-200 MHz band HERA observes, starting from 109.8 MHz to 187.9 MHz. The means for each 10 MHz band is greater than 1, indicating that there is non-redundancy across the band. However, the variances in the middle frequency ranges tend to be closer to the expected value of 2/DoF = 0.0031.	94
3.10	Fractional errors for four baseline lengths and orientations from the OMNICAL run without the shortest spacing baselines. Each curve in each subplot is a different instance of the baselines of length and orientation as indicated in the title (refer to Figure 3.1 for exact baselines). The shortest spacing baselines (14.7 meter baselines in the east-west and north-west directions) shown in the top two panels have the largest fractional errors, 25% to 50%, where the long baselines (bottom row) have the smallest errors hovering around 25%. Additionally, the structure in the fractional errors arises from the structure in the true visibilities solved for by OMNICAL.	95

- 3.11 RFI flags generated from the watershed algorithm described in section 3.3.4 applied to the raw visibilities (first row), **OMNICAL**visibilities (second row), **OMNICAL**chisq (third row), and **OMNICAL**gain solutions (4th row) for an instance to the shortest east west baseline (12,13). The total flag (bottom row) shows the OR of the flags all the preceding flags. The application of the watershed algorithm on each data product picks out different characteristics of RFI. 97

List of Tables

2.1 SIGNAL LOSS VERSUS ANALYSIS STAGE	54
---	----

Acknowledgments

Even though the research presented in this thesis was conducted and written by me, it is really the work of many people who have helped and supported me over my graduate career. It's inevitable that I will forget to mention somebody here, and for that, please accept my sincerest apologies.

First of all, without the PAPER and HERA projects, this thesis would not exist. PAPER was funded by the National Science Foundation (awards 0804508, 1129258, and 1125558) along with HERA (awards 1440343 and 1636646). I also want to thank SKA SA for dedicating resources for the construction and upkeep of both PAPER and HERA. I would like to particularly thank William Walburgh for organizing the PAPER deployments, Kathryn Rosie for being an excellent manager of all things HERA in South Africa, Andre Walker and Matthys Maree for on-site support in the Karoo, including making sure the air conditioners were always working, and all of the interns for making both PAPER and HERA successful instruments. Without them, none of this would be possible.

I would also like to thank the PAPER and HERA teams for all the support and knowledge they have shared with me. In particular, I would like to thank Adrian Liu, Jonathan Pober, Danny Jacobs, and Joshua Dillon for being excellent mentors and for answering my embarrassing questions. Additionally, I would also like to thank Matthew Dexter and David MacMahon for helping me get started when I joined RAL as a Junior Specialist some 8 years ago. I have learned so much about computers, FPGA's, analog systems, and saving my ass on deployments from them.

Of course, I wouldn't be here today without my advisor, Aaron Parsons. Thank you for taking a chance on a graduating senior with a curiosity for engineering and radio astronomy. I had no idea I would end up on this path. You have supported me over the past 8 years, especially during the rough times, and for that I'm grateful. I also want to thank my qualification and dissertation committee, Adrian Lee, Chung-Pei Ma, and Carl Heiles, for their constructive feedback.

For making sure I was paid and that all my forms were pushed through the wonderful UC system, thank you Dexter Stewart. You deserve so much recognition for everything you have done for all the graduates and undergrads in astronomy. We wouldn't be here without you. Have a spectacular retirement, you deserve it!

The friends I have made, both inside and outside of the department, over the past 6 years have been my support system. I wouldn't be as sane as I am without them. Thank you all.

I especially would like to thank Carina for always lending an ear when I needed someone to talk to and helping me celebrate my accomplishments. You have provided me so much support, both professionally and personally.

Finally, I would like to thank my parents. Thank you for being supportive of my life decisions and always letting me follow my own path, no matter how obscure. In addition to providing me with a solid foundation for who I am today, you are both lifelong role models. Dad, I wish you were here. I miss you.

Chapter 1

Introduction

1.1 Introduction

Our understanding of the evolution of the Universe has come a long way over the past century. From the discovery of the Cosmic Microwave Background (CMB) anisotropies that seed the distribution of matter in the Universe to the unexpected revelation that our Universe is accelerating, the past three decades have shown an exponential increase in our knowledge of the Universe we occupy. We can now trace the timeline of how our Universe evolved from its origins during the Big Bang to today's Universe, where discrete galaxies and the ionized Intergalactic Medium (IGM) locally dominate. Observations of our Universe at multiple wavelengths have provided us with the necessary data and evidence for our understanding of the Universe.

As of today, we have only observed a small fraction of the Universe, with the volume between $z \approx 1$ and $z \approx 1100$ mostly unobserved as shown in Figure 1.1. This volume can be broken up into four regimes. The first is the *Dark Ages*, which is characterized by the absence of luminous sources and lasted from $z \approx 1100 \sim 50$. Lasting from $z \approx 50 \sim 12$, the *Cosmic Dawn* is defined as the era when the first luminous sources turned on. The Epoch of Reionization (EoR), the third regime, is the last global phase transition in the Universe, when the IGM became fully ionized, and spans $z \approx 12 \sim 6$. Finally, once the IGM is fully ionized, the post-reionization era starts. This regime is characterized by complex structures and objects, as we see them today. All the regimes until post-reionization are largely unobserved, and can provide an abundance of information on galactic evolution. Specifically, we can begin to answer questions like how did the first stars and galaxies form? What kinds of radiation did they emit and what were their spectral distributions? How did they effect other objects and the surrounding baryons? What effect do they have on the CMB? This thesis provides the foundational observations and measurements needed in order to start answering some of these questions.

For the rest of this introduction, I will present a brief overview of the history of the Universe from the Big Bang to its current state, providing context for the observations and measurements of the Epoch of Reionization (EoR). I will then provide a brief review of 21 cm

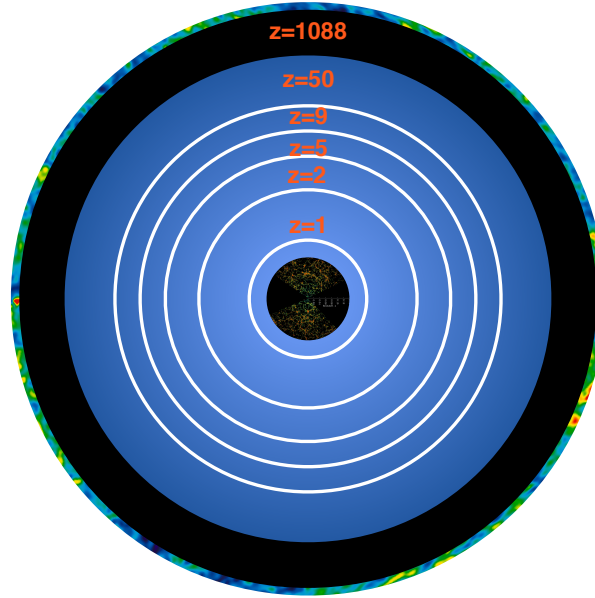


Figure 1.1: A schematic diagram of the observable Universe. The CMB is located at $z = 1088$. Between the surface of last scattering and $z \sim 200$ is a region where no observable radiation is emitted and hence remains unobservable. However, during these Dark Ages, the first signals after Recombination are theoretically observable between $z \sim 200$ and $z \sim 50$. The center of the diagram consists of the SDSS galaxy sample, which has systematically surveyed over 3 million objects (with spectra) over a large fraction of the sky up to $z \sim 1$. This is the extent of the locally observed Universe. A number of objects have been detected at redshifts greater than 1, but these are one off detections and does not provide a picture of the Universe as a whole. However, the blue region is a still yet unobserved volume of the Universe that will be able to be mapped out by highly red-shifted neutral hydrogen using the 21 cm spin flip transition (Mao et al. 2008).

physics, including how to observe the transition in cosmological observations and what we will expect to see. Finally, I will end with a description of the low frequency radio telescopes used to make these observations.

1.2 A Brief History

In this section, I briefly present our current understanding of the history of the Universe from the Big Bang in its hot dense state to today's local Universe comprised of bound, collapsed objects. This will put the work discussed in this thesis on 21 cm cosmology into context. I start by discussing the first 380,000 years in the Universe, and then the era of the formation of the first stars and galaxies known as the *Cosmic Dawn* ending with the Epoch of Reionization, and finally a brief discussion of the post reionization era.

1.2.1 From the Big Bang to Decoupling

The Big Bang Theory describes the expansion of the Universe from an extremely compact initial state. Although direct detection of this compact state is physically impossible, there is very strong evidence that this is how the Universe began. The detection of the expanding Universe ([Hubble 1929](#); [Riess et al. 1998](#); [Perlmutter et al. 1999](#)), agreement between Big Bang Nucleosynthesis (BBN) and the observed baryon composition, and finally the CMB ([Penzias & Wilson 1965](#); [Smoot et al. 1992](#); [Jarosik et al. 2011](#); [Bennett et al. 2013](#); [Planck Collaboration et al. 2014, 2015](#)), all provide strong evidence for the Big Bang.

During the Big Bang, a period of inflation ravaged space with exponential expansion and laid the seeds of over-densities in space through quantum fluctuations in space. These fluctuations would result in regions of over-densities and under-densities of matter in the Universe. After exponential expansion ended and gave way to adiabatic expansion under the influence of radiation and matter, the first light elements were formed through BBN. The relative abundances of hydrogen, helium, and (trace amounts of) lithium were set in this era and fueled the first stars a billion years later. See [Dodelson \(2003\)](#) for a full review of BBN.

Even though hydrogen (deuterium, helium, and lithium) nuclei were able to form in the early Universe shortly after conception, it was still far too hot for electrons to be captured by nuclei to form neutral atoms. Therefore, the Universe existed in a highly ionized state with radiation ionizing neutral atoms immediately after electron capture, setting up acoustic waves in the Universe known as Baryon Acoustic Oscillations ([Eisenstein et al. 2005](#); [Anderson et al. 2014](#)). It took another 380,000 years for the Universe to cool enough so that electrons could be captured by nuclei and radiation could free stream out as their mean free paths became nearly infinite due to the greatly reduced number of free electrons needed for Compton scattering. This marks the end of Recombination, when the Universe phase transitioned from an ionized state to a neutral state, and is the origin of the CMB radiation.

After matter-radiation equality ($z \approx 2700$), dark matter perturbations on all scales grew

linearly, and this continued after recombination. Specifically, the dark matter perturbations, defined as $\delta_{dm} = \frac{\rho_{dm} - \bar{\rho}_{dm}}{\bar{\rho}_{dm}}$, where ρ_{dm} is the density of dark matter in the Universe and $\bar{\rho}_{dm}$ is the mean density of dark matter, scale as $\delta_{dm} \propto a$, where a is the scale factor, for all modes within the horizon. Therefore, as the over-densities of dark matter grow, their ability to attract more baryonic matter grows, and as we'll discuss below, the first stars are able to form once the densities of the baryonic matter grew large enough. For completeness, prior to matter-radiation equality, over-densities grew logarithmically with the scale factor: $\delta_{dm} \propto \ln(a)$. In this regime, over-densities grew less drastically than in the matter dominated era due to the oscillating gravitational potential.

1.2.2 Post-Recombination

Once neutral hydrogen was able to form after recombination, an era known as the Dark Ages had begun. Starting at $z \approx 1100$ and ending sometime between $z \approx 50$, the Dark Ages are characterized by the lack of luminous sources. However, after the first luminous sources were able to form, an era known as the Cosmic Dawn commenced. This era is characterized by the formation of the first collapsed structures in the Universe (stars and galaxies). This era is rich with astrophysics and conjures new questions. How were the first stars formed? How massive were they? How did metals form? How did the first super massive black holes form? These types of questions can be answered through the direct detection of this era.

Once the first stars and galaxies formed, their UV radiation began to ionize the neutral hydrogen around them, ending in the EoR, when the Universe transitioned from completely neutral to ionized IGM. This unobserved period of the Universe occurred roughly between $z \sim 6 - 12$, and bridges the current day Universe, as we see it today, and the early Universe. Measurements of the EoR will provide a wealth of information about the evolution of galaxies and the Universe as a whole.

In order to observe the Cosmic Dawn, a number of major challenges must be overcome. The Hubble Space Telescope (HST) has detected galaxies as high as $z = 8$ and beyond in its Hubble Ultra Deep Field (HUDF) observation (Beckwith et al. 2006). These observations have provided us with the deepest images in the visible spectrum and give us a glimpse in to the properties of the first galaxies. Various techniques are used to detect the earliest galaxies using the HUDF, including making use of narrow band imaging (looking at highly red-shifted Lyman- α emission; e.g. Zheng et al. 2017b), looking for Lyman-break galaxies (e.g. Rhoads et al. 2013), using gravitational lensing, and looking for individual Gamma-Ray bursts. Each of these techniques have their benefits and trade offs, but they all provide powerful ways to detect the first galaxies.

In the near future, new instruments will be able to probe the UV emission from the first stars (which is red-shifted into the infrared). The James Webb Space Telescope¹ (JWST; Gardner et al. 2006) will be HST's successor, its 6.5 m aperture, and the fact that it is positioned at L2, outside of Earth's atmosphere, will be well suited for the detection of distant

¹www.jwst.nasa.gov

first galaxies. In addition to JWST, a number of ground based IR telescopes are on the horizon, including the European Extremely Large Telescope² (EELT), the Giant Magellan Telescope³ (GMT), and the Thirty Meter Telescope⁴ (TMT; [Skidmore et al. 2015](#)).

In contrast to optical telescopes looking for the redshifted Lyman- α transition, a complementary idea has been proposed to look at highly red-shifted 21 cm emission from neutral hydrogen. This complementary probe is looking for the neutral IGM surrounding the first stars and galaxies. The morphology of this gas can tell us a great deal about the first sources. We discuss the experiments used to detect this signal in Section 1.3.4.

Once the final phase transition in the Universe occurred, neutral hydrogen only existed inside the high density and self-shielded regions of galaxies ([Madau et al. 1997](#); [Barkana & Loeb 2007](#)). The IGM was, and remains, highly ionized. The Universe entered the peak of star formation between $z \sim 2 - 3$, where individual galaxies were producing stars using the left over and reprocessed neutral hydrogen and molecular gas as their fuel. During this time, a feedback balance stemming from outflows (supernova winds) and inflows (gas accretion) was critical in evolving galaxies to what we see today. The post reionization era is a complicated and erratic time for galaxies, and we are still trying to understand how they evolved into galaxies like our own Milky Way.

1.3 21 cm Cosmology

In this section, I discuss the fundamentals of 21 cm cosmology. 21 cm cosmology is a subfield of cosmology that uses the hyperfine transition of neutral hydrogen to observe the Universe on large scales. Specifically, observing the 21 cm transition allows the use of intensity mapping, where large angular regions on the sky are observed in aggregate. Additionally, because 21 cm radiation is a line transition, it can provide a tomographic mapping of the Universe, where different observing frequencies correspond to different redshifts. This method of observing the Universe provides a probe of large volumes of the Universe prior to and including the EoR, when the cosmos was dominated by neutral hydrogen. For a full review of 21 cm cosmology, see [Furlanetto et al. \(2006b\)](#) and [Pritchard & Loeb \(2012\)](#).

1.3.1 The Hyperfine Transition of Hydrogen

The 21 cm spin flip transition of hydrogen occurs when the proton and ground state electron change polarity from an aligned state to an anti-aligned state. This change causes a decrease in energy of $\Delta E = 5.874 \mu\text{eV}$ in the ground state of hydrogen, which gets emitted as a photon of wavelength

$$\lambda_0 = \frac{1}{\nu_0} c = \frac{h}{E} c = 21.106 \text{ cm.} \quad (1.1)$$

²www.eso.org/sci/facilities/eelt

³www.gmto.org

⁴www.tmt.org

The spin flip transition is a forbidden line transition, spontaneously emitting a 21 cm photon once every 10.934 million years. However, since the abundance of neutral hydrogen in the early Universe ($z \approx 12 - 1100$) is so high and the optical depth of this line is so low, 21 cm photons easily travel through the IGM and are theoretically easily observable during the Dark Ages and Cosmic Dawn. In practice, however, it is hard to measure the red-shifted 21 cm line due to other factors as we will see later in this section.

The hyperfine ground states in neutral hydrogen are usually parameterized using the Boltzmann distribution such that

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} e^{\frac{-T_\star}{T_s}}, \quad (1.2)$$

where n_0 and n_1 are the number densities of neutral hydrogen in the lower energy spin anti-aligned state and the higher energy spin aligned state, respectively, the g_i 's are the statistical degeneracy factors ($\frac{g_1}{g_0} = 3$ for the hyperfine ground state), $T_\star = .068 K$ is the temperature associated with the energy difference of the two states, and T_s is defined as the *spin temperature* and parameterizes the relative abundances of hydrogen in the excited and ground states. Using a line transition to observe the Universe has the benefit that different observing frequencies map to different redshifts in the Universe. 21 cm emission at redshift z will map to frequencies

$$\nu_{obs} = \frac{\nu_0}{1+z} = \frac{1420.405 MHz}{1+z}. \quad (1.3)$$

In this way, observers can map the evolution of neutral hydrogen as a function of time and position until it is fully ionized after the EoR. This tomographic mapping has the potential to provide 3-dimensional maps of neutral hydrogen in our early Universe.

Observation of the 21 cm signal requires looking at the differential brightness temperature, defined as

$$\delta T_b \equiv T_S(1 - e^{-\tau_{\nu_0}}) + T_R e^{-\tau_{\nu_0}}, \quad (1.4)$$

where T_S is the spin temperature, τ_{ν_0} is the optical depth of the 21 cm transition, and T_R is a background radiation source, usually taken to be the CMB (the dominant global emission in the Universe at the observation frequencies). After accounting for the expansion of the Universe, Taylor expanding the exponentials (since the optical depth of the 21 cm transition is small), and setting T_R to the CMB temperature, T_γ , the differential brightness temperature can be written as

$$\delta T_b \approx \frac{T_s - T_\gamma(z)}{1+z} \tau_{\nu_0} \quad (1.5)$$

$$\approx 9x_{HI}(1+\delta)(1+z)^{1/2} \left[1 - \frac{T_\gamma(z)}{T_s} \right] \left[\frac{H(z)/(1+z)}{dv_{\parallel}/dr_{\parallel}} \right] \text{ mK}, \quad (1.6)$$

where x_{HI} is the neutral fraction hydrogen, δ is the fractional over-density of hydrogen relative to the mean, T_γ is the background radiation source taken to be the CMB, T_s is the spin temperature, $H(z)$ is the Hubble constant, and $dv_{\parallel}/dr_{\parallel}$ is the velocity gradient along the line of sight. The last term encapsulates redshift space distortions relative to the expansion of the Universe. Redshift space distortions can introduce anisotropies on the order 10% (Jensen et al. 2013) to an otherwise isotropic signal, and are neglected in most analyses. For the rest of this section, I will discuss the evolution of the δT_b over time. For a thorough derivation of 21 cm line physics refer to chapter 12 of Loeb & Furlanetto (2013).

1.3.2 The Evolution of the 21 cm signal

The emission we see from the 21 cm line at high redshift is governed by Equation 1.5. There are two ways to look at the evolution of the signal as a function of redshift: a globally averaged signal or the spatially varying signal. I will first provide a brief description of the evolution of the globally averaged signal and then discuss the spatially varying signal.

The Global Signal

Each term in Equation 1.5 has a redshift (z) and spatial dependence ($\mathbf{r}$). If we average δT_b over all angles on the sky, we end up with the globally averaged signal as a function of z . Figure 1.2 shows how the globally averaged differential brightness temperature is expected to evolve as a function of time (bottom panel). The evolution of the globally averaged signal is primarily dominated by the ratio of the CMB temperature, T_γ , to the spin temperature, T_s , but the neutral fraction becomes important when reionization commences. The CMB temperature, T_γ , has a redshift dependence that goes as $T_\gamma \propto (1+z)$, due to redshifting of radiation as it travels through the expanding Universe. The spin temperature has a much more complicated evolution that depends on various coupling factors between the gas and the CMB, which primarily drive the evolution of δT_b . Without any heating source, the redshift dependence of the gas temperature goes as, $T_{gas} \propto (1+z)^2$, which can be derived by considering an monotonic adiabatically cooling gas in an expanding Universe.

We briefly describe the evolution of the globally averaged signal starting with recombination. After recombination, the spin temperature is coupled to the CMB temperature through Compton scattering, via residual free electrons. In this regime $T_\gamma = T_s$ and lasts from $z \approx 1100$ to $z \approx 200$, when the heating rate via Compton scattering from the decreasing fraction of free electrons becomes negligible compared to the cooling rate from the expansion of the Universe. There is no detectable signal during this period since $\delta T_b = 0$.

The first possible detection of 21 cm emission from the early Universe occurs after $z \approx 200$ (labeled as the Dark Ages in Figure 1.2). At this point the coupling from Compton scattering of T_s to T_γ becomes negligible and the gas temperature cools adiabatically ($T_{gas} \propto (1+z)^2$). Collisional coupling sets T_s to the gas temperature, making $T_s < T_\gamma$. Therefore, $\delta T_b < 0$, and we can observe the cosmological signal in absorption.

Collisional coupling starts to become ineffective around $z \approx 40$ due to the decreasing

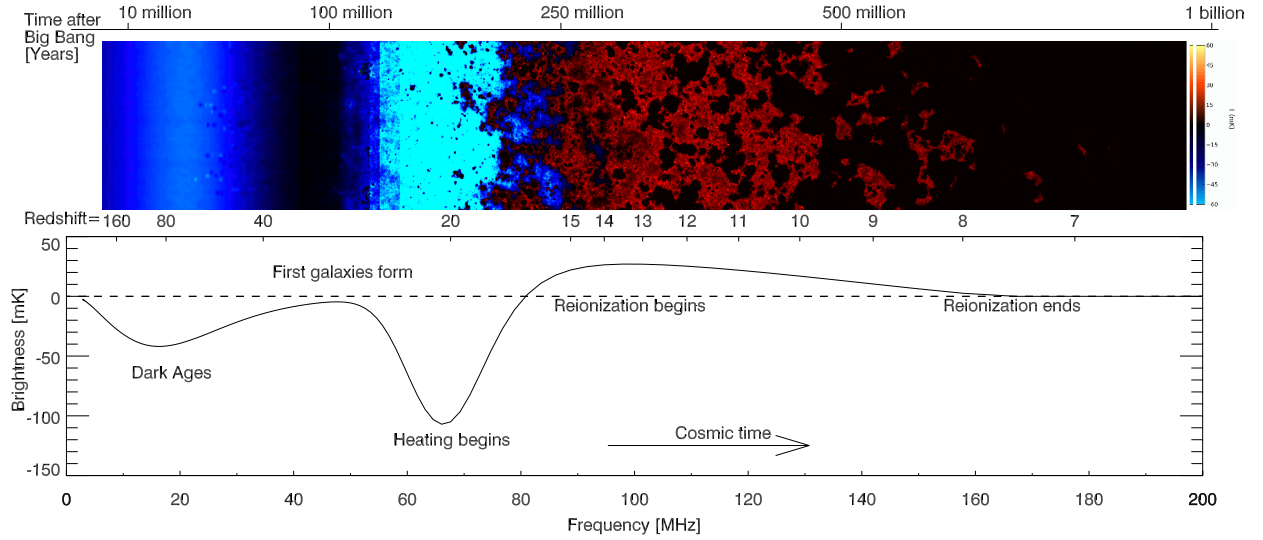


Figure 1.2: The evolution of the 21 cm hydrogen signal from a 21 cm simulation. *top:* The evolution of the spatial fluctuations in the differential brightness temperature as defined in Equation 1.5. The color indicates whether the signal is in absorption (blue), emission (red), or no signal (black). As a function of redshift, fluctuations in different fields are dominating the spatial fluctuations, such as the spin temperature or the neutral fraction. These spatial variations encode information on the nature of the first sources such as their spectral energy distributions and what type of sources they were (black holes, population III stars, etc...). *bottom:* The global evolution of the signal which is defined as the spatial average of the differential brightness temperature at a given redshift (black solid curve). The widely varying features in the global signal are due to different physical effects such as collisional coupling, the Wouthuysen-Field effect, heating from the first stars, and finally ionization of neutral hydrogen. The evolution is further described in Section 1.3.2. Figure used with permission from (Pritchard & Loeb 2012).

gas density ($n_{gas} \propto (1+z)^3$), making radiative coupling the dominant coupling mechanism driving T_s to T_γ . There is a brief period of no observable signal, because $\delta T_b = 0$, until the first sources turn on. Emitting Lyman- α photons, the first sources begin to couple T_s to T_{gas} via the Wouthuysen-Field effect (Field 1958), which mixes the ground state hyperfine levels in neutral hydrogen. This coupling eventually saturates, tightly binding the gas and spin temperatures. Initially, the gas temperature in this regime is still cooling adiabatically. Therefore $T_s < T_\gamma$ and δT_b is in absorption. The absolute value of δT_b is increasing. This era corresponds to the sharp decline in δT_b after the formation of the first galaxies in Figure 1.2.

As the first sources continue forming, UV and X-ray photons travel far into the IGM, heating the gas before being redshifted into Lyman-limit resonance and being absorbed. Because of this heating, the global temperature of the gas begins to increase. The spin temperature is still tightly coupled to the gas, and we begin to see a decrease in the absolute amplitude of δT_b . Specifically, the derivative of δT_b changes sign, but $T_s < T_\gamma$ still holds. The exact point when the gas temperature begins to increase and the rate at which it is increasing depends on the exact nature of the first sources and their spectral shapes at X-ray frequencies. In Figure 1.2, this regime is shown at $z \approx 20$ when heating begins.

In the most likely scenarios, the gas temperature increases until $T_s = T_{gas} > T_\gamma$, and we begin to see the 21 cm signal in emission ($\delta T_b > 0$). The differential brightness temperature saturates as T_{gas} continues to grow. After saturation, the dominant term in Equation 1.5 for affecting the global amplitude of the signal is the globally averaged neutral fraction. UV radiation from the present sources begin to ionize the neutral hydrogen around them, generating pockets of ionized regions. As more sources turned on, these ionized regions began to merge and coalesce, decreasing the neutral fraction, x_{HI} , until the Universe was fully ionized, culminating in the Epoch of Reionization.

Much of the exact locations and specifications of the evolution of the global signal are unknown and require observational evidence. Depending on the sources of heating, the absorption trough can be deeper, wider, and turnover at different redshifts (Mirocha et al. 2018, 2017; Fialkov et al. 2017; Fialkov & Loeb 2016; Fialkov & Barkana 2014). Additionally, depending on the nature of the sources heating the IGM, the neutral gas may be ionized before it is heated above the CMB temperature, making the emission signal undetectable (Fialkov et al. 2014a).

The Spatial Signal

In the previous section, I focussed on the global amplitude of the 21 cm signal as measured through δT_b , the differential brightness temperature. However, if we do not average over the entire sky, we can access a rich dataset that provides information on how the IGM is evolving spatially, and temporally. This extra axis provides us with further information about the nature of the first ionizing sources, giving us insight into what types of sources they were and if there was any exotic physics taking place. For example, we could constrain dark matter decay or annihilation by comparing the observed spatial distribution of power against

simulations (Furlanetto et al. 2006c). The top panel of Figure 1.2 shows a slice through space of the spatially varying cosmological signal as a function of redshift.

The measurement of the spatially varying signal is made using the 21 cm power spectrum defined such that

$$\langle \tilde{\delta}_{21}(\mathbf{k}_1) \tilde{\delta}_{21}(\mathbf{k}_2) \rangle = (2\pi)^3 \delta_D(\mathbf{k}_1 - \mathbf{k}_2) P_{21}(\mathbf{k}_1) \quad (1.7)$$

$$= (2\pi)^3 \delta_D(\mathbf{k}_1 - \mathbf{k}_2) \int e^{-i\mathbf{k}_1 \cdot \mathbf{r}} \xi(\mathbf{r}) d^3\mathbf{r}, \quad (1.8)$$

where $\tilde{\delta}_{21}(\mathbf{k}_1)$ is the Fourier dual to $\delta_{21}(\mathbf{r})$, the fractional perturbation to the differential brightness temperature, δ_D is the Dirac delta function, $P_{21}(\mathbf{k}_1)$ is the 21 cm power spectrum, and $\xi(\mathbf{r})$ is the real space correlation function. Therefore, $P_{21}(\mathbf{k}_1)$ is a measure of the variance of the spatial fluctuations on a certain length scale $\mathbf{k}_1$ and is directly related to the Fourier transform of the real space correlation function. This provides us with a picture of what scales are most important (largest power) at a given redshift. It is worthwhile to note that the power spectrum doesn't contain all the information of the spatially varying signal. Most 21 cm experiments measuring the cosmic dawn and EoR promise to make 3D imaged maps of the emission using 21 cm tomography, a technique that uses the fact that different observing frequencies measure different distances for making 3D image cubes. However, this endeavour requires precisely removing foregrounds, an increase in collecting area for experiments, and access to measurements on the finest scales, amongst other difficulties. In the near term, the power spectrum will access the spatial variance in the 21 cm signal.

The spatial signal probes different underlying distributions depending on the dominant components at the redshift of observation. The fluctuating signal is due to a number of parameters in equation 1.5, including the fractional over-density of matter, δ , the neutral fraction, x_{HI} , and the local variation in the spin temperature, T_s , which is usually broken down even further into the Lyman- α coupling coefficient, and the temperature of the gas, T_{gas} . Each of these parameters will dominate in different regimes during Cosmic Dawn.

During the first observable epoch during Cosmic Dawn, around $z \approx 200$, when collisional coupling drives a temperature difference between T_s and T_γ , the underlying dominant spatial field is the matter distribution. Regions of more matter have higher collisional coupling due to larger number densities leading to the 21 cm power spectrum tracing the matter power spectrum. Observing this period will provide a way to get at the initial distribution of matter in the Universe. Although the CMB does provide this through observations of anisotropies, the CMB radiation is damped due to free electrons after reionization and the signal we receive has been modified (Hu & White 1997). Probing $P_{21}(k)$ at these redshifts will provide a cleaner signal to corroborate the matter distribution.

As the first sources turned on in the largest dark matter halos, their Lyman- α radiation couples neighboring neutral gas to the gas temperature first; that is $T_s \rightarrow T_{gas}$ around the first sources. Most of the IGM at this point is still coupled to the CMB since Lyman- α coupling, via the Wouthuysen Field effect, has not yet saturated. Therefore the spatial variation in

the signal is driven by the first galaxies, which tend to be clustered. As a function of time (or redshift), the Wouthuysen Field effect (also known as Lyman- α pumping) comes in and out of importance and is dominant just prior to heating from X-rays, when δT_b is at a minimum. Concretely, this epoch is dominated by spatial variations in Lyman- α coupling (Wouthuysen Field effect).

As heating begins due to X-rays, Lyman- α coupling is expected to be strong everywhere and therefore spatial variations are negligible. The spin temperature is coupled to the gas temperature, which is at a minimum and therefore the IGM is cold. However, gas around the first X-ray sources have already been heated as stronger X-rays make their way further into the IGM. Therefore, $T_s > T_\gamma$ near the strongest X-ray sources (mini-quasars, for example) and the signal is in emission. The temperature fluctuations in the gas dominate the large amplitudes in $P_{21}(k)$ once the heating is underway.

As heating saturates, so that $T_s \gg T_\gamma$, the IGM has a roughly constant differential brightness temperature. The signal is in emission. Fluctuations in the spin temperature are now insignificant since δT_b has saturated, and the power spectrum once again follows the matter distribution. There is a period of low signal prior to the beginning of reionization.

As reionization begins, the dominant component in the spatial fluctuations comes from the neutral fraction, x_{HII} . The morphology of reionization, or the pattern of ionized bubbles, can provide information into what types of sources ionized them (Furlanetto et al. 2006a,d). Once reionization is in full force, power is expected to scale as k^{-3} in $P(\mathbf{k})$ (flat in $\Delta^2(\mathbf{k} \equiv \frac{P(\mathbf{k})k^3}{2\pi^2})$).

1.3.3 Challenges of 21 cm EoR Experiments

To reach the thermal noise level required to detect the 21 cm signal involves integrating on the sky with a modest low frequency radio telescope for a few hours. However, detecting neutral hydrogen from the early Universe remains one of the toughest challenges in cosmology today. This is due to sources of contamination that must be modeled and subtracted before accessing the cosmological signal. In this section we discuss some of the dominant sources of contamination and the potential ways to mitigate them.

Foregrounds

Foreground radio emission from sources between Earth and the cosmological signal at high redshift, is the biggest contaminant in 21 cm experiments. Synchrotron emission from our own Galaxy and extragalactic objects is the dominant source of low frequency emission for Cosmic Dawn experiments ⁵ (Santos et al. 2005; Ali et al. 2008; de Oliveira-Costa et al. 2008a; Jelić et al. 2008; Bernardi et al. 2009, 2010; Ghosh et al. 2011). Spiraling electrons

⁵Theoretical models suggest that the EoR occurs somewhere between $z \approx 6 - 12$. For 21 cm photons emitted at these redshifts, observing frequencies are between $\nu \approx 109 - 177$ MHz. If we open up observations to the Dark Ages, where the first theoretical observable signal comes from, the observing frequency range widens to $\nu \approx 50 - 177$ MHz.

in magnetic fields cause synchrotron emission, which is characterized by smooth and power law spectral energy distributions (Rybicki & Lightman 1979). For our own galaxy, which is dominated by synchrotron radiation, the brightness temperature follows the form:

$$T_{sync} = 180 \left(\frac{\nu}{180 \text{ MHz}} \right)^{-2.6} \text{ K}. \quad (1.9)$$

Galactic synchrotron emission is at least 4 orders of magnitude larger than the cosmological signal, which is of order 10 mK. This poses a very big problem for EoR experiments, requiring precision calibration and subtraction of these smooth spectrum foregrounds (Liu et al. 2010; Zheng et al. 2014; Patil et al. 2016; Eastwood et al. 2017).

A number of different ways have been suggested to remove or suppress foreground emission including, parametric fits (Wang et al. 2006; Bowman et al. 2006; Liu et al. 2009b,a), non-parametric fits (Harker et al. 2009; Chapman et al. 2013, 2012), principal component analysis (Paciga et al. 2011; Liu & Tegmark 2012; Masui et al. 2013), frequency filtering (Gleser et al. 2008; Petrovic & Oh 2011; Parsons & Backer 2009), and quadratic methods (Liu & Tegmark 2011; Dillon et al. 2013; Shaw et al. 2013) (in which case foregrounds can be down-weighted, but not directly modeled and subtracted). However, most of these methods boil down to using spectral differentiation of foregrounds and the cosmological signal. Foregrounds are spectrally smooth and the EoR signal is spectrally un-smooth. This distinction may allow the two signals to be separated. However, since foregrounds are several orders of magnitude brighter than the cosmological signal, we require subtraction with a precision of 1 part in 10^4 , requiring precise calibration and a deep understanding of the instrument.

The Wedge

Removing foregrounds will be necessary for making 21 cm maps. However, using the power spectrum to detect the cosmological signal requires a closer look at foregrounds. Consider the 2-dimensional power spectrum, where we can measure $P_{21}(\mathbf{k})$ as a function of $\mathbf{k}_\perp$, the wave number perpendicular to the line of sight, and $\mathbf{k}_\parallel$, the wave number parallel to the line of sight, $P_{21}(\mathbf{k}_\perp, \mathbf{k}_\parallel)$. In this domain, it can be shown that smooth spectrum foregrounds are localized to a region known as the “wedge”.

To see this, consider the measured frequency dependent visibility (Thompson et al. 2001) for a given baseline,

$$V_\nu(\mathbf{b}) = \int I_\nu(\hat{\mathbf{s}}) e^{2\pi i \frac{\mathbf{b} \cdot \hat{\mathbf{s}}}{\lambda}} d\Omega, \quad (1.10)$$

where I_ν is the sky intensity at a specific location on the sky, $\hat{\mathbf{s}}$, $\mathbf{b}$ is the baseline vector, λ is the wavelength, and the integral is taken over the whole sky. The Fourier transform of the baseline visibility (Parsons & Backer 2009), $V_\nu(\mathbf{b})$, is given by

$$\begin{aligned}
\tilde{V}_\tau(\mathbf{b}) &= \int V_\nu e^{-2\pi i \tau \nu} d\nu \\
&= \int I_\nu(\hat{\mathbf{s}}) e^{2\pi i \frac{\mathbf{b} \cdot \hat{\mathbf{s}}}{\lambda}} e^{-2\pi i \tau \nu} d\nu d\Omega,
\end{aligned} \tag{1.11}$$

where τ is the geometric delay of a source at position $\hat{\mathbf{s}}$ on the sky. Therefore, the maximum possible delay a flat spectrum source can have is determined by the length of the projection of the baseline vector on to the source vector. For a perfectly east-west baseline and a source at the baseline latitude, the maximum delay is given by the length of the baseline and occurs when the source is at the horizon. Longer baselines, will extend foreground emission to larger delays, and foreground emission on shorter baselines is contained to smaller delays. For sources with a power law spectrum, foreground emission will leak outside the horizon limits, determined by the convolution of Fourier transform of the power law (Parsons & Backer 2009). These convolution kernels are localized in delay space, maintaining a small footprint.

In the power spectrum domain, $\mathbf{k}$ -modes perpendicular to the line of sight are probed by the different baseline lengths. That is $\mathbf{k}_\perp \propto |\mathbf{b}|$. Delays, τ , are the Fourier dual to frequency, ν , which probes the line of sight direction. Therefore $\mathbf{k}_\parallel \propto \tau$. Using the fact that for smooth spectrum foregrounds, the maximum delay of a source is given by the length of the baselines, we find that there is linear relation for the extent of foreground emission. Namely, for smooth spectrum foregrounds, $\mathbf{k}_\parallel \propto \mathbf{k}_\perp$, creating a wedge of foreground emission (Datta et al. 2010; Parsons et al. 2012b; Vedantham et al. 2012; Hazelton et al. 2013; Liu et al. 2014a,b).

This important result implies that we do not need to remove foregrounds at all in order to detect the 21 cm power spectrum. We can follow a foreground avoidance strategy where we try to measure the 21 cm signal outside of the wedge. This is in contrast to the foreground removal strategies, which try to precisely remove foregrounds to work within the wedge; the benefit being that the 21 cm signal has a larger signal to noise in this regime.

The wedge has been detected in a number of experiments including the Murchison Wide-field Array (MWA) (Dillon et al. 2014) and PAPER (Pober et al. 2012). Both of these experiments use dipole antennas. As discussed in Thyagarajan et al. (2015), using dipole antennas, which inherently have a wide beams, the entire region within the wedge is contaminated. However, narrow field beams create a pitchfork structure, highly contaminating all $\mathbf{k}_\perp$ modes and low $\mathbf{k}_\parallel$ and $\mathbf{k}_\parallel$ modes near the horizon. However, as shown in Kohn et al. (2018), the narrow field beam of HERA (discussed below) does not seem to exhibit the pitchfork behavior, rather only the low $\mathbf{k}_\parallel$ modes seem to be contaminated. The wedge provides a powerful tool to separate foregrounds and the cosmological signal in the power spectrum. For a full derivation of the wedge see (Liu et al. 2014a).

Other Contaminants

Detecting the redshifted 21 cm emission from the early Universe requires using radio telescopes that observe from, roughly, 50 MHz to 200 MHz, right in the middle of the radio

communication frequencies on Earth. This leads to another challenge when trying to observe the cosmological signal: radio frequency interference (RFI). RFI sources include FM radio, radio communication, flight communications, satellite transmissions, and local electronics in the instrument, to name a few. As an escape from RFI, radio telescopes are usually built in the remotest areas of the planet including the Karoo Desert, ZA, Marion Island, ZA, and Western Australia. However, this is not enough to escape the extent of man made radio signals, driving the necessity to develop RFI detection algorithms ([Offringa 2012](#)). RFI manifests itself in data as many different shapes and forms, but is generally localized in frequency or time, and has very large amplitudes. However, the most problematic RFI are those below the noise floor. These RFI can potentially imitate the spectral nature of EoR in Fourier space and cause potential false detections and introduce systematics. RFI detection techniques use techniques such as derivatives, watershed algorithms, and even machine learning for detection.

I have described two major challenges that need to be overcome in order to detect the signature of the neutral hydrogen in the early Universe, however there are several other phenomenon that must be addressed before a detection could be made and proven. One of these contaminants include the ionosphere ([Yatawatta et al. 2008](#); [Datta et al. 2014](#); [Vedantham et al. 2014](#)), which acts as a refractor for low frequency radiation, causing sources to move on a timescale of a few seconds and requiring precision calibration to undo its effects. In a similar phenomenon, Faraday rotation of polarized emission from astrophysical sources can leak into the unpolarized EoR signal ([Moore et al. 2013](#); [Asad et al. 2015](#); [Moore et al. 2017](#); [Nunhokee et al. 2017](#)), generally known as polarization leakage. Faraday rotation effects the linear components of Stokes parameters such that the intrinsic stokes Q and U are modulated by a complex phasor that is dependent on λ^2 and the rotation measure (RM). Faraday rotation effects the frequency structure of line of sight Fourier modes used in estimating the power spectrum with EoR experiments. Depending on rotation measures (RM), this effect can contaminate modes used to measure EoR. Polarization leakage can cause low level contaminants that look like the EoR signal and may be one of the biggest challenges experiments will have to face in order to prove an EoR detection.

Both the ionosphere and polarization leakage through the interstellar medium (ISM) may pose a problem to the measurement of the 21 cm signal, as discussed above. However, as discussed in Chapter 2, observations with radio interferometers are usually averaged together in sidereal time to increase sensitivity to the cosmological signal. The authors of [Moore et al. \(2017\)](#) showed that the distribution of RM's at a given sidereal time over an observing season has a large variance, attenuating the polarized signal when observations over the season are averaged together. Therefore, there are ways to mitigate polarization leakage in observations of the 21 cm signal through the ionosphere and the ISM.

1.3.4 Experiments to Detect Cosmic Hydrogen

A number of dedicated 21 cm experiments are promising to detect neutral hydrogen gas during the Dark Ages and Cosmic Dawn. There are two classes of experiments: global signal

experiments, aimed at detecting the spatially averaged signal, and interferometers aimed at characterizing the spatial fluctuations in the cosmological signal. In some cases, these two experiment designs can overlap, such that the benefits of interferometers can be used for measuring the global signal (Presley et al. 2015).

Global signal experiments are trying to detect the global evolution of δT_b , aiming specifically to detect the Dark Ages, or the expected trough when Lyman- α coupling (Wouthuysen Field effect) has saturated and heating is just beginning. These experiments usually consist of a very well characterized single dual polarization dipole element, such as the Experiment to Detect the Global EoR Step (EDGES; Bowman & Rogers 2010), Shaped Antenna measurement of the background RAdio Spectrum (SARAS; Patra et al. 2015), SARAS 2 (Singh et al. 2018a,b), the Probing Radio Intensity at high-Z from Marion (PRIZM; Philip et al. 2018) experiment, and the Dark Ages Radio Explorer (DARE; Burns et al. 2012). Other global signal experiments like the Low Wavelength Array (LWA; Taylor et al. 2012), and the Owens Valley Array (Eastwood et al. 2017) use interferometers to precisely calibrate their instrument and then make a measurement of the spatially averaged signal using total power from the individual dipole antennas.

At the time of writing this thesis, EDGES has released a result that claims to have detected the global signal (Bowman et al. 2018). After calibration and foreground removal, they find an absorption feature centered at 78 MHz ($z = 17$) with an amplitude of 0.5 K. The remarkable consequence of this measurement is that the amplitude is far higher than expected and seems to disobey known physics. To see this consider Equation 1.5 and lets also hypothesize that T_s is coupled to the gas temperature since Compton scattering became negligible at $z \approx 200$. In the absence of heating, the gas adiabatically cools to $T_{gas} \approx 8$ K and the background temperature, T_γ , cools to ≈ 49 K at $z \approx 17$. Plugging into equation 1.5 we find that the absolute amplitude of the global signal should be $\delta T_b \approx 0.22$ K. This is the maximum possible absolute absorption amplitude according to our understanding of physics during the period. However, the detected signal by EDGES is a factor of 2 larger than the theoretical maximum absolute amplitude. This points to possible exotic physics going on in the Dark Ages and there have been numerous theoretical papers since the publishing of this result trying to explain the discrepancy. Everything from new insights into dark matter (Barkana et al. 2018; Clark et al. 2018; Liu & Slatyer 2018), dark energy (Costa et al. 2018), to a different radio background (Ewall-Wice et al. 2018) have been used to explain this difference. Additional work has also been done to characterize the nature of early galaxies using the timing of the trough (Mirocha & Furlanetto 2018). As the first potential detection, the result will need verification by other experiments in order to be a confirmed detection. The EDGES result has driven a lot of interest and hope into the 21 cm community and a confirmation of the result in the near future will provide a necessary milestone for the community.

In contrast to global signal experiments, low frequency radio interferometers are being used to measure the spatial fluctuations in δT_b as a function of redshift. These experiments aim to detect EoR by first measuring the power spectrum of the brightness temperature fluctuations and then making images of EoR as a function of redshift. Interferometers are

ideal for measuring the spatial fluctuations because they inherently access spatial modes, which depend on how one places antennas on the ground. Longer antenna separations correspond to smaller spatial scales and shorter antenna separations correspond to longer spatial scales.

Interferometers such as LOw Frequency Array (LOFAR; [van Haarlem et al. 2013a](#)), the Giant Meter-Wave Radio Telescope (GMRT; [Paciga et al. 2013](#)), the Murchison Widefield Array (MWA; [Tingay et al. 2013](#)), the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER; [Parsons et al. 2010](#)), and the Hydrogen Epoch of Reionization Array (HERA; [DeBoer et al. 2017](#)) are all trying to detect the EoR. LOFAR is a general purpose observatory based in Western Europe with tiles spread out around the continent and baselines extending up to hundreds of kilometers. However, long baselines are not suited for measuring the 21 cm power spectrum due to low SNR on the smallest scales. Therefore, a dense core with a large amount of collecting area provides enough sensitivity for the detection of EoR if a handle on precision foreground removal, calibration and RFI mitigation can be obtained. GMRT is another general purpose radio observatory just outside Pune, India. It consists of thirty 45-meter dishes that contribute to its high sensitivity. However, the RFI environment make EoR science a hard measurement to make. The MWA is the last of the general purpose low frequency instruments having a number of science drivers including making source catalogs and detecting space debris, in addition to detecting EoR. The large field of view of MWA dipoles make it a strong contender for detecting EoR, however issues with the digital signal chain and calibration currently make EoR science difficult. We discuss PAPER and HERA in more detail below, as they are a focus of this thesis.

All of the interferometers described above are using imaging techniques to subtract the several orders of magnitude of foreground signal required to make a detection of EoR. In effect, they are all trying to work within the wedge of foreground emission, making precise foreground removal even more important. Working within the wedge is critical for providing 3 dimensional maps of the EoR. In the near term, the use of foreground avoidance strategies could be enough to detect the 21 cm power spectrum, which is the first goal of EoR experiments and will provide critical progress to understanding the astrophysics driving EoR. The primary driver for these experiments is to build up sensitivity to certain spatial scales since images are not required (but they do help in calibration and learning about the instrument). This is the idea behind PAPER and HERA, both based in the Karoo Desert of South Africa at the SKA Radio Reserve. We discuss both of these instruments in the following sections as they are the instruments used in this thesis.

1.3.5 PAPER

PAPER was a dedicated experiment to detect the EoR. Initially conceived by Donald Backer at the University of California, Berkeley and Richard Bradley at the National Radio Astronomy Observatory, PAPER first deployed in Western Australia, near the MWA site, but soon transitioned to the Karoo Desert in South Africa. In addition to the science array in South Africa, a technical development array was also built in Green Bank, WV, a

useful second site for experimenting with different configurations and upgrades on the array. PAPER followed an incremental deployment, doubling the number of antennas every two years since its first deployment in 2008 (Parsons et al. 2010). This strategy allowed the incremental characterization of the array so that issues with the design could be addressed in the early stages before building all the elements at once. The PAPER experiment ended in 2015 after its final two seasons of observation with 128 antennas. Previous iterations consisted of 32 and 64 antennas in various configurations (Parsons et al. 2014; Jacobs et al. 2015; Ali et al. 2015).

The PAPER elements were comprised of sleeved dipole antennas as observing elements, each with their own ground screens. The dual polarization feeds consist of cross dipole copper tubes sensitive to 150 MHz sandwiched between two aluminum circular plates that increase the bandwidth to 100 - 200 MHz and produce a spatially and spectrally smooth beam profile. This was a critical design feature as it allows foregrounds to be cleanly separated from the cosmological signal. Following the feed, a differential dual polarization amplifier consisting of 28 dB of gain amplifies sky signal before travelling over 150 m, 75 Ω coaxial, F-type cables. A second stage of amplification (60 dB) and bandpass filtering (100 - 200 MHz) occur in receiver modules. Each antenna's signal chain ends with digitization using analog to digital converters (ADC) before they are all correlated in the digital correlator. The results were visibilities that instantaneously measure 100 - 200 MHz at a resolution of 97.65 kHz and time integrations averaged over 10.7 seconds.

The PAPER correlator went through multiple iterations using hardware and firmware based on work by the Collaboration for Astronomy Signal Processing and Electronics Research (CASPER; Parsons et al. 2006; Parsons et al. 2008; Hickish et al. 2016). The CASPER collaboration designs open source field programmable gate array (FPGA) hardware in addition to developing kernels for Graphics Processing Units (GPU) and control software for digital instrumentation, providing a flexible and quick development cycle for radio instrumentation. The PAPER correlator initially used interconnect Break out Boards (iBOB's) for channelizing voltage streams and Reconfigurable Open Architecture Computing Architecture (ROACH) boards for the X-engine, all connected through a 10 Gb switch. However, for PAPER-128, the move to using GPU's as the X-engines was made, leading to a heterogeneous system (using multiple computing architectures).

As the instrument with the least collecting area of all the interferometers mentioned earlier, PAPER decided to focus its observation on large spatial scales (small wave numbers), where the SNR of EoR is the largest. This strategy led PAPER to arrange its elements as close as possible and ultimately led to a grid configuration (see Figure 1.3). This configuration had the added benefit of building up sensitivity quickly by repeatedly measuring the same spatial Fourier modes (Parsons et al. 2012a). In this highly redundant configuration, repeated measurements of the sky could be made and combined prior to forming power spectra, therefore increasing sensitivity as N , the number of repeated baselines, rather than $\sqrt{N}$. This provided PAPER with similar sensitivity to other experiments, which have array configurations more conducive to forming images and cannot stack repeated measurements.



Figure 1.3: The Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER) in its 128 antenna redundant configuration in September 2013. Located in the Karoo Desert in South Africa on the SKA Radio Reserve, PAPER arranged its antennas in a grid configuration to boost sensitivity in the Power Spectrum due to its limited collecting area.

Arranging antennas in a grid like formation eliminated the production of quality images for subtraction of foregrounds. Instead, PAPER pioneered delay filtering, a technique which uses the localization of smooth spectral foregrounds to a maximal delay based on the length of a baseline to remove them (Parsons & Backer 2009). This key technique allowed the suppression of foreground emission (being agnostic to any particular source) to 1 part in 10^4 in temperature units (Parsons et al. 2014; Ali et al. 2015). See Chapter 2.3.3 for more detail. The two previously described techniques allowed PAPER to deliver upper limits on the 21 cm power spectrum by providing a way to access the cosmological signal in k -space and by delivering enough sensitivity to be competitive with other experiments.

1.3.6 HERA

First generation EoR experiments, such as GMRT, LOFAR, MWA, and PAPER, tried different ways to access the cosmological signal and learned many lessons along the way. These lessons and techniques would be used to build a new experiment to definitively detect and further characterize the EoR. This led to the development of HERA, the second generation EoR array. Three key lessons from early experiments were prioritized when designing HERA:

1. Sensitivity: Measuring the cosmological signal to high confidence requires a dramatic increase in telescope sensitivity. This translates to increasing overall collecting area. Rather than building thousands of dipoles, HERA's strategy is to build large parabolic dishes, which increases collecting area quickly and minimizes the cost-performance trade-off as discussed in DeBoer et al. (2017). Building dishes rather than dipoles has the effect of reducing the size of the beam (from roughly 60° to 10° FWHM), which in turn helps suppress foreground emission within the wedge (Parsons et al. 2012b; Vedantham et al. 2012; Thyagarajan et al. 2013, 2015, 2016).
2. Redundancy: A key development from the PAPER experiment was that redundancy is a powerful tool, both for increasing sensitivity and performing precision calibration. In this context, redundancy is defined as the instantaneously repeated measurement of the same spatial scale on the sky. HERA is therefore a highly redundant array with many repeated baselines of varying lengths. Additionally, HERA is also a close packed array, where antennas are placed as close as possible leading to a hexagonal configuration (Dillon & Parsons 2016). The short spacings are necessary in order to measure the lowest Fourier modes, where SNR to the cosmological signal is highest and spectral contamination by the instrument is the lowest.
3. Spectral Smoothness: As discussed previously, in order to separate foregrounds from the cosmological signal, the instrument needs to be spectrally smooth so as not to introduce spectral structure and contaminate the cosmological signal. Minimizing spectral structure includes suppressing reflections within the signal chain on scales that can contaminate EoR modes. This specification translated to suppressing reflections to be



Figure 1.4: The Hydrogen Epoch of Reionization Array (HERA) in the Summer of 2018. Over 100 dishes have been built with old original feeds.

~ -60 dB down at ~ 60 ns. To ensure the feed, dish, and signal chain are spectrally smooth, HERA is running EM simulations, testing antennas in the field, and iterating on its design (Neben et al. 2016; Ewall-Wice et al. 2016; Patra et al. 2018).

With these priorities in mind, HERA was designed to be a 350 element, close packed, and highly redundant interferometer built in the Karoo Desert, ZA at the SKA radio reserve. The core will consist of 330 close packed dishes split into 3 sectors each dithered by $1/3$ of a dish diameter (Dillon & Parsons 2016). There will also be 20 outriggers around the interior hexagon to increase sensitivity to the smallest spatial scales for imaging. At the time of writing this thesis, HERA has completed its second season of observing. The first season consisted of a 19 element array using modified dipoles from the PAPER instrument as feeds. The second season consisted of an updated feed design and partially updated signal chain consisting of 47 antennas. Currently there are over 100 antennas built and awaiting new feeds (Figure 1.4).

Each HERA element is a 14.7 m parabolic dish made with reflective wire mesh (at meter long wavelengths), 2x4 wooden posts, PVC pipes, and concrete. The feeds are modified Vivaldi antenna (Hall 1988), consisting of two orthogonal polarizations placed at the dish focal point of 4.7 m. They are sensitive from 50 - 250 MHz which allows for the detection of the Dark Ages and EoR. From the feed, analog signals travel over 30 m RF over fibre cables to Nodes where the analog signals are digitized using Smart Network ADC Processor (SNAP)⁶ boards. Once cross correlated data are stored, a real time processing system (RTS) begins reducing data. There are two purposes to HERA's RTS. The first is to characterize the array and the observing environment by detecting bad antennas (and quickly relaying that information to on site engineers), calibrating the array, and measuring the RFI environment. These activities occur during the commissioning stages of HERA. Once fully built, HERA

⁶<https://casper.berkeley.edu/wiki/SNAP>

will be producing ~ 500 TB of data a night. This leads to the second purpose of the RTS: a data reduction pipeline. Specifically, this will enable data compression and reduced data products for further analysis. For more details, see Chapter 3

Following in the footsteps of PAPER’s proven methodology for power spectrum estimation, HERA aims to initially measure the 21 cm power spectrum using the delay spectrum approach outlined in Parsons et al. (2012b) and demonstrated in Parsons et al. (2014); Ali et al. (2015). In parallel, HERA will also develop algorithms and techniques for measuring the power spectrum using an imaging based pipeline, as done by LOFAR and MWA. These techniques involve precisely modeling and subtracting foregrounds from low frequency source catalogs before making a power spectral measurements (Dillon et al. 2014; Beardsley et al. 2016; Patil et al. 2017). Additionally, they potentially allow working within the wedge of foreground emission where SNR for EoR is higher.

HERA aims to characterize the Dark Ages and the EoR as a function of time and space; first delivering the power spectrum and eventually promising 3D maps of EoR. With its large sensitivity, HERA will provide highly significant detections of EoR (23σ to 90σ depending on analysis strategy)(Pober et al. 2015). In conjunction with semi-analytical models, these measurements will allow for constraints on astrophysical parameters that describe the ionization history such as, the escape fraction of ionizing photons from their host galaxies, f_{esc} , the minimum virial temperature of halos producing ionizing photons, T_{vir} , and the mean free path of ionizing photons in HII regions, R_{mfp} (Greig & Mesinger 2015; Kern et al. 2017).

In addition to these primary science drivers, a number of secondary science objectives are also being pursued. HERA will be able to improve upon CMB constraints (Liu & Parsons 2016) by helping to constrain the optical depth parameter, τ , and the sum of the neutrino masses, $\sum m_\nu$ (Liu et al. 2016). HERA will also strive to provide opportunities for cross-correlating with other reionization experiments and tracers of EoR (Lidz et al. 2009; Silva et al. 2015; Vrbancic et al. 2016). These include, but are not limited to maps from CO, Lyman- α , and the CMB. Finally, HERA will also provide opportunities for detection of exoplanetary radio bursts due to highly polarized magnetosphere bursts around exoplanets and provide follow up measurements of fast radio bursts.

1.4 Outline of Thesis

This thesis is outlined as follows: In Chapter 2, I describe the PAPER-64 dataset and the analysis pipeline used to reduce data to a power spectrum. The analysis techniques described include, redundant calibration, delay filtering, fringe-rate filtering, and optimal quadratic estimators using empirical covariances. Chapter 3 switches gears to the development of a real time system (RTS) for the HERA experiment. I argue for the use of an RTS in astronomy in general, and provide a methodology for the development of an RTS for a new low frequency, highly redundant array, like HERA. I discuss the RTS of HERA along with providing an assessment of the first season of HERA data. Finally, I conclude in Chapter 4.

Chapter 2

Results from the Donald C. Backer Precision Array for Probing the Epoch of Reionization

In this chapter, we discuss limits on the 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 (PAPER) in South Africa. This work extends the analysis techniques presented in [Parsons et al. \(2014\)](#) with improved redundancy-based calibration, improved fringe-rate filtering, and an updated power-spectral analysis using optimal quadratic estimators. We present the effects of these new analysis techniques on the power spectrum and develop a formalism for putting limits on the spin temperature given a 21 cm power spectrum. The limits initially published in this Chapter have been revised due to significant signal loss through the power spectrum pipeline. The details of this revision are discussed in Section 2.6 of this Chapter. We conclude with a discussion of implications for future 21 cm reionization experiments, including the newly funded Hydrogen Epoch of Reionization Array.

2.1 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 Ly α by the neutral IGM toward distant quasars (the ‘Gunn-Peterson’ effect) (Fan et al. 2006), the demographics of Ly α emitting galaxies (Schenker et al. 2013; Treu et al. 2013; Faisst et al. 2014), and the Ly α 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., JWST 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. 2006b; 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. 2013b), 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 21cm reionization signal. This approach focuses on spectral smoothness as the primary discriminant between foreground emission and the 21cm 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 $\Delta^2(k) \leq (41 \text{ mK})^2$ at $k = 0.27h \text{ Mpc}^{-1}$ ([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 Chapter, 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 $(22.4 \text{ mK})^2$ in the range $0.15 < k < 0.5h \text{ Mpc}^{-1}$ at $z = 8.4$. This result places constraints on the spin temperature of the IGM, and as is shown in a forthcoming paper, [Pober et al. \(2015\)](#), this supports and extends previous evidence against extremely cold reionization scenarios. In Section 2.2 we describe the observations used in this analysis. In Sections 2.3 and 2.4, 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 2.5. We present our results in Section 2.7 along with new constraints on the 21cm power spectrum. We discuss these results in Section 2.8 and conclude in Section 2.9.

2.2 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

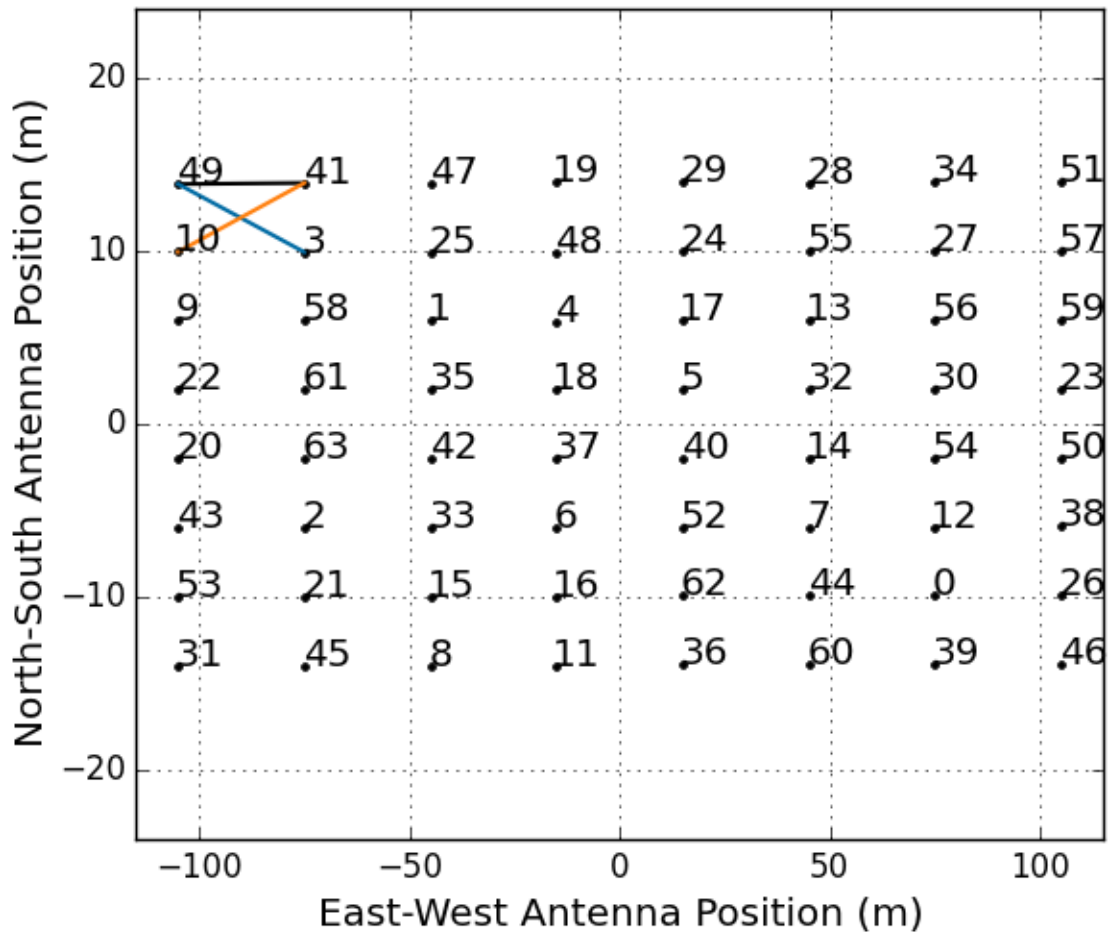


Figure 2.1: Antenna position within the PAPER-64 array. This analysis only makes use of east-west baselines between adjacent columns that have row separations of zero (black; e.g. 49-41, 41-47, 10-3, ...) one in the northward direction (orange; e.g. 10-41, 3-47, 9-3, ...) or one in the southward direction (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.

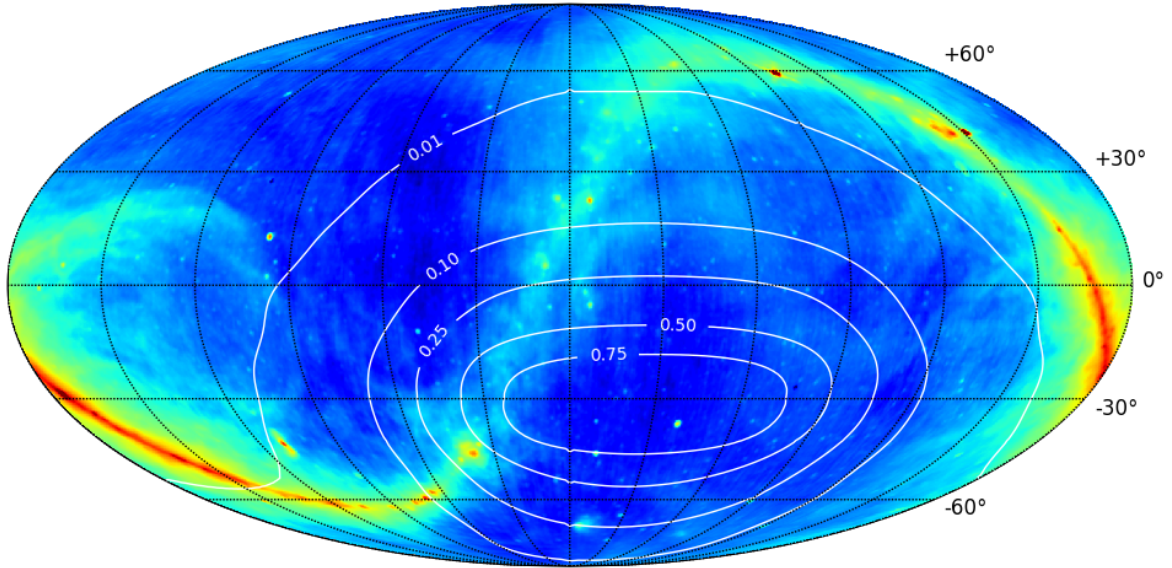


Figure 2.2: The Global Sky Model (de Oliveira-Costa et al. 2008a), illustrating foregrounds to the 21cm cosmological signal, with contours indicating beam-weighted observing time (relative to peak) for the PAPER observations described in Section 2.2. The map is centered at 6:00 hours in RA.

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 2.1, 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 2.1; hereafter referred to as *fiducial baselines*), 30-m east-west baselines whose eastern element is staggered one row up (e.g. 10-41, orange in Figure 2.1), and those whose eastern element is one row down (e.g. 49-3, blue in Figure 2.1). 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

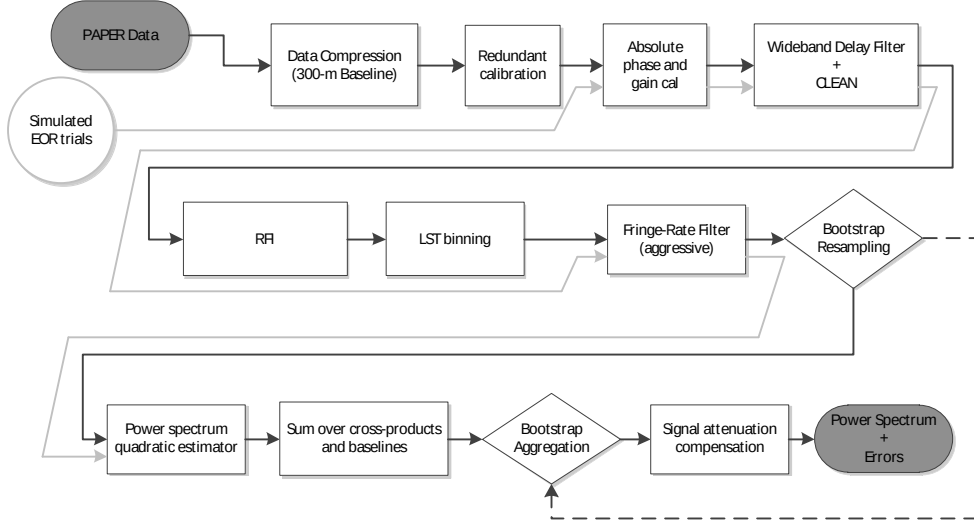


Figure 2.3: The stages of power-spectrum analysis. Black lines indicate data flow; red lines indicate Monte Carlo simulations used to measure signal loss. Yellow boxes indicate stages that by construction have negligible signal loss. Signal loss in other stages is tabulated in Table 2.1.

between 0:00 and 8:30 hours 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 2.2 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 CASPER¹ 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.

¹<http://casper.berkeley.edu>

2.3 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. 2008a; 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 Chapter 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,b), and has been confirmed in simulations (Datta et al. 2010; Hazelton et al. 2013) and observationally (Pofer 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 2.3, 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 $|\tau| \lesssim 1\mu\text{s}$ and fringe-rate scales of $f \lesssim 23\text{ mHz}$, 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 $-0.5 \leq k_{\parallel} \leq 0.5h\text{ Mpc}^{-1}$ 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 2.3.1) uses instantaneous redundancy to solve for the majority of the per-antenna internal dof in the array. In the second stage (Section 2.3.2), 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 2.3.3) and additional RFI flagging and crosstalk removal, we average the data in LST (Section 2.3.4) and apply a fringe-rate filter (Section 2.3.5) to combine time-domain data. Finally, we use an OQE (Section 2.5) to make our estimate of the 21 cm power spectrum.

2.3.1 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 (P14; P12a; 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,

$$v_{ij} = g_i^* g_j y_{i-j} + n_{ij}^{res}, \quad (2.1)$$

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

$$\log v_{ij} = \log g_i^* + \log g_j + \log y_{i-j}, \quad (2.2)$$

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 biased by the asymmetric behavior of additive noise in the logarithm (Liu et al. 2010).

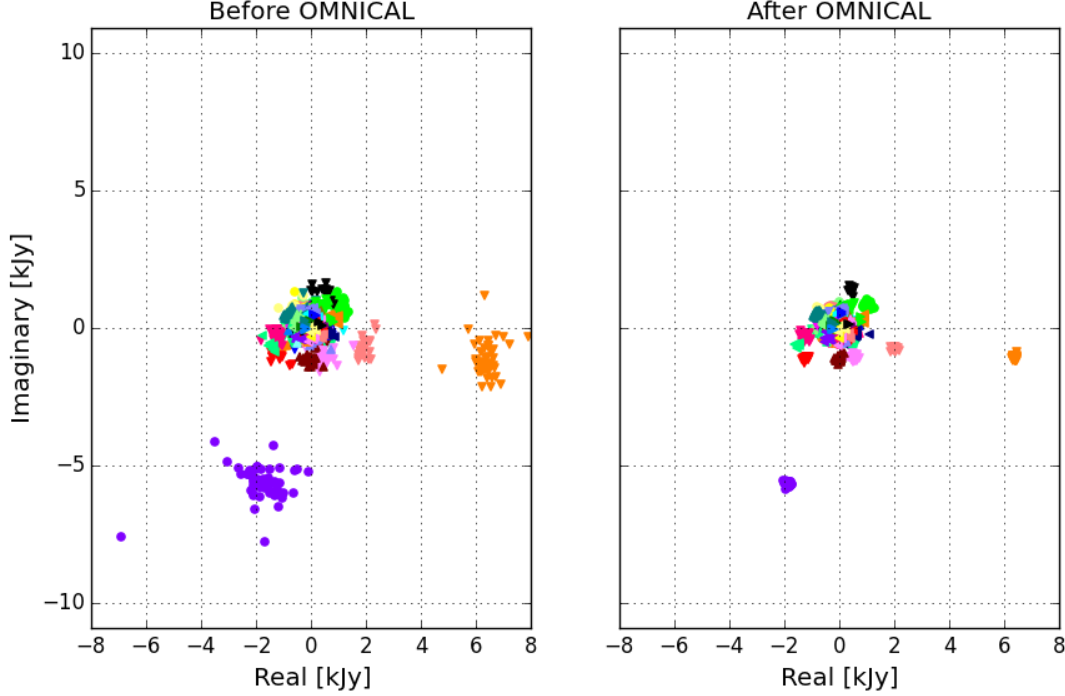


Figure 2.4: PAPER visibilities plotted in the complex plane before (left) and after (right) the application of the improved redundancy-based calibration with OMNICAL (Zheng et al. 2014). All baselines in the array measured at 159 MHz for a single time integration are plotted. Instantaneously redundant baselines are assigned the same symbol/color. The tighter clustering of redundant measurements with OMNICAL indicates improved calibration.

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

$$v_{ij} = g_i^{0*} g_j^0 y_{i-j}^0 + g_i^{1*} g_j^0 y_{i-j}^0 + g_i^{0*} g_j^1 y_{i-j}^0 + g_i^{0*} g_j^0 y_{i-j}^1, \quad (2.3)$$

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 OMNICAL² — 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 3.6.

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.

²<https://github.com/jeffzhen/omnical>

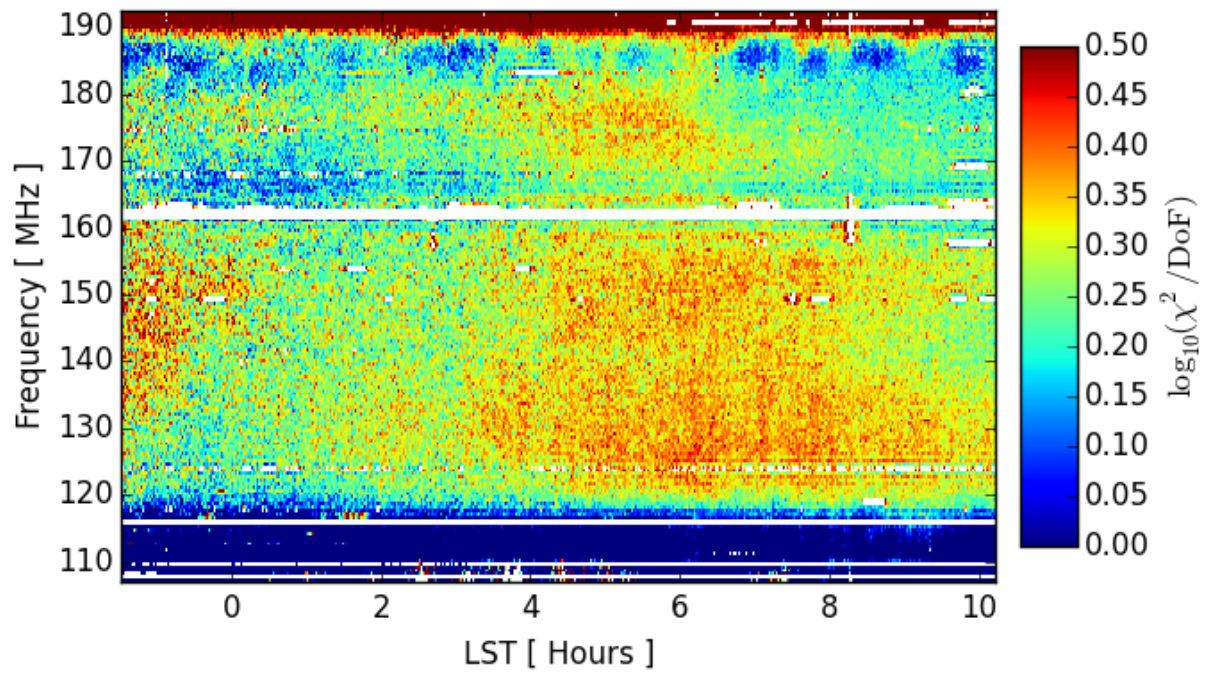


Figure 2.5: Log of χ^2 per degree of freedom of all baseline residuals after the application of OMNICAL. The plot comprises a observations over one day, with a frequency resolution of 493 kHz and a time resolution of 42.9 s.

(2014),

$$\chi^2 = \sum_{ij} \frac{|v_{ij} - y_{i-j} g_i^* g_j|^2}{\sigma_{ij}^2}, \quad (2.4)$$

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

$$\begin{aligned} \text{dof} &= N_{\text{measurements}} - N_{\text{parameters}} \\ &= 2N_{\text{baselines}} - 2(N_{\text{antennas}} + N_{\text{unique baselines}}), \end{aligned} \quad (2.5)$$

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 $\sigma^2 = 2/\text{dof}$. 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 2.5, 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 2.5, crosstalk is a defect in the 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.

2.3.2 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 2.6 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. \(2013\)](#),

$$S_\nu = S_{150} \times \left(\frac{\nu}{150 \text{ MHz}} \right)^\alpha, \quad (2.6)$$

where $S_{150} = 381.88 \text{ Jy} \pm 5.36$ 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 2.3.3, for the hour before and after the transit of Pictor A. We image a $30^\circ \times 30^\circ$ 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 CASA³ gives similar performances as the PAPER array is essentially instantaneously coplanar. A source spectrum is derived for each snapshot by fitting a two-dimensional Gaussian to Pictor A by using the PyBDSM⁴ 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 2.7 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 of 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 $2 \times 10^{-7}\%$. We apply the latter correction factor for all modes outside the width of the delay filter to the final power spectrum.

2.3.3 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

³<http://casa.nrao.edu>

⁴http://www.lofar.org/wiki/doku.php?id=public:user_software:pybdsm

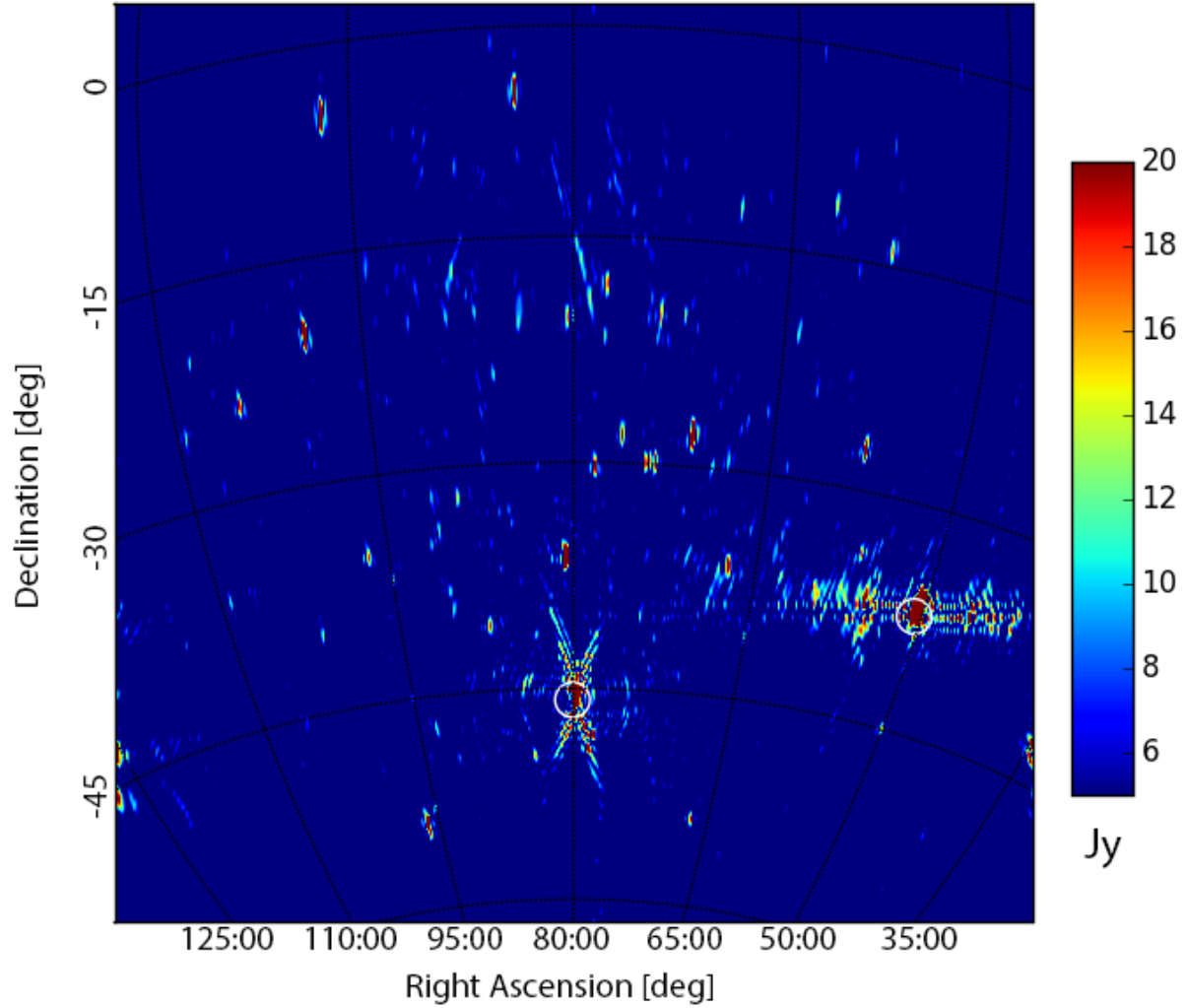


Figure 2.6: PAPER-64 image of a field including Pictor A and Fornax A, with white circles indicating catalog positions (Jacobs et al. 2011). Image was synthesized with two hours of visibilities while Pictor A was in transit and 53 MHz of instantaneous bandwidth from 120 to 173 MHz. Image quality is limited 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.

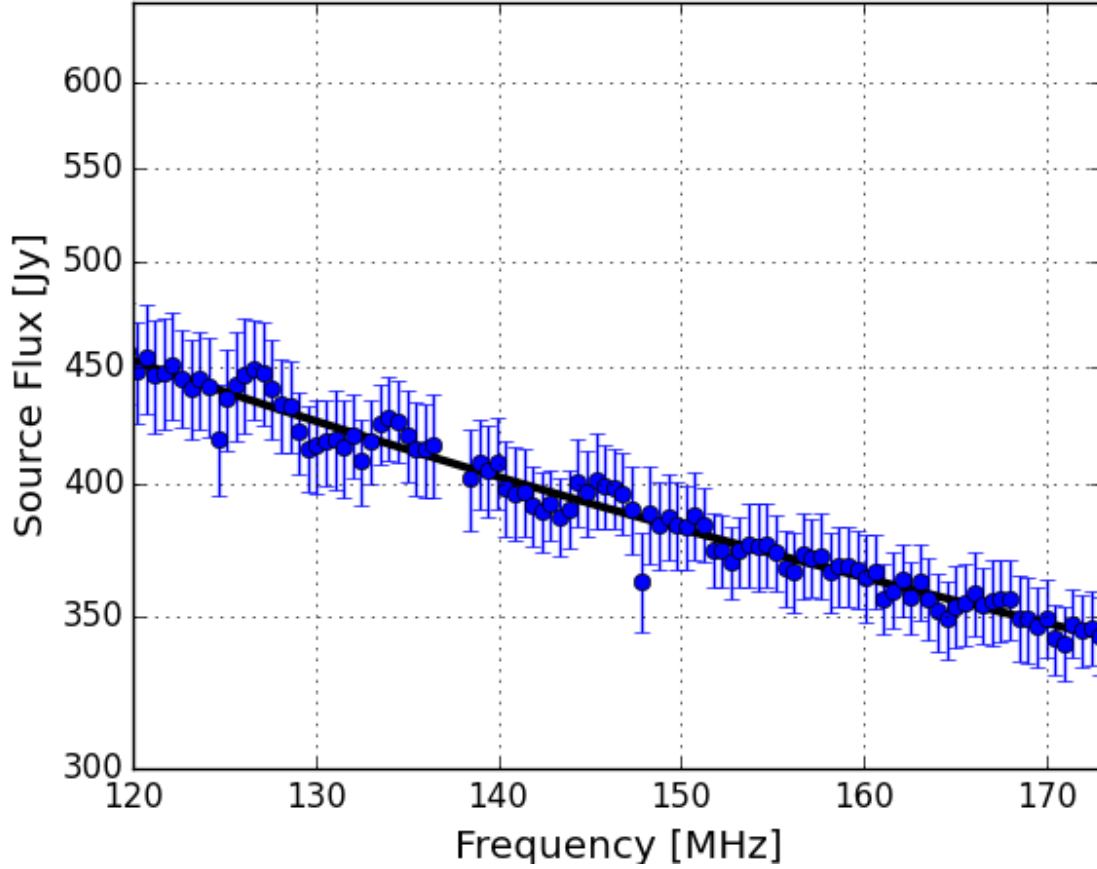


Figure 2.7: Measured spectrum of Pictor A in Stokes I (blue) relative to its catalog value (black; [Jacobs et al. 2013](#)). Flux density measurements are extracted from images of Pictor A, made independently for each frequency channel in 10 minutes snapshots as Pictor transits between hour angles of -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.

estimated as

$$V_I = \frac{1}{2}(V_{XX} + V_{YY}), \quad (2.7)$$

where V_{XX} and V_{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 (2.7). 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 into 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 into 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 ([Paciga et al. 2011](#); [Liu & Tegmark 2011](#); [Paciga et al. 2013](#); [Masui 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,b](#)), 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

$$\begin{aligned} \tilde{V}_\tau &= \int W_\nu A_\nu I_\nu e^{-2\pi i \tau g} \cdot e^{2\pi i \tau \nu} d\nu \\ &= \tilde{W}_\tau * \tilde{A}_\tau * \tilde{I}_\tau * \delta(\tau_g - \tau), \end{aligned} \quad (2.8)$$

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 (2.8). 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

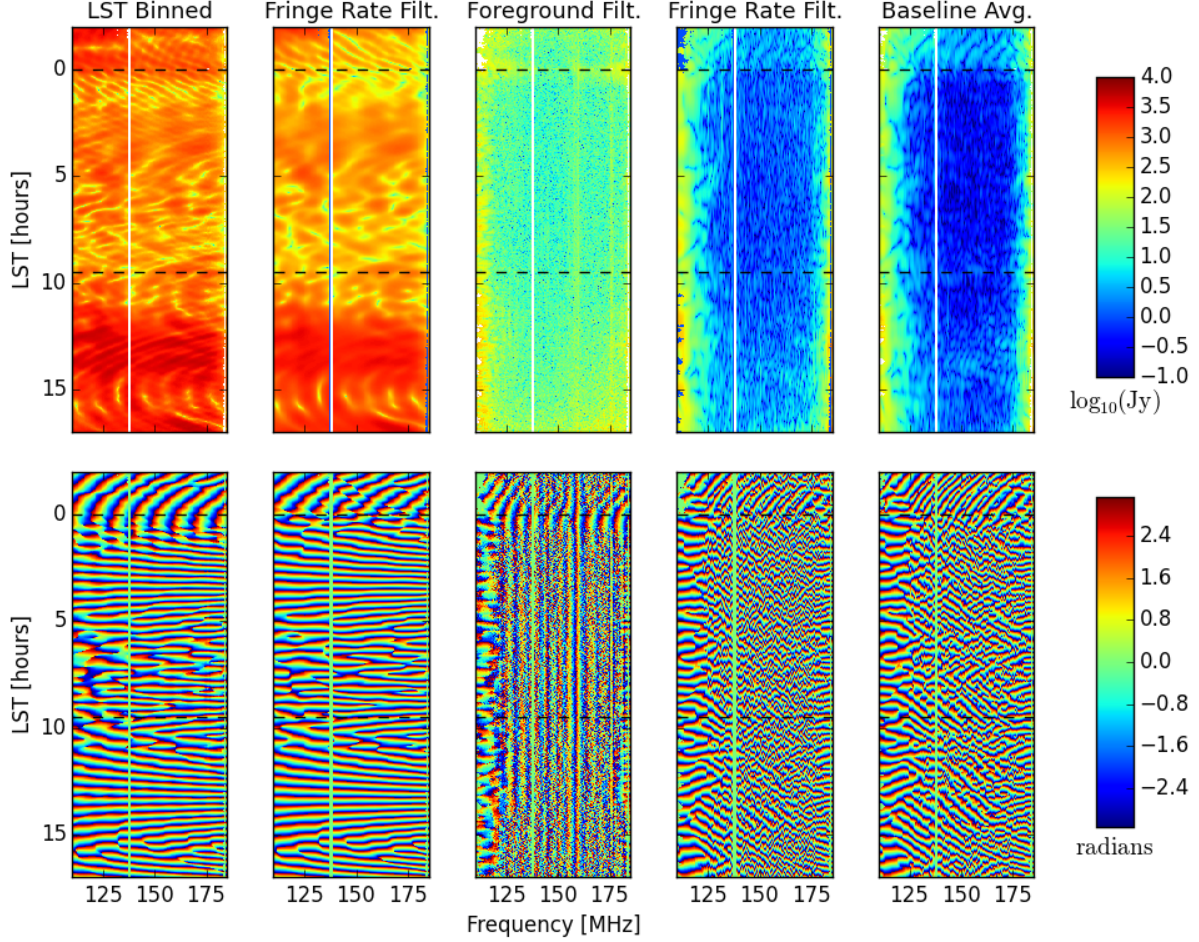


Figure 2.8: Visibilities measured by a fiducial baseline in the PAPER-64 array, averaged over 135 days of observation. From left to right, columns represent data that: (1) contain foregrounds prior to the application of a wideband delay filter or fringe-rate filtering, (2) are fringe-rate filtered but 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.

suppression, as seen in Figure 2.8. 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.

2.3.4 Binning in LST

After the wideband delay filter, we remove a second layer of RFI which 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” dataset. 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 estimated empirically through bootstrapping (see Section 2.5.4), the slight increase in measurement error associated with truncating the tails of the Gaussian distribution are naturally accounted for.

2.3.5 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 root-mean-squared

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 response⁵, $A(l, m)$ ([Parsons et al. 2015](#)). By utilizing prior 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_{fr} , the effective integration time can be calculated by comparing the variance statistic before and after filtering:

$$t_{\text{int,after}} = t_{\text{int,before}} \frac{\int \sigma_f^2 df}{\int \sigma_f^2 f_{\text{fr}}^2 df}, \quad (2.9)$$

where $t_{\text{int,before}}$ 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 (2.9), 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 2.5.3, 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 2.9, 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 2.1, 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

⁵The angular area in Equation (2.24) will reflect the new angular area corresponding to the change in beam area.

as crosstalk. We convolve the time-domain visibilities with the Fourier transform of the resulting fringe-rate filter, shown in Figure 2.9, to produce an averaged visibility. The effect on the data can be seen in Figure 2.8.

2.4 Instrumental Performance

2.4.1 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 2.10 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 $T_{\text{sys}} \sim 500$ K. We get more samples per LST bin in the range of .5–2.5 radians 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 2.11, as time scales decrease (temporal frequency increase), 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.

2.4.2 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.

$$T_{\text{rms}} = T_{\text{sys}} / \sqrt{2\Delta\nu t}, \quad (2.10)$$

where $\Delta\nu$ is the bandwidth, t is the integration time, and T_{rms} is the RMS temperature, or the variance statistic described above. Figure 2.12 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 $T_{\text{sys}} = 560$ K, at $T_{\text{sys}} = 500$ K at 160 MHz. However, this estimate

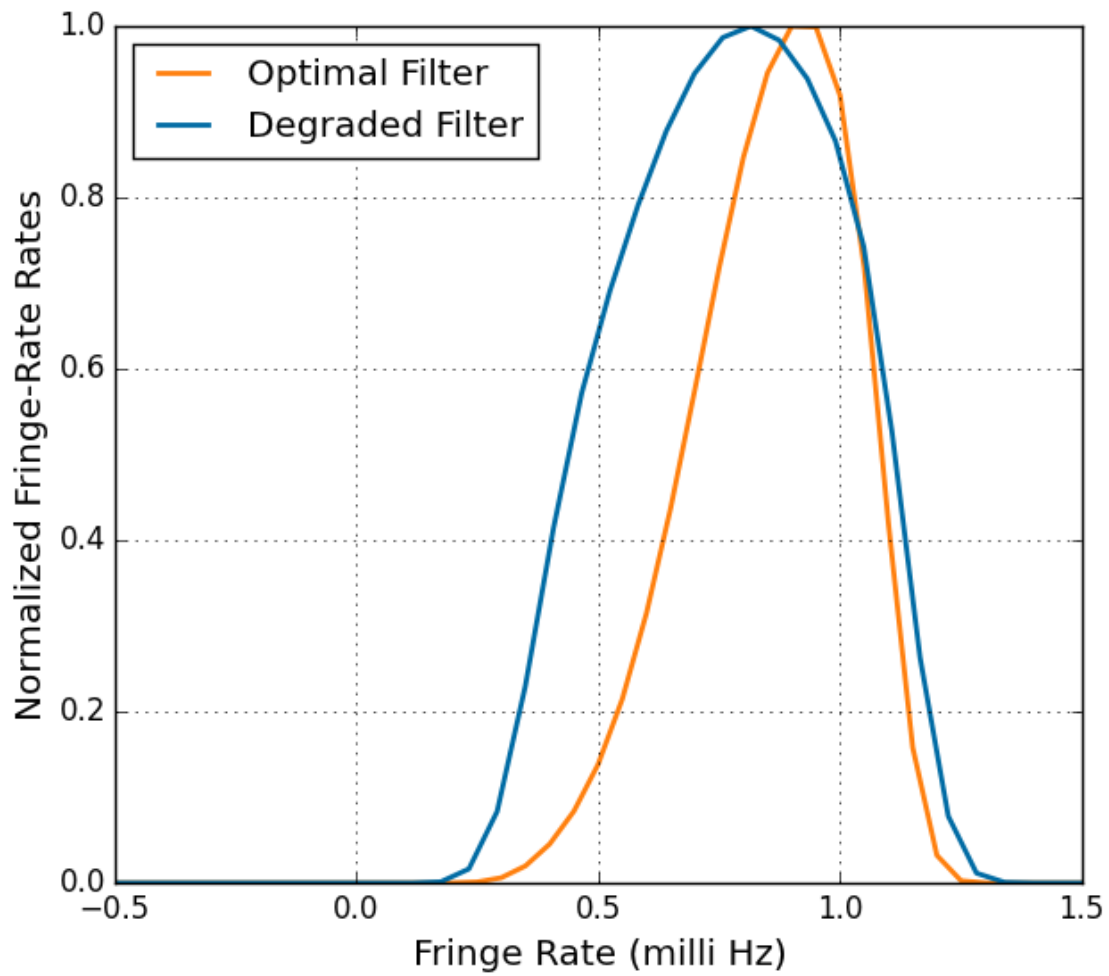


Figure 2.9: The optimal fringe-rate filter (orange) that and the degraded fringe-rate filter (blue) actually used in the analysis at 151 MHz, normalized to peak at unity.

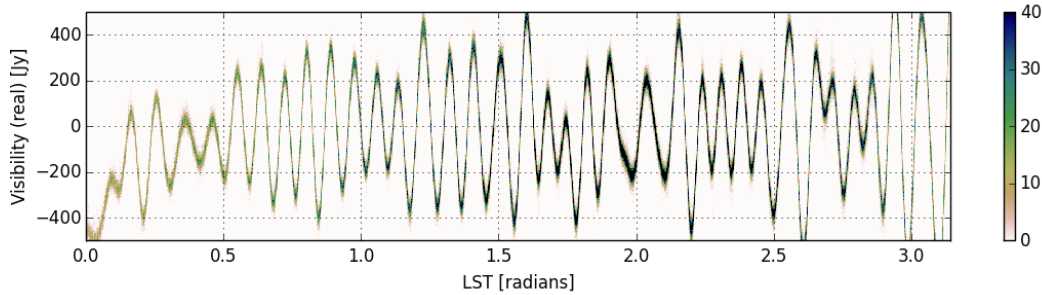


Figure 2.10: Histogram of the real component of all calibrated visibilities measured over 135 days with every redundant instance of the fiducial baseline at 150 MHz. Color scale indicates the number of samples falling in an LST/flux-density bin. This plot serves to illustrate the stability of the PAPER instrument and the precision of calibration. The temporal stability of a single LST bin over multiple days is shown in Figure 2.11.

is more consistent with the results derived in (Moore et al. 2015), where $T_{\text{sys}} = 505 \text{ K}$ 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 2.2.

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.

2.5 Power Spectrum Analysis

In this section we first review the optimal quadratic estimator (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.

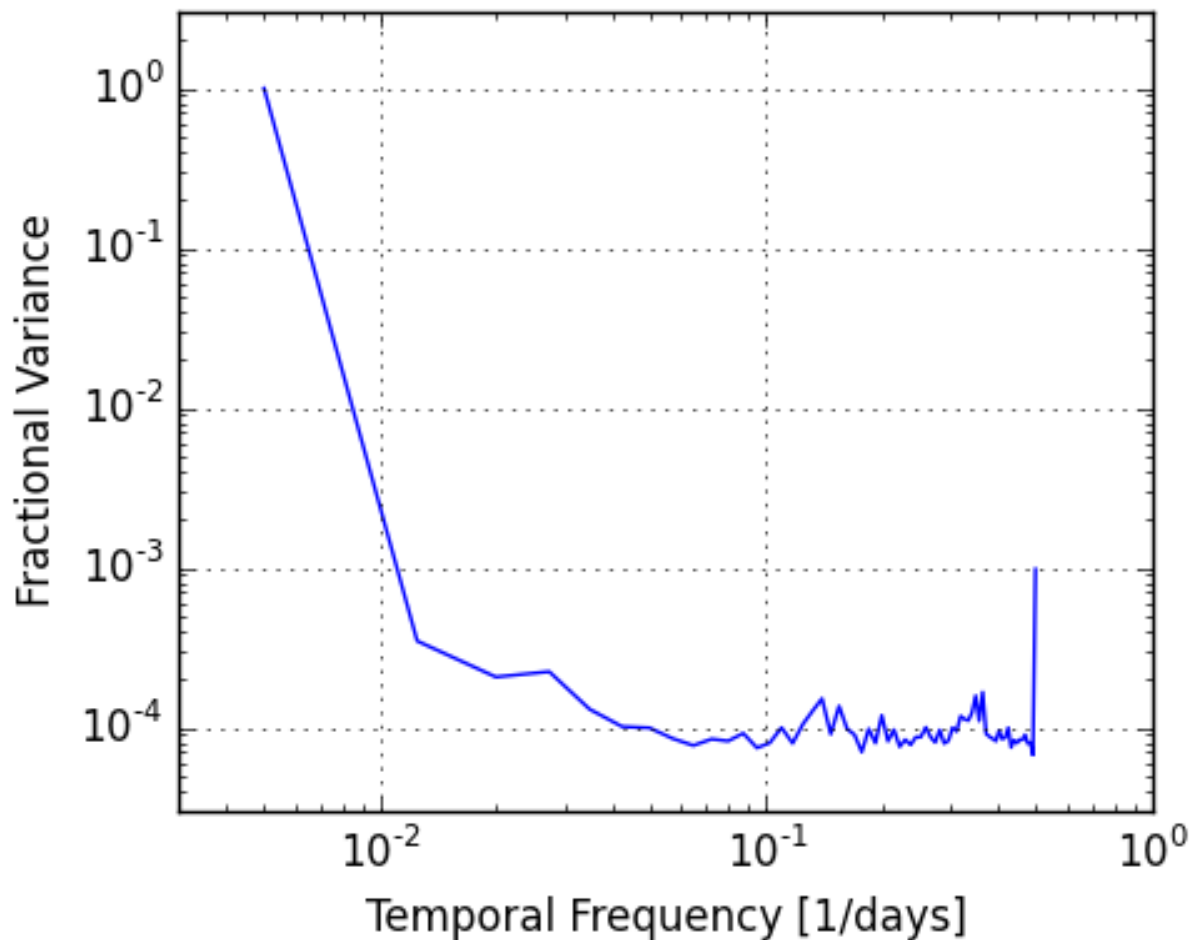


Figure 2.11: Power spectrum of 135 days of time-series data contributing to a single LST bin, illustrating the stability of measurements over the observing campaign. Relative to the average value, variation in the measured value across days (quantified by variance as a function of time period) is orders 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.

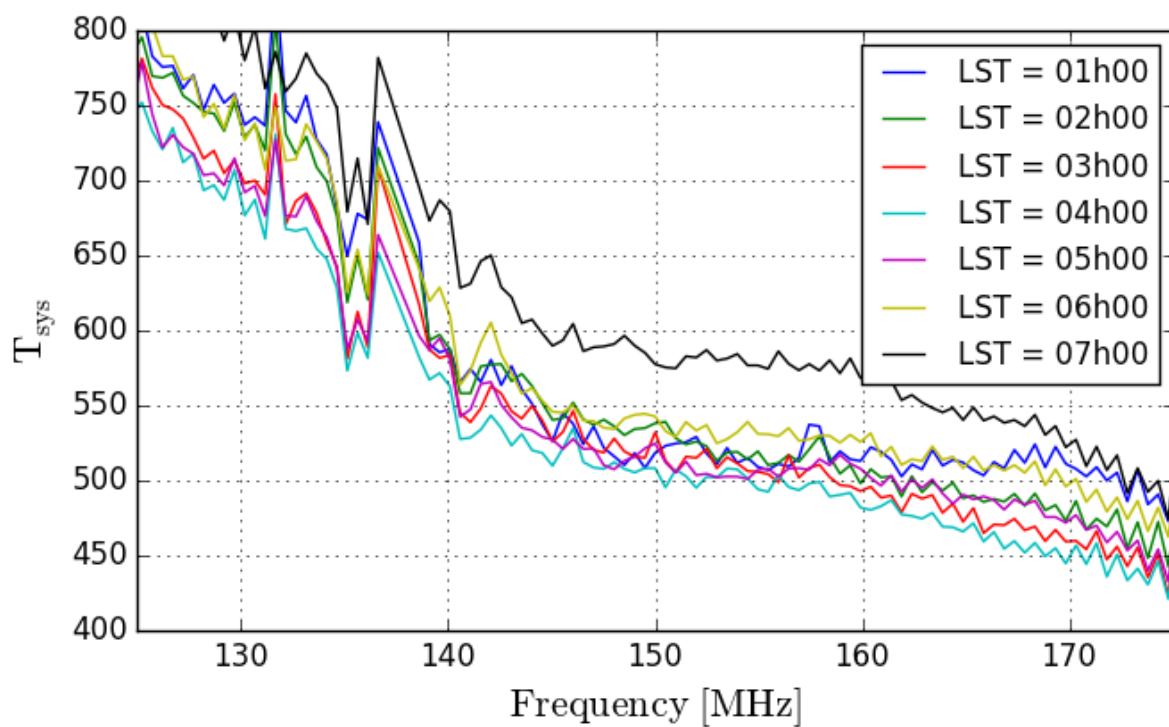


Figure 2.12: System temperature, inferred from the variance of samples falling in an LST bin, averaged over one-hour intervals in LST. The measured value in the 150–160 MHz range is consistent with previous determinations of system temperature (Jacobs et al. 2015; P14).

2.5.1 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, $P_{21}(\mathbf{k})$, defined such that

$$\langle \tilde{T}_b(\mathbf{k}) \tilde{T}_b^*(\mathbf{k}') \rangle = (2\pi)^3 \delta^D(\mathbf{k} - \mathbf{k}') P_{21}(\mathbf{k}), \quad (2.11)$$

where $\tilde{T}_b(\mathbf{k})$ is the spatial Fourier transform of the brightness temperature distribution on the sky, $\langle \rangle$ denotes an ensemble average, and δ^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 $\mathbf{x}$. 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,

$$\hat{q}_\alpha = \frac{1}{2} \mathbf{x}^\dagger \mathbf{C}^{-1} \mathbf{Q}_\alpha \mathbf{C}^{-1} \mathbf{x} - b_\alpha, \quad (2.12)$$

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

The bias term b_α in Equation (2.12) 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 (2.12). 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 $\mathbf{x}_1$ and $\mathbf{x}_2$ and computing

$$\hat{q}_\alpha = \frac{1}{2} \mathbf{x}_1^\dagger \mathbf{C}^{-1} \mathbf{Q}_\alpha \mathbf{C}^{-1} \mathbf{x}_2, \quad (2.13)$$

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 $\mathbf{C}^{-1}$). To normalize our results, we group

the unnormalized bandpowers into a vector $\hat{\mathbf{q}}$ and apply a matrix $\mathbf{M}$ (whose exact form we specify later), so that

$$\hat{\mathbf{p}} = \mathbf{M}\hat{\mathbf{q}} \quad (2.14)$$

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

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

$$\hat{\mathbf{p}} = \mathbf{W}\mathbf{p}, \quad (2.15)$$

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

$$\mathbf{W} = \mathbf{M}\mathbf{F}, \quad (2.16)$$

where

$$\mathbf{F}_{\alpha\beta} = \frac{1}{2}\text{tr}(\mathbf{C}^{-1}\mathbf{Q}_{\alpha}\mathbf{C}^{-1}\mathbf{Q}_{\beta}), \quad (2.17)$$

which we have suggestively denoted with the symbol $\mathbf{F}$ 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 $\mathbf{M}$ is chosen in such a way that $\mathbf{W}$ satisfies this criterion, the resulting bandpower estimates $\hat{\mathbf{p}}$ will be properly normalized.

Beyond the normalization criterion, a data analyst has some freedom over the precise form of $\mathbf{M}$, which effectively also re-bins the bandpower estimates. One popular choice is $\mathbf{M} = \mathbf{F}^{-1}$, which implies that $\mathbf{W} = \mathbf{I}$. 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

$$\mathbf{\Sigma} = \text{Cov}(\hat{\mathbf{p}}) = \langle \hat{\mathbf{p}}\hat{\mathbf{p}}^{\dagger} \rangle - \langle \hat{\mathbf{p}} \rangle \langle \hat{\mathbf{p}} \rangle^{\dagger}. \quad (2.18)$$

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

$$\mathbf{\Sigma} = \mathbf{M}\mathbf{F}\mathbf{M}^{\dagger}. \quad (2.19)$$

The choice of $\mathbf{M} = \mathbf{F}^{-1}$ tends to give rather large error bars. At the other extreme, picking $\mathbf{M}_{\alpha\beta} \propto \delta_{\alpha\beta}/\mathbf{F}_{\alpha\alpha}$ (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 QEs in the following sections, we will pick an intermediate choice for $\mathbf{M}$, one that is carefully tailored to avoid the leakage of foreground power from low k modes to high k modes.

2.5.2 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 $\mathbf{x}$, $\mathbf{C}$, $\mathbf{Q}_\alpha$ for our analysis pipeline.

First, we consider $\mathbf{x}$, 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 2.2), 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 2.5.4. Note that because our data have been fringe-rate filtered in the manner described in Section 2.3.5, 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 2.3.3), 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

$$\mathbf{x}^{ri}(t) = \begin{pmatrix} V^{ri}(\nu_1, t) \\ V^{ri}(\nu_2, t) \\ \vdots \end{pmatrix}, \quad (2.20)$$

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 (2.12) is to weight the data by their inverse covariance. To do so, we of course require the covariance matrix $\mathbf{C}$, which by definition, is the ensemble average of $\mathbf{xx}^\dagger$, namely $\mathbf{C} = \langle \mathbf{xx}^\dagger \rangle$. Unfortunately, in our case the covariance is difficult to model from first principles, and we must resort to an empirically estimated $\mathbf{C}$. We make this estimation by taking the time average of the quantity $\mathbf{xx}^\dagger$ 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 2.5.3.

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

$$\mathbf{z}^{ri} = (\mathbf{C}^{ri})^{-1} \mathbf{x}^{ri} \quad (2.21)$$

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 2.13 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 2.3.3. In the figure, the top row corresponds to the data vector $\mathbf{x}^{ri}$ 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 $\mathbf{x}$ with itself and averaging over the time axis. Finally, the last row shows the results of inverse covariance weighting the data, namely $\mathbf{z}^{ri}$. 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 $\mathbf{x}^{ri}$ in Figure 2.13, 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 2.14 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 2.13 that look more noise-dominated. With the foreground-suppressed portion (right half) of Figure 2.13, the initial $\mathbf{x}^{ri}$ vectors already appear noise dominated (which is corroborated by the relatively noisy form of the eigenvalue spectra in Figure 2.14). The final $\mathbf{z}^{ri}$ 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 $\mathbf{C}^{-1}$, we may group our identical baselines into five different sets and average together $\mathbf{z}^{ri}$ vectors for baselines within the same set. That is, we form

$$\mathbf{z}_A^r = \sum_{i \in A} (\mathbf{C}^{ri})^{-1} \mathbf{x}^{ri}, \quad (2.22)$$

where A ranges from 1 to 5 and indexes the baseline set. At this point, we have 10 weighted data vectors $\mathbf{z}$ (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

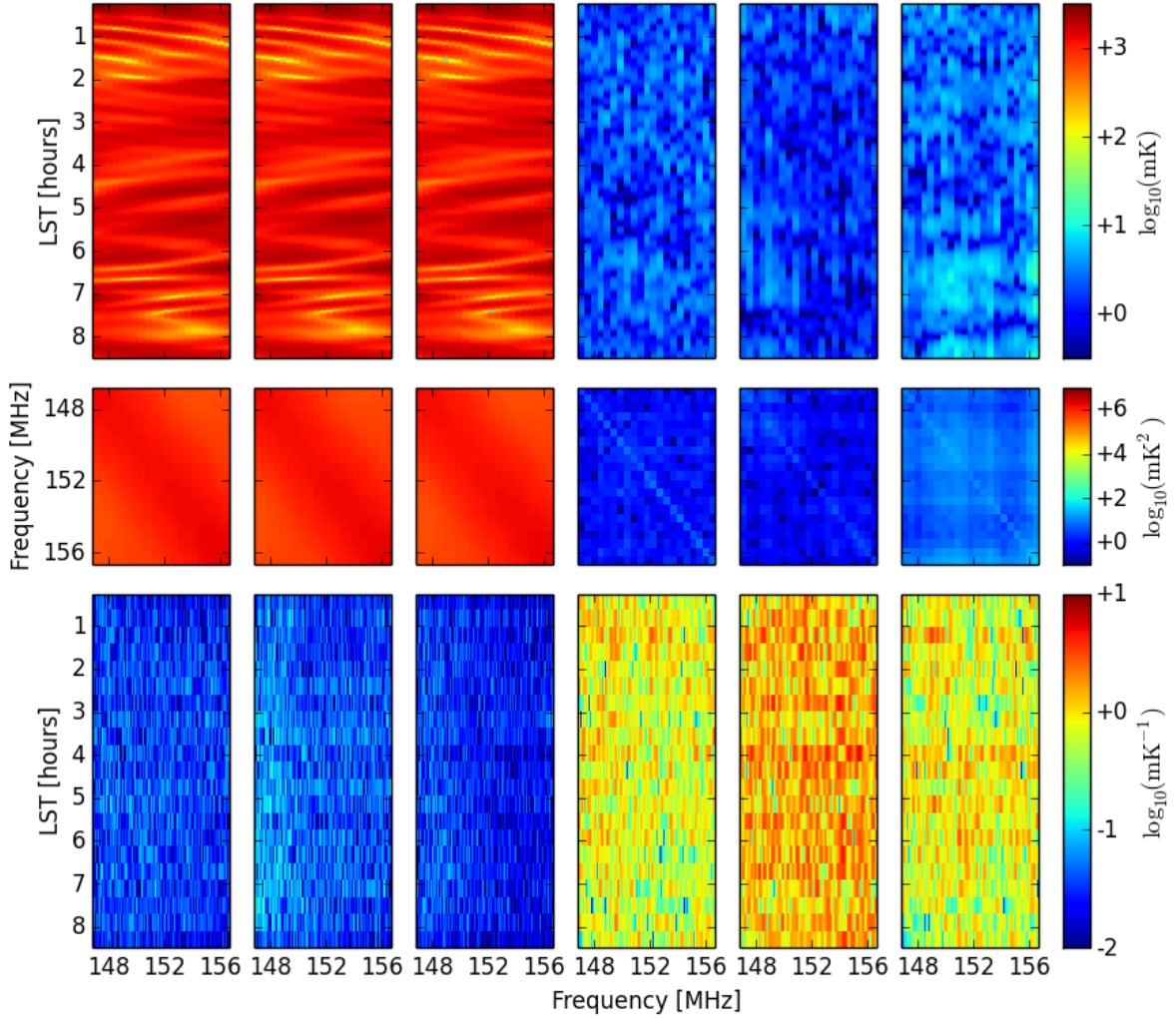


Figure 2.13: Visibilities before (top row) and after (bottom row) inverse covariance weighting. Signal covariance (middle row) is estimated empirically, averaging over LST. The three left/right columns show visibilities from three different baselines in a redundant group before/after delay filtering, respectively.

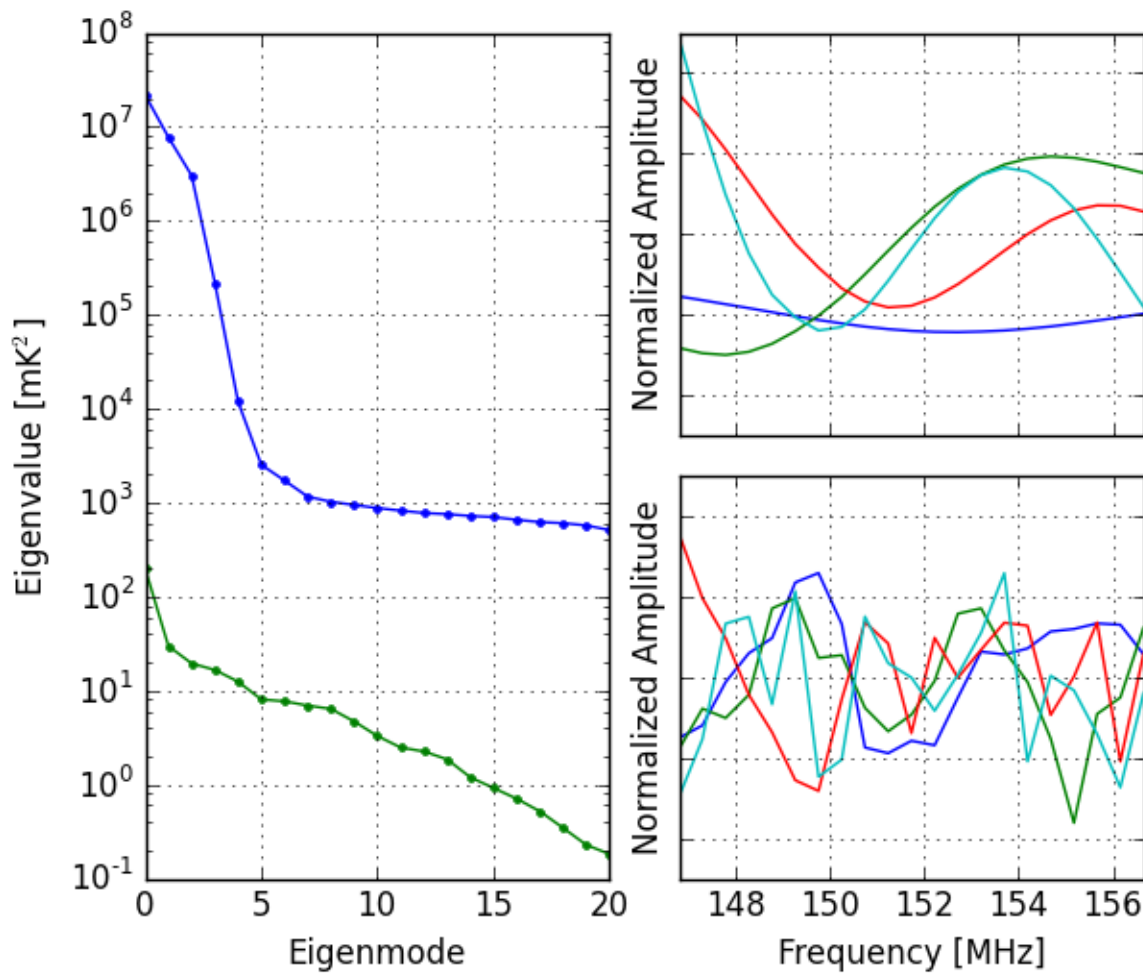


Figure 2.14: Eigenvalue spectrum of covariance matrices (left) empirically estimated from visibilities before (blue) and after (green) delay filtering. The four strongest eigenmodes of the filtered/unfiltered data are plotted on the top/bottom panels on the right, respectively.

Equation (2.13) to our present case where we have 10 different data vectors, we have

$$\hat{\mathbf{q}}_\alpha = \sum_{\substack{A,B,r,s \\ r \neq s, A \neq B}} \mathbf{z}_A^{r\dagger} \mathbf{Q}_\alpha \mathbf{z}_B^s, \quad (2.23)$$

so that auto-power contributions from identical baseline groups or identical even/odd indices never appear. Residual foreground bias will remain in Equation (2.23), 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 (2.23), it is necessary to derive a form for $\mathbf{Q}_\alpha \equiv \partial \mathbf{C} / \partial p_\alpha$. To do so, we follow the delay spectrum technique of P12a, where it was shown that

$$P(\mathbf{k}_{t\tau}) \approx \left(\frac{\lambda^2}{2k_B} \right)^2 \frac{X^2 Y}{\Omega B} \langle \tilde{V}_i(t, \tau) \tilde{V}_j^*(t, \tau) \rangle, \quad (2.24)$$

where $V_i(t, \tau)$ is the delay transform of baseline visibilities given by Equation (2.8), 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_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 $\mathbf{Q}_\alpha$ in Equation (2.23) is to perform a frequency Fourier transform on each copy of $\mathbf{z}$. It is therefore a separable matrix of the form $\mathbf{Q}_\alpha = \mathbf{m}_\alpha \mathbf{m}_\alpha^\dagger$, where $\mathbf{m}_\alpha$ is a complex sinusoid of a specific frequency corresponding to delay mode α . We may thus write

$$\hat{\mathbf{q}}_\alpha = \sum_{\substack{A,B,r,s \\ r \neq s, A \neq B}} \mathbf{z}_A^{r\dagger} \mathbf{m}_\alpha \mathbf{m}_\alpha^\dagger \mathbf{z}_B^s. \quad (2.25)$$

With an explicit form for $\mathbf{Q}_\alpha$, one now also has the necessary ingredients to compute the Fisher matrix using Equation (2.17).

Having computed the $\hat{\mathbf{q}}_\alpha$ s, we group our results into a vector $\hat{\mathbf{q}}$. This vector of unnormalized bandpowers is then normalized to form our final estimates of the power spectrum $\mathbf{p}$. As noted above, the normalization occurs by the $\mathbf{M}$ matrix in Equation (2.14), 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 $\hat{\mathbf{q}}$ (and therefore in $\mathbf{F}$, $\mathbf{M}$, and $\hat{\mathbf{p}}$ 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 $\mathbf{F} = \mathbf{L}\mathbf{L}^\dagger$, where $\mathbf{L}$ is a lower triangular matrix. Following that, we pick $\mathbf{M} = \mathbf{D}\mathbf{L}^{-1}$, where $\mathbf{D}$ is a diagonal matrix chosen to adhere to the normalization constraint that $\mathbf{W} = \mathbf{M}\mathbf{F}$ has rows that sum to unity. In this case,

the window function matrix becomes, $\mathbf{W} = \mathbf{D}\mathbf{L}^\dagger$. This means that $\mathbf{W}$ 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 $\hat{\mathbf{p}}$ prevents leakage from this edge bin that will contain aliased power. Figure 2.15 shows the resulting window functions.

Our choice of normalization matrix also has the attractive property of eliminating error correlations between bandpower estimates. Using Equation (2.19), we have that

$$\mathbf{\Sigma} = \mathbf{D}\mathbf{L}^{-1}\mathbf{L}\mathbf{L}^\dagger\mathbf{L}^{-1}\mathbf{D} = \mathbf{D}^2. \quad (2.26)$$

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.

2.5.3 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.,

$$\mathbf{C} = \langle \mathbf{x}\mathbf{x}^\dagger \rangle, \quad (2.27)$$

where $\mathbf{x}$ 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 $\mathbf{C}$, 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 2.3.5 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,

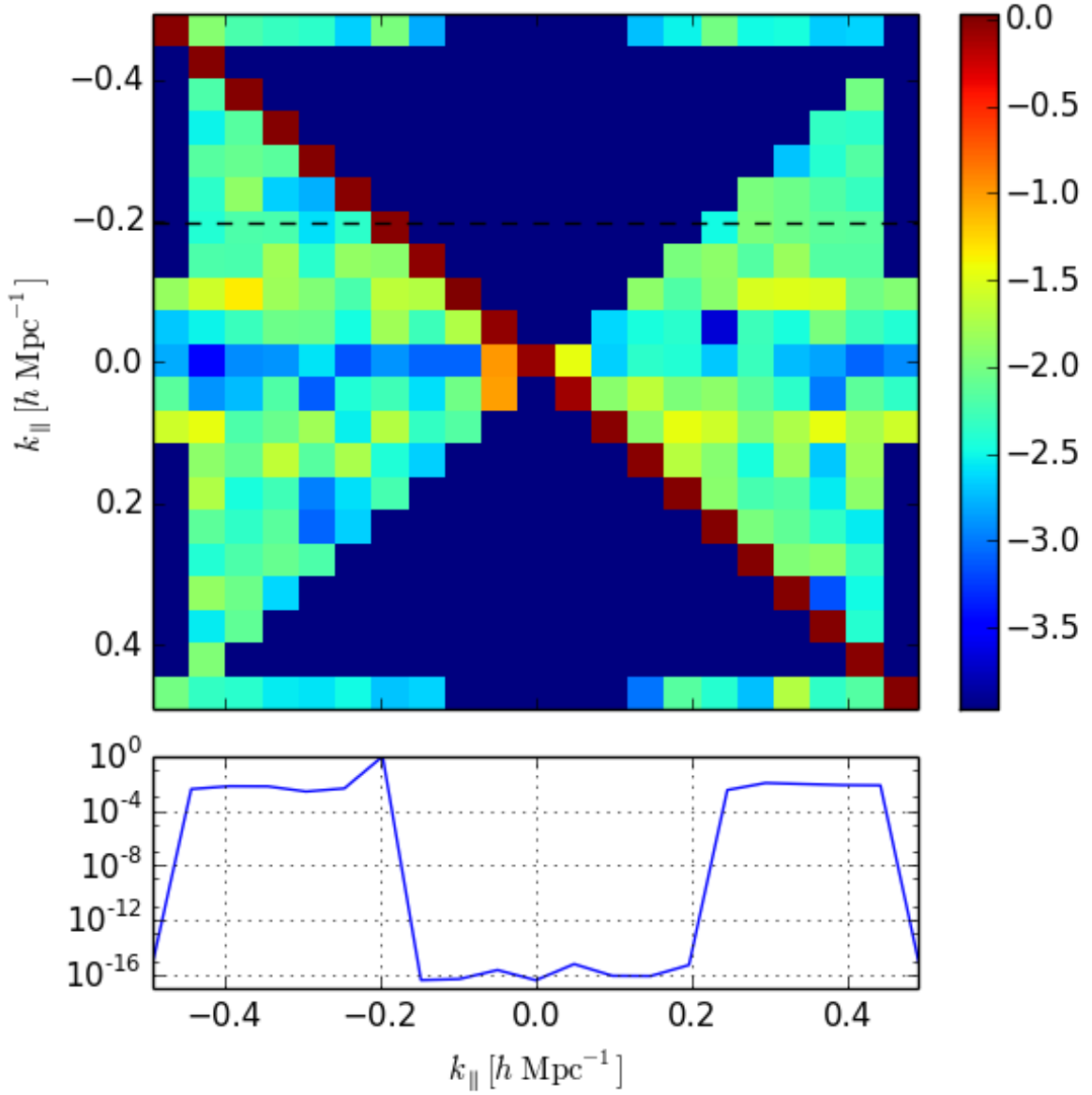


Figure 2.15: The window function matrix $\mathbf{W}$, as defined in Equation (2.15). The i^{th} row corresponds to the window function used in the estimate of the power spectrum for the i^{th} k -mode. Color scale indicates $\log_{10} \mathbf{W}$. The inset plot illustrates the window function along the dashed line in the upper panel. As described in Section 2.5.2, $\mathbf{M}$ in Equation (2.16) 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.

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 $\mathbf{C}$ 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 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 2.16 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 2.16, we see that deviations from unity signal transfer begin at power spectrum amplitudes of $10^7 \text{mK}^2 (h^{-1} \text{Mpc})^3$. For the range of output power spectrum amplitudes in our final estimate of the 21 cm power spectrum (Figure 2.18), we show that signal loss is $< 2\%$ at 95% confidence.

Table 2.1: SIGNAL LOSS VERSUS ANALYSIS STAGE

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

As shown in Table 2.1, 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 2.3). 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,

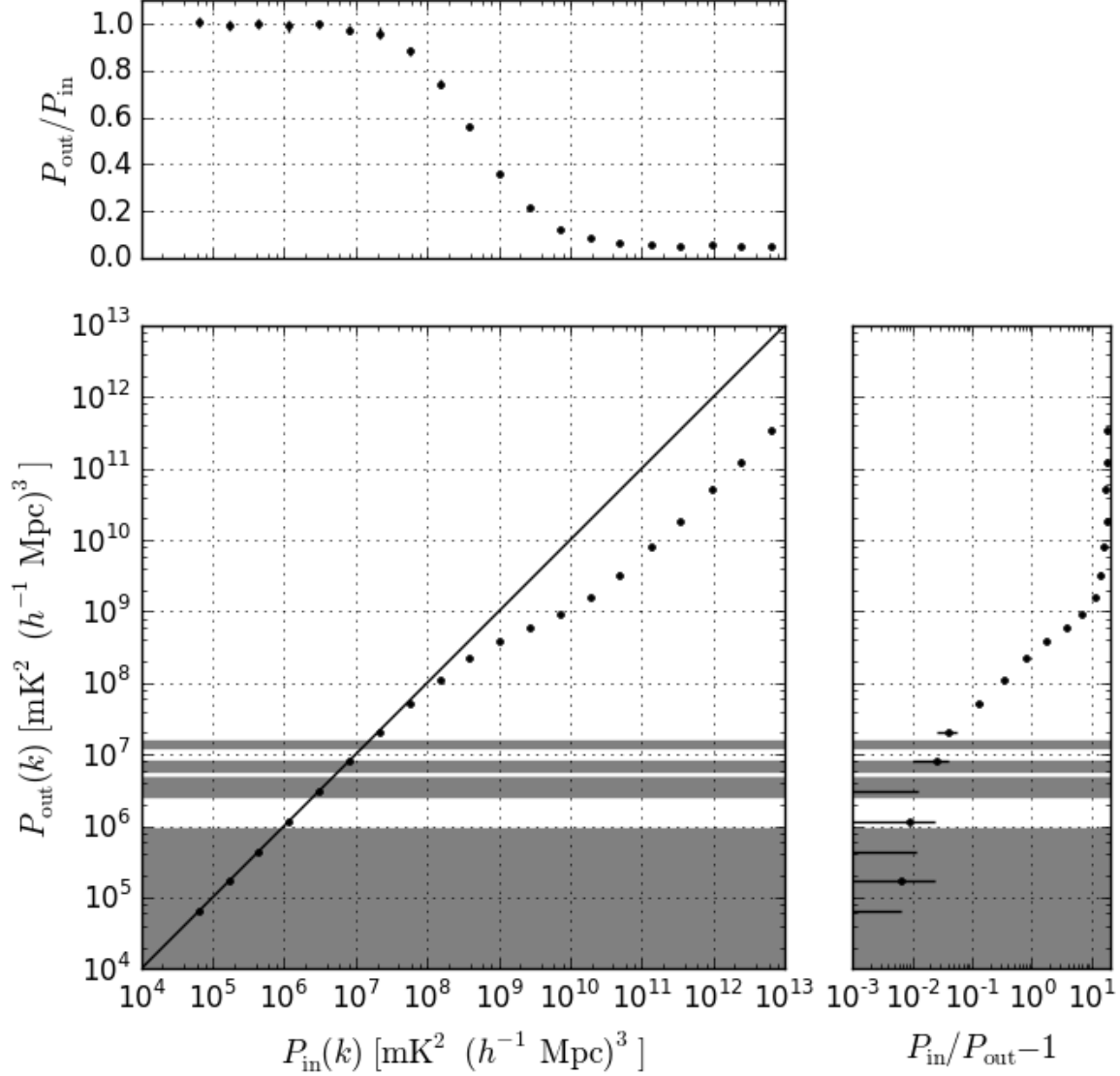


Figure 2.16: Recovered power spectrum signal as a function of injected signal amplitude. Shaded regions indicate the range in measured amplitude of power spectrum modes in Figure 2.18. Error bars indicate 95% confidence intervals as determined from the Monte Carlo simulations described in Section 2.5.3. 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 $P_{\text{in}}/P_{\text{out}}$ 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.

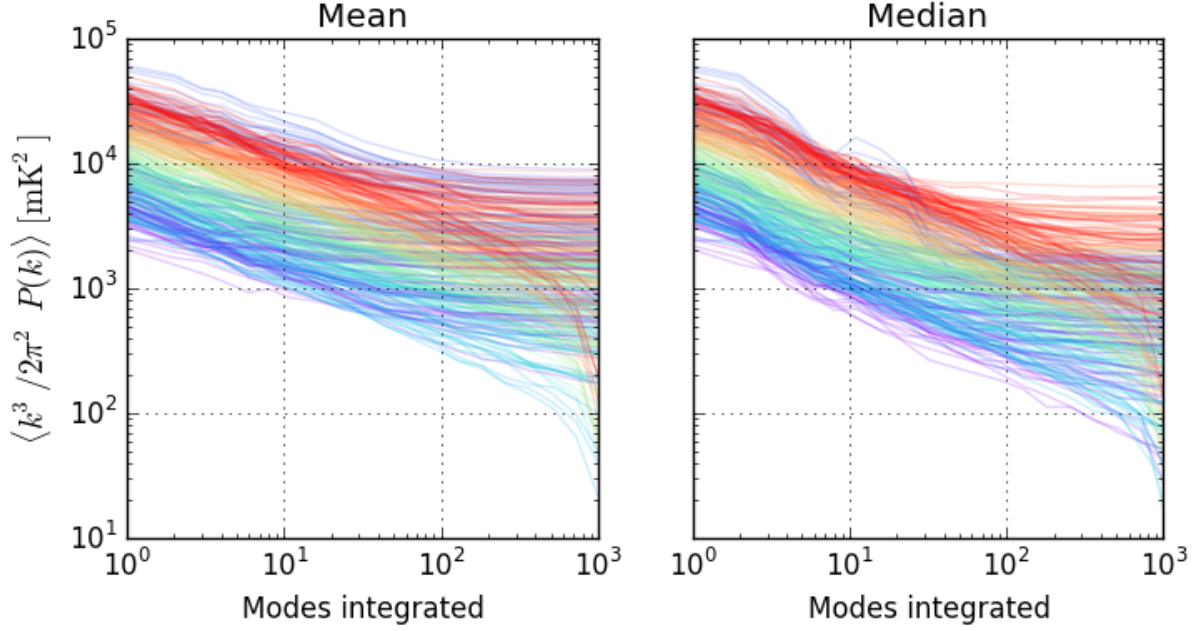


Figure 2.17: Absolute value of the cumulative mean (left) and median (right), as a function of number of modes of the power spectrum band power for $k_{\parallel}$ modes ranging from -0.49 (red) to $0.44h \text{ Mpc}^{-1}$ (violet). Here, modes are defined as samples from different redundant baseline 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 2.18, along with a correction factor compensating for the difference between the mean and median in estimating variance.

where we apply the maximum signal loss correction to be conservative.

2.5.4 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 (2.22))
2. Power spectrum estimates from each of the three redundant baseline types (described in Section 2.2) 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 $P(\mathbf{k})$. Repeating random draws allows one to quantify the inherent scatter—and hence the error bars—in our estimate of $P(\mathbf{k})$. When plotting $\Delta^2(k) \equiv k^3 P(k)/2\pi^2$ instead of $P(\mathbf{k})$, 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 2.17, 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 2.1, we apply a $1/\ln 2$ correction factor to our power spectrum estimates to infer the mean from the median of a Rayleigh distribution.

2.6 An Updated Understanding of Signal Loss and Power Spectrum Estimation

In this section we discuss our updated understanding of signal loss as it pertains to the power spectrum pipeline discussed in the previous section. The calculation of cosmological signal loss done in the previous section was found to be incorrect; using the same analysis it underestimated the amount of loss through the pipeline by a factor of ~ 10000 in power spectrum units.

In addition to underestimating the amount of signal loss through our analysis, errors in the estimation of the theoretical noise variance of the power spectrum and in computing bootstrapped errors led us to believe that the original estimate of the power spectrum was correct. Specifically, we overestimated the sensitivity of our dataset by a factor of 3 (in

mK), leading to a lower noise level. We also underestimated the error in the measurement by a factor of 2 (in mK). These two errors were of similar magnitude, leading us to believe our measurement was self-consistent. However, the effects of signal loss were the reason the power spectrum limits were incorrect.

The following sections detail our current understanding of signal loss and the errors that were made in the original analysis pipeline. For more details on signal loss estimation, the errors in this measurement, and a new method for estimating the power spectrum that minimizes signal loss, see Cheng et al. (2018, submitted).

Updated Understanding of Signal Loss

In Section 2.5 we outlined signal loss estimation through the power spectrum pipeline using an injection and recovery methodology. Since the initial publication of the results in this chapter, an error in this signal loss estimation methodology was found, leading to more signal loss than previously expected. In this section, I will discuss the errors in the original signal loss methodology.

We first set up the mathematical framework for estimating signal loss in our power spectrum pipeline, as described in section 2.5.3. To estimate signal loss, we use a signal injection and recovery method, similar to Masui et al. (2013). We inject a mock EoR signal with known power into the quadratic estimator pipeline and measure the output recovered power. An output power that is lower than the input power is indicative of signal loss. Let $\mathbf{x}$ be the data vector and $\mathbf{e}$ be the known mock EoR signal. We define the data vector $\mathbf{r}$ as

$$\mathbf{r} \equiv \mathbf{x} + \mathbf{e}, \quad (2.28)$$

the sum of the data and input mock EoR signal. The mock EoR signal we use is a Gaussian random signal that rotates with the sky and is added to every baseline used in estimating the power spectrum. The input power in the α^{th} $\mathbf{k}$ bin is given by

$$\hat{P}_{in}^{\alpha} \equiv M_{in}^{\alpha} \mathbf{e}^{\dagger} \mathbf{I} \mathbf{Q}^{\alpha} \mathbf{I} \mathbf{e}, \quad (2.29)$$

where M_{in}^{α} is the normalization matrix for the α^{th} bin, $\mathbf{I}$ is the identity matrix, and $\mathbf{Q}^{\alpha}$ encodes the response of covariance matrix to the α^{th} power spectrum bin. In our case, $\mathbf{Q}^{\alpha}$ encodes the Fourier transform. The input power spectrum defined in Equation 2.29 is the true power spectrum of the injected mock EoR signal only. The estimate of this power spectrum does not have signal loss since an identity weighting is used. In the uniform weighted limit, the estimator is equivalent to the delay spectrum estimate of the power spectrum in Equation 2.24. As shown in Parsons et al. (2012b), this estimator is lossless.

To assess the amount of signal loss through the system, we need to compare the input power to the output power, defined as

$$\hat{P}_{out}^{\alpha} \equiv M_r^{\alpha} \mathbf{r}^{\dagger} \mathbf{C}_r^{-1} \mathbf{Q}^{\alpha} \mathbf{C}_r^{-1} \mathbf{r}, \quad (2.30)$$

where $\mathbf{C}_r^{-1}$ is the empirical covariance estimated using the $\mathbf{r}$ data vector defined in Equation 2.28, and M_r^α is the normalization matrix for the α^{th} bin. Since we are interested in the frequency-frequency covariance, the covariance matrix is estimated by taking the ensemble average of the data vector over time: $\mathbf{C}_r^{-1} = \langle \mathbf{r}\mathbf{r}^\dagger \rangle$ and crucially, is not equivalent to the true covariance since it is estimated with a finite sample of data. The amount of signal loss is then calculated by comparing $\hat{P}_{out}$ to $\hat{P}_{in}$. Specifically, it is calculated as a function of input signal strength by varying the input amplitude of the mock EoR signal, as shown in Figure 2.16 (however, note that this curve does not take into account cross-terms as discussed below). We run multiple simulations of the mock EoR signal for a specified injection level to estimate the error on the signal loss. In the original analysis, 100 simulations were used.

To gain a better understanding of where signal loss is coming from, we plug in the constituents of $\mathbf{r}$ into Equation 2.30 and expand:

$$\hat{P}_{out}^\alpha = M_r^\alpha (\mathbf{x} + \mathbf{e})^\dagger \mathbf{C}_r^{-1} \mathbf{Q}^\alpha \mathbf{C}_r^{-1} (\mathbf{x} + \mathbf{e}) \quad (2.31)$$

$$\propto \mathbf{x}^\dagger \mathbf{C}_r^{-1} \mathbf{Q}^\alpha \mathbf{C}_r^{-1} \mathbf{x} + \mathbf{e}^\dagger \mathbf{C}_r^{-1} \mathbf{Q}^\alpha \mathbf{C}_r^{-1} \mathbf{e} + \mathbf{x}^\dagger \mathbf{C}_r^{-1} \mathbf{Q}^\alpha \mathbf{C}_r^{-1} \mathbf{e} + \mathbf{e}^\dagger \mathbf{C}_r^{-1} \mathbf{Q}^\alpha \mathbf{C}_r^{-1} \mathbf{x} \quad (2.32)$$

$$\propto \mathbf{x}^\dagger \mathbf{C}_r^{-1} \mathbf{Q}^\alpha \mathbf{C}_r^{-1} \mathbf{x} + \mathbf{e}^\dagger \mathbf{C}_r^{-1} \mathbf{Q}^\alpha \mathbf{C}_r^{-1} \mathbf{e} + 2\mathbf{x}^\dagger \mathbf{C}_r^{-1} \mathbf{Q}^\alpha \mathbf{C}_r^{-1} \mathbf{e}, \quad (2.33)$$

where we drop the normalization term and the last line uses the fact that $\mathbf{C}_r^{-1}$ is hermitian. The three terms that contribute to the output power are the data-data, EoR-EoR, and data-EoR terms. The data-EoR cross-terms were neglected in the original analysis, leading to significant signal loss as we'll show below. Depending on the input amplitude of the mock EoR signal, the shape of the transfer curve⁶ varies. We briefly explain and highlight three interesting regimes of the transfer curve. A detailed analysis of these different regimes are presented in (Cheng et al. 2018, submitted).

When the input mock EoR signal is low (compared to the original data, $\mathbf{x}$), both $\mathbf{r}$ and $\mathbf{C}_r^{-1}$ are dominated by the original data vector. In this case the cross-term, defined as those terms with both $\mathbf{x}$ and $\mathbf{e}$ vectors, is normally distributed around zero because $\mathbf{x}$ and $\mathbf{e}$ are not correlated to one another. Additionally, $\mathbf{C}_r^{-1}$ is dominated by $\mathbf{x}$ and does not correlate $\mathbf{x}$ with $\mathbf{e}$. In this regime, the output power spectrum amplitude is set by the input data $\mathbf{x}$. There is negligible signal loss in this regime, and the transfer curve is linear ($\beta = 1$).

In the regime where $\mathbf{e}$ is much larger than the data, $\mathbf{x}$, the output power spectrum is driven by $\mathbf{e}$. The cross-term in this case is also normally distributed around zero because $\mathbf{x}$ and $\mathbf{e}$ are not correlated to one another and $\mathbf{C}_r^{-1}$ is dominated by $\mathbf{e}$. In this case, the transfer function is also linear ($\beta = 1$).

The interesting case occurs when the input mock EoR signal is of order the data amplitude. In this regime, two things happen: $\mathbf{C}_r^{-1}$ is not dominated by any one component and more crucially, the expectation of the cross-term becomes negative. The negative cross-term is the result of couplings between $\mathbf{C}_r^{-1}$ and $\mathbf{x}$. In the other regimes discussed above, the

⁶We define the transfer curve as Figure 2.16. Namely, it is the ratio of P_{out} to P_{in} as P_{in} varies in magnitude. Generally, the transfer curve has the functional form $P_{out} = \beta P_{in}$. If there is no signal loss, $\beta = 1$. However, if signal loss does occur, β can follow a more complicated non-linear function.

cross-terms were normally distributed around zero, but in this case they are systematically negative as discussed in Cheng et al. (2018, submitted) and Switzer et al. (2015). For a deeper investigation of this systematic negative correlation we refer the reader to Cheng et al. (2018, submitted). In this section, we use this negative correlation result when discussing signal loss.

With the mathematical framework outlined, we now discuss the errors in the original signal loss framework discussed in Section 2.5.3, which we briefly review here. Recall that we measured the input power spectrum, P_{in} as defined in Equation 2.29, and compared it to the output power spectrum, P_{out} . The crucial error in estimating the signal loss occurred in the definition of P_{out} . In the original analysis, P_{out} was defined to be simply

$$P'_{out} = \mathbf{e}^\dagger \mathbf{C}_r^{-1} \mathbf{Q}^\alpha \mathbf{C}_r^{-1} \mathbf{e}. \quad (2.34)$$

That is, we were only looking at the effect of the covariance matrix on the injected data vector. The crux of the error was to assume that the cross-terms in Equation 2.30 averaged to zero and were therefore ignored in the calculation of signal loss. However, significant negative correlations in the cross-terms bias the power spectrum low. Accounting for the cross-terms leads to a power spectrum limit that is 4 orders of magnitude higher using the same methodology outlined in this chapter.

Errors in Noise and Bootstrapping

In addition to errors in the estimate of signal loss through the power spectrum pipeline, two additional smaller errors also affected the results presented in the original work. The first of these errors was the calculation of the theoretical error in the power spectrum, namely, the noise power spectrum. As defined in Parsons et al. (2012a), Pober et al. (2013), and Chenget al. (2018, submitted), the noise prediction is given by

$$n(k) = \frac{X^2 Y \Omega_{eff} T_{sys}^2}{\sqrt{2 N_{lst} N_{seps} t_{int} N_{days} N_{bls} N_{pols}}}, \quad (2.35)$$

where $X^2 Y$ is the conversion from observing coordinates to cosmological coordinates, Ω_{eff} is the effective primary beam (which can be modified by the fringe-rate filter), T_{sys} is the system temperature, N_{lst} is the number of independent lst samples that go into the power spectrum estimate, N_{seps} is the number of baseline separations used, t_{int} is the integration time per independent lst sample, N_{days} is the number of days used, N_{bls} is the number of baselines contributing to the power spectrum estimate, and N_{pols} is the number of linear polarizations used in the estimate of the power spectrum. Equation 2.35 was used to estimate the noise of the power spectrum in Figure 2.18 (cyan curve), but the original analysis mis-estimated some of the factors, significantly underestimating the theoretical noise (overestimating the sensitivity of the data set).

More specifically, the original calculation of the noise curve mis-estimated N_{days} and N_{bls} . The original estimate used $N_{days} = 135$, but it ignored the fact that those 135 days were

split into two data sets (even and odd), that cross-correlations between the two datasets was used, and that not every lst had an equal number of samples due to sidereal drift. Using an updated expression for N_{days} derived in (Cheng et al. 2018, submitted), they find that the effective $N_{days} \sim 47$.

The number of baselines, N_{bls} , used in the original sensitivity calculation ignored the fact that baselines were averaged into 5 groups before being cross multiplied so that not all baseline cross-multiples were used to form the power spectrum. Accounting for this fact significantly reduces the number of effective baselines that are actually used in the calculation, going from $N_{bls} = 51$ in the original analysis to $N_{bls} \sim 32$. The exact equation for N_{bls} can be found in (Cheng et al. 2018, submitted). Accounting for the errors in N_{bls} , N_{days} , and a factor of $\sqrt{2}$ for combining positive and negative k -bins together, the sensitivity calculation decreases (increases the noise floor) by a multiplicative factor of 7 mK^2 .

The third main error in the power spectrum analysis presented in the original work was in the estimation of the error bars on the power spectrum from bootstrapping (as discussed in Section 2.5.4). As discussed in (Cheng et al. 2018, submitted), a number of errors in the way bootstrapping was performed led to an underestimate of the errors on the power spectrum results. We enumerate these errors below:

- The first of the errors pertains to the fact that bootstrapping is only valid when the dataset consists of independent samples. In the original analysis (Section 2.5.4), bootstrapping occurred over the baseline and time axes. Over the time axis, the number of draws per bootstrap was equal to the number of integrations in the original data vector, $\mathbf{x}$. However, since visibilities were fringe-rate filtered, the number of independent time samples was reduced to ~ 8 . Bootstrapping thus oversampled the time axis, leading to an underestimate of the error on the power spectrum by a multiplicative factor of $\sim 3 \text{ mK}^2$.
- In the original analysis, bootstrapping occurred over the baseline axis in addition to the time axis. Random draws of baselines were sampled, averaged together in groups (to save on computation), and then the power spectrum was estimated. However, bootstrapping should occur over the sample of interest, which in our case is the power spectrum, not the baseline axis. As discussed in Cheng et al. (2018, submitted), bootstrapping over the baseline axis produces variances that are inconsistent with simulations, but bootstrapping over the final cross power spectrum axis produces variances that are consistent with simulations. This effect underestimates the power spectrum errors by a multiplicative factor of ~ 2 .

2.7 Power Spectrum Results

In this section we present the power spectrum results from the original lossy power spectrum pipeline that includes 4 orders of magnitude of signal loss. The power spectrum result in Figure 2.18 is not correct and is left as a reference for the discussion in Section

2.7.2 on the constraints of the spin temperature. The first part of this section presents the power spectrum result and discusses the effect of the different analysis stages on the power spectrum. This discussion does not change due to the lossy power spectrum estimator and merely captures the qualitative assessment of choosing different analysis methods on the power spectrum.

2.7.1 Power Spectrum Constraints

Before presenting the lossy power spectrum results, we review how it was produced to produce it. We follow the power spectrum analysis procedure outlined in Section 2.5.2 (which we now know to be lossy; see Section 2.6), we incoherently combine independent power spectrum measurements made at different times and with different baseline groups using the median statistic. As described in Section 2.5.4, we bootstrap over all of these (not so) independent measurements, as well as over the selection of baselines included in the power spectrum analysis for each baseline group, in order to estimate 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 2.1. However, the errors in the original signal loss calculation are not corrected; see Section 2.6.

The results from the lossy analysis are plotted in Figure 2.18. 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 (2.8) (either from the chromaticity of the antenna response, or from the inherent spectrum of the foregrounds themselves). As discussed in Cheng et al. (2018, submitted), these detections remain in the updated power spectrum methodology and is attributed to foreground leakage. Somewhat more difficult to interpret are the 2.4σ excess at $k \approx 0.30h \text{ Mpc}^{-1}$ and the 2.9σ excess at $k \approx 0.44h \text{ Mpc}^{-1}$. 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 2.19), 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 with using the P14-type time-domain filtering (blue in Figure 2.19) decrease significantly using the improved fringe-rate filter. As discussed in Parsons et al. (2015), the normalization applied to Ω_{eff} for fringe-rate filtering correctly compensates for the effect of this filtering on power-spectral

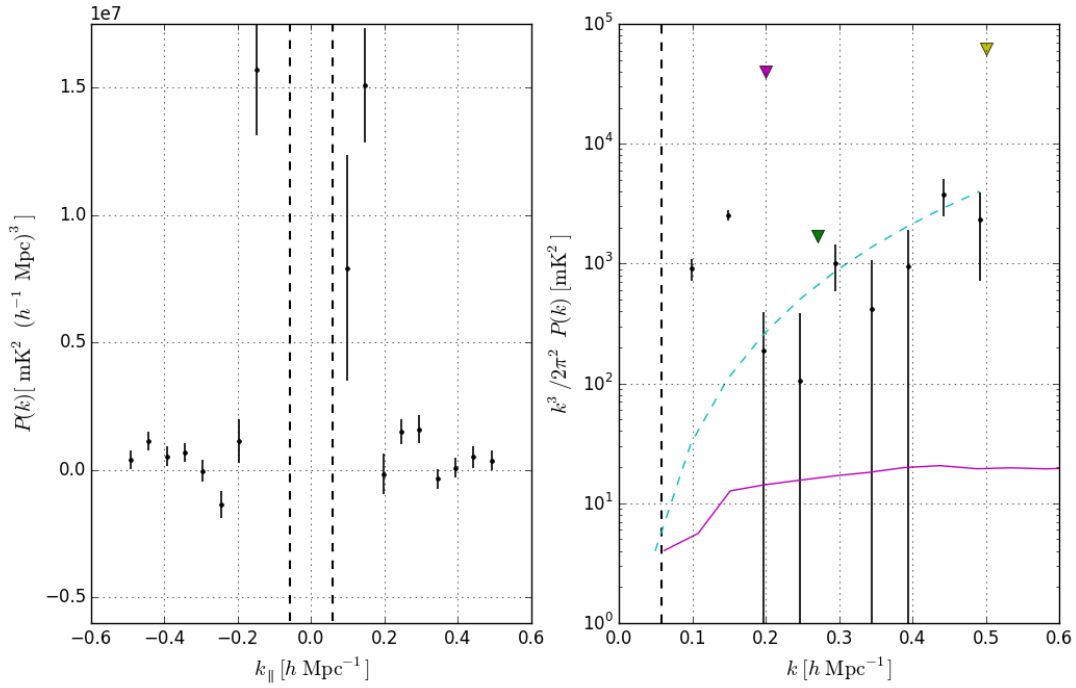


Figure 2.18: Measured lossy power spectrum (black dots with 2σ error bars) at $z = 8.4$ resulting from a 135 day observation with PAPER-64. The dashed vertical lines at $0.06h \text{Mpc}^{-1}$ show the bounds of the delay filter described in Section 2.3.3. The predicted 2σ upper limit in the absence of a celestial signal is shown in dashed cyan, assuming $T_{\text{sys}} = 500\text{K}$. The triangles indicate 2σ 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).

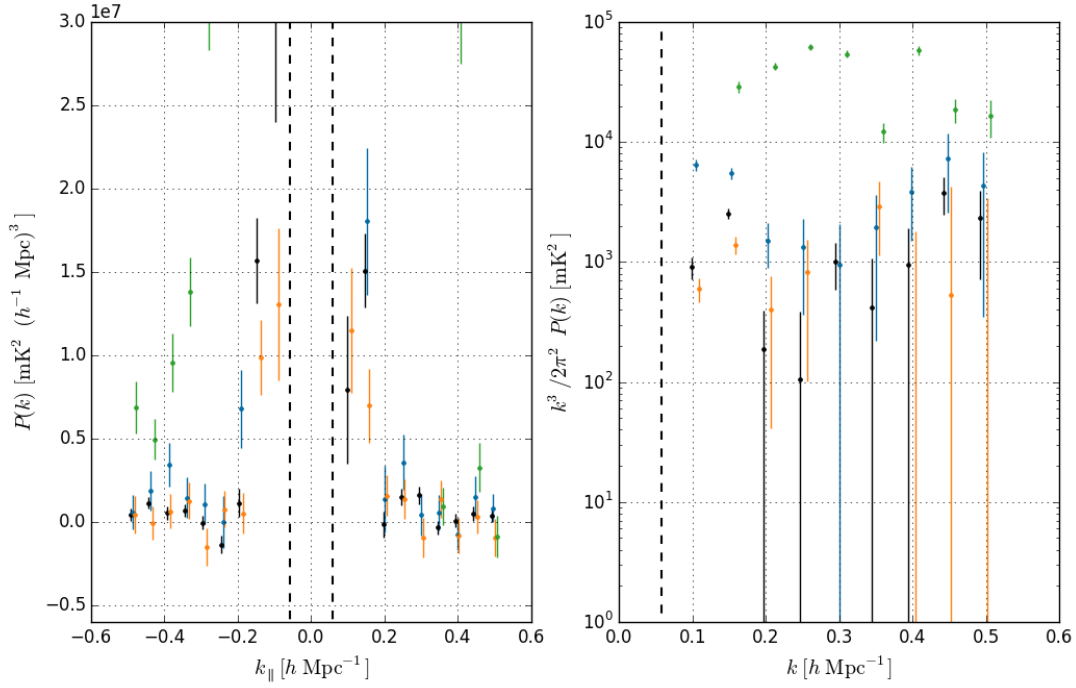


Figure 2.19: Diagnostic power spectra in the style of Figure 2.18 illustrating the impact of various analysis stages. The blue power spectrum uses the P14 fringe-rate filter combined with crosstalk removal. Green illustrates the result using the improved fringe-rate filter, but without crosstalk removal. 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 2.18.

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 $< 100 \text{ mK}^2$ for all k .

That the excesses at $k \approx 0.30$ and $0.44h \text{ Mpc}^{-1}$ are relatively unaffected by the filtering could be an indication that they are more isotropically distributed, but more likely, it may mean that they 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 2.19). 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 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 lossy power spectrum (Figure 2.18) into a single upper limit, we fit a flat $\Delta^2(k)$ model to measurements in the range $0.15 < k < 0.5h \text{ Mpc}^{-1}$. We use a uniform prior of amplitudes between -5000 and 5000 mK^2 , and assume measurement errors are Gaussian. Figure 2.20 shows the posterior distribution of the fit. From this distribution, we determine a mean of $(18.9 \text{ mK})^2$ and a 2σ upper limit of $(22.4 \text{ mK})^2$. The measured mean is inconsistent with zero at the 4.7σ level, indicating that we are detecting a clear power spectrum excess at $k > 0.15h \text{ Mpc}^{-1}$.

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 $(22.4 \text{ mK})^2$ at $z = 8.4$, continues to serve as a conservative upper limit.

2.7.2 Spin Temperature Constraints

In this section we show the potential implication of the results shown in Figure 2.18 on the spin temperature of the 21cm line at $z = 8.4$. Note that the results presented in this section are potential results that could be made if a power spectrum similar to the one presented in Figure 2.18 was made. We focus here on a simple parameterization of the shape of the 21cm power spectrum signal to put constraints on the IGM. For a more thorough analysis of constraints on the IGM using a simulation based framework, we refer the reader to Pober et al. (2015).

The brightness temperature of the 21cm signal, δT_b , arising from the contrast between

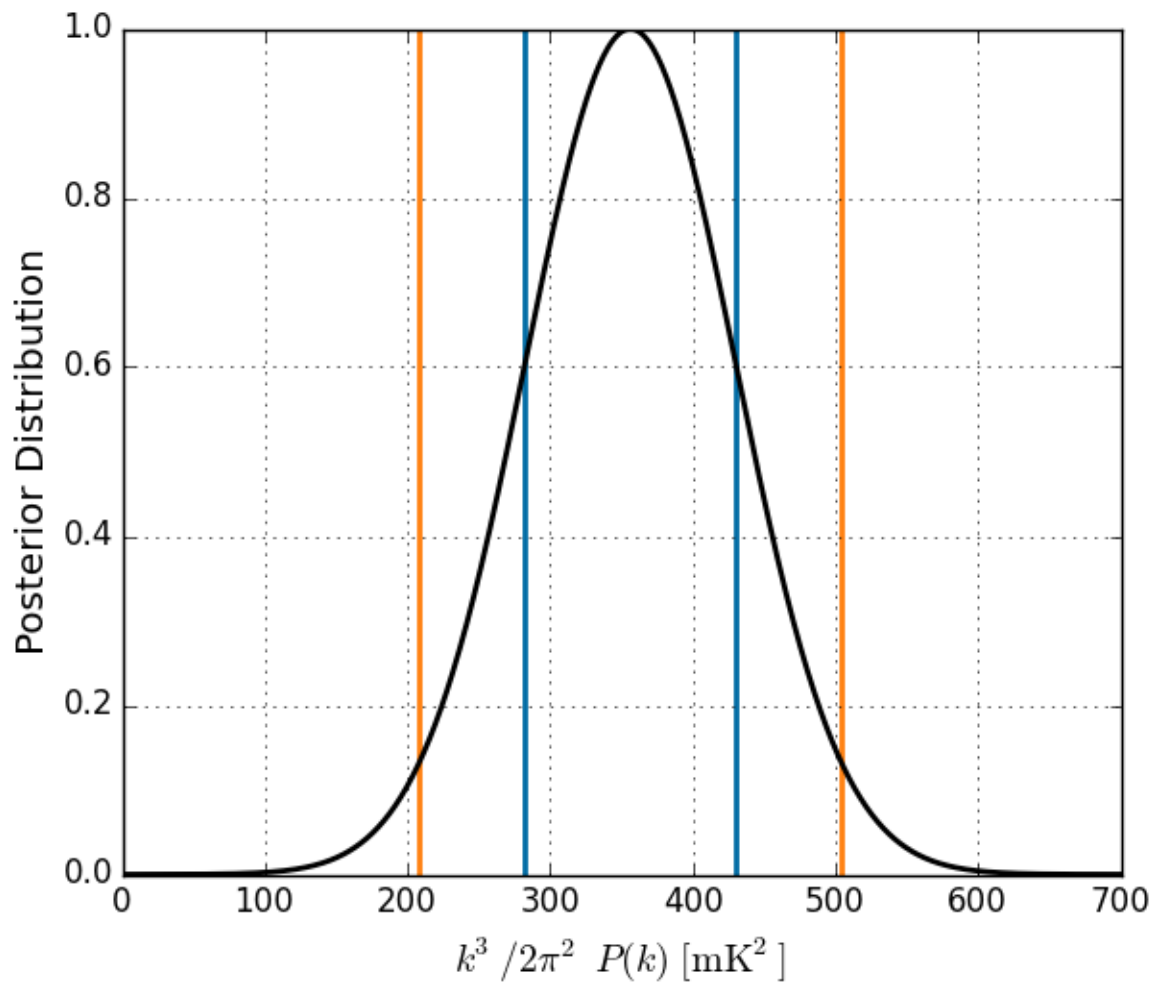


Figure 2.20: Posterior distribution of power spectrum amplitude for a flat $\Delta^2(k)$ power spectrum over $0.15 < k < 0.5h \text{ Mpc}^{-1}$ (solid black), assuming Gaussian error bars. The blue and orange vertical lines correspond to the 1σ and 2σ bounds, respectively. This posterior was calculated using the result in Figure 2.18 and is therefore biased low, but represents the type of information aggregation that can be used in a power spectrum estimate.

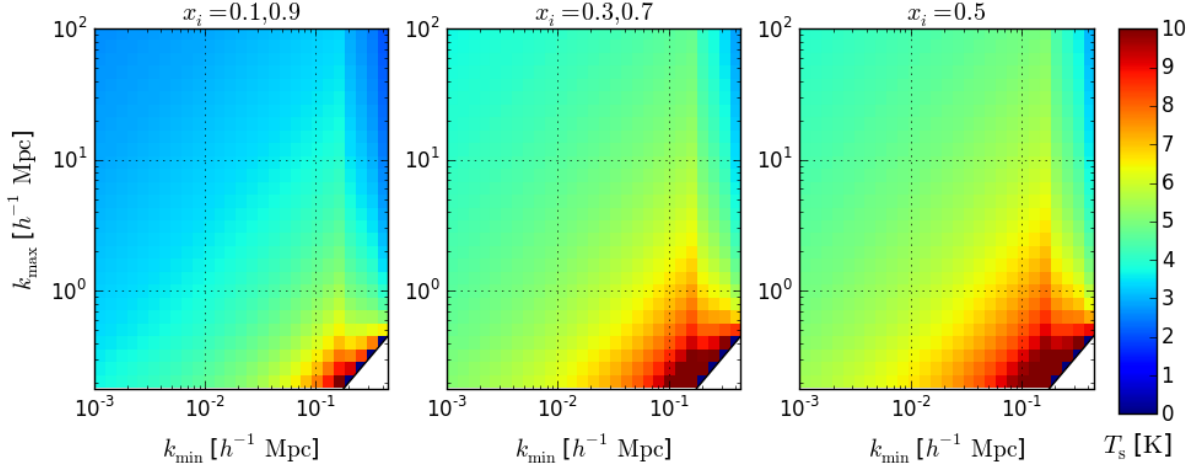


Figure 2.21: Constraints on the 21cm spin temperature at $z = 8.4$, assuming the patchy reionization model in Equations (2.39) and (2.41), which hold in the limit that $T_s < T_{\text{CMB}}$.

the cosmic microwave background, T_γ , and the spin temperature, T_s , is given by

$$\delta T_b = \frac{T_s - T_\gamma}{1 + z} (1 - e^{-\tau}) \approx \frac{T_s - T_\gamma}{1 + z} \tau, \quad (2.36)$$

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)

$$\tau = \frac{3c^3 \hbar A_{10} n_{\text{HI}}}{16k\nu_0^2 T_s H(z)} \quad (2.37)$$

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

$$\delta T_b \approx T_0 x_{\text{HI}} (1 + \delta) \xi, \quad (2.38)$$

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

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 21cm power spectrum:

$$\Delta_{21}^2(k) = T_0^2 \xi^2(z) \Delta_i^2(k), \quad (2.39)$$

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

As shown in P14, the maximum value of the prefactor in Equation (2.39) 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 $z_{\text{dec}} \approx 150$ (Furlanetto et al. 2006b) and then cools adiabatically as $(1+z)^2$. In this case, ξ is given by

$$\xi = 1 - \frac{1 + z_{\text{dec}}}{1 + z} \approx -\frac{150}{1 + z}. \quad (2.40)$$

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 $\Delta_{21}^2(k)$, we can use models for $\Delta_i^2(k)$ in Equation (2.39) 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:

$$\Delta_i^2(k) = (x_{\text{HI}} - x_{\text{HI}}^2) / \ln(k_{\text{max}}/k_{\text{min}}). \quad (2.41)$$

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 2.18. Figure 2.21 shows the results of these bounds for neutral fractions of $x_{\text{HI}} = 0.1, 0.3, 0.5, 0.7$, and 0.9 . In almost all cases (excepting $x_{\text{HI}} = 0.1, 0.9$ for $k_{\text{min}} < 0.1h \text{ Mpc}^{-1}$), we find that $T_s \gtrsim 3$ 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 $k_{\text{min}} \sim 0.1$ and $k_{\text{max}} \sim 30$ 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 $T_s \gtrsim 4$ 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. 2006b; Pritchard & Loeb 2008; Fialkov et al. 2014b). 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.

2.8 Discussion

This chapter lays out three advances in the estimation of the 21 cm power spectra:

1. the use of OMNICAL for redundant calibration significantly improves the clustering of measurements over the previous implementation of LOGCAL used in P14,
2. fringe-rate filtering further improves power spectrum sensitivity and suppresses systematics associated with foregrounds low in the primary beam, and
3. moving from a lossless quadratic estimator targeting difference modes in redundant measurements to an OQE formalism opens up a toolbox of analysis strategies.

Figure 2.19 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 3.6 and 2.10), 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 21cm cosmology.

2.9 Conclusions

In this chapter, we analyzed a season of PAPER-64 data. We presented calibration techniques, foreground removal using the delay transform, optimal time integrating using fringe-rate filters, and discussed instrumental performance. We additionally showed a method for estimating the 21 cm power spectrum using optimal quadratic estimators. Using this technique requires care, as empirically estimated covariance matrices cause cosmological signal loss. The limits presented in this Chapter incur 4 orders of magnitude of signal loss. We also

presented a method for putting a lower limit on the spin temperature for hydrogen in the IGM. If an experiment were to produce limits that were on the order of the lossy, uncorrected limits presented in the Figure 2.18, we can show that $T_s > 4\text{ K}$ using a patch reionization model. A more detailed analysis of power spectrum constraints on spin temperature using semi-analytic reionization/heating simulations is presented in (Pober et al. 2015).

The power spectrum results are 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 chapter, we extended 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.

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 Hydrogen Epoch of Reionization Array (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 has been fully funded for the construction of the instrument through the National Science Foundation's *Mid-Scale Innovations Program* and the Moore Foundation. Additionally, HERA was recently fully funded to deliver science data products to the community and explore imaging based power spectrum techniques through the National Science Foundation. 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 (2.8). Sensitivity forecasts for a 331-element HERA array and SKA-Low show that they can deliver detections of the 21cm reionization signal at a significance of 39σ and 21σ , respectively, using the same the conservative foreground avoidance strategy employed in this Chapter (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.

Acknowledgments

A version of this chapter was previously published in Ali et al. [AJ 809 61](#), and is reproduced with the permission of all coauthors and the copyright holder. Copyright 2015 American Astronomical Society.

Chapter 3

Real-Time Systems in Astronomy: An Application to the Hydrogen Epoch of Reionization Array (HERA)

Modern telescopes are beginning to produce petabytes of data, posing a significant challenge for data analysts. As data volume increases, experiments will become limited by manual data processing, and as a consequence, data reduction will be significantly slowed. Therefore, low latency data reduction pipelines with the absence of human intervention, which we define as real time systems, are becoming increasingly necessary. In this Chapter, we present a methodology for the development of real time systems and discuss the two aspects of real time systems: one as a low latency system producing a set of pre defined data products in the absence of outside intervention, and two, as a data reduction system. We proceed by discussing radio arrays, specifically low frequency redundant radio arrays, by using the Hydrogen Epoch of Reionization Array (HERA) as a driver for the application of a real time system. In doing so, we present results from the initial version of HERA's real time system. Initial results from HERA's realtime system indicate that the degree to which baselines are redundant depends on the length of the baseline, where shorter baselines are less redundant to one another than longer baselines. We conclude this Chapter by discussing future improvements to HERA's real time system.

3.1 Introduction

In the past few decades, modern radio telescopes have been generating an increasing amount of data due to the development of larger and faster computer processors and digitizers, fueled by Moore's Law. The increase in data volume has surpassed reasonable storage capacity, implying a need for an increase in data processing and data compression to effectively analyze and write data to disk.

As radio telescopes and experiments generate larger amounts of data, practical challenges

of data storage need to be addressed. The amount of raw data produced and the limitations of long term data storage necessitate the need for data compression, where various axes of the dataset can be collapsed upon. Data compression is a critical driver for the development of new algorithms and analysis techniques that enable combining of data.

As new radio telescopes begin observing, response time for performance issues becomes a critical challenge. Solving this challenge will minimize downtime and eliminate wasted operation costs. With the increase in data volume, manually looking at all the data to assess possible issues with the instrument starts to become infeasible. How should telescopes synthesize their raw data to reveal instrumental shortcomings? Is there anything that can be done to mitigate them? These sorts of questions prompt radio astronomers to develop algorithms that provide feedback on instrumental performance, which is a critical component in quickly and correctly detecting issues with the instrument and reducing chances of missing out on opportunities for the next big scientific discovery. These opportunities can range from follow up observations of a source to providing observational support for other instruments.

One way to solve these challenges facing modern radio telescopes is to implement a real time system (RTS). Generically, an RTS is a computational system that algorithmically processes data in the absence of human intervention, thus outputting a predefined collection of data products. In particular, there are two aspects to an RTS: it is an automated system that produces data products with low latency, and more critically, it is a data reduction pipeline. As a result, RTS's are ubiquitous and key programs built into modern radio observatories and experiments. Such systems provide a way for producing routine analyses on astronomical data in real time, such as producing data calibrations, monitoring telescope performance, and compressing raw data.

The first aspect of an RTS is as a low latency processor that produces a predefined set of data products in the absence of human intervention. The quick delivery of data products can be used for various applications such as finding instrumental performance issues, characterizing the observing environment, or determining whether an observational follow up is necessary. The exact definition of low latency, the data products, and its applications depends on the type of experiment that the RTS is built for. For example, experiments interested in detecting transient phenomenon such as compact binary mergers ([Acernese et al. 2015](#); [LIGO Scientific Collaboration et al. 2015](#)) or fast radio bursts (FRBs) ([Newburgh et al. 2016](#); [The CHIME/FRB Collaboration et al. 2018](#)), require latencies on the order of milliseconds in order to detect and follow up on potential events. However, for experiments trying to detect the Epoch of Reionization (EoR) using the 21 cm line, latencies of 24 hours are reasonable to determine instrumental performance issues since these experiments usually take drift scan observations. The applications of an RTS in these two examples are different. In the transient case, the application is the detection of a possible event and a trigger that automatically decides if a followup observation is needed. In the case of the 21 cm experiment, the application is more about determining and relaying instrumental performance issues to observers. The common thread in both of these examples is that the RTS is independent of human intervention and has low latency, which is determined by the nature of the experiment.

The second, and more critical aspect of an RTS is as a data reduction pipeline. As stated previously, radio telescopes are beginning to produce upwards of a petabyte of data per observing season, leading to a problem of data storage and analysis. There is too much data to save and analyze in a realistic post processing pipeline. As such, there is a balance between disk space, how fast it will fill up (latency), and the cost of the data storage system. It is possible to throw an indefinite amount of money at the problem and store all the data all the time. However, for any realistic experiment, the amount of disk storage is limited by the cost, implying a disk storage-latency trade off, where not all the raw data can be stored all the time. RTS's need to be defined so that data compression can occur with a high degree of confidence so that a sufficient amount of data storage is available for the rest of the observing run.

The data reduction aspect of an RTS can be thought of as having the ultimate goal of data compression. The final outputs of an RTS are reduced data products which can quickly assess the scientific quality of the raw data. Data compression can either be lossy or lossless. Almost all data compression in radio astronomy is lossy; some information about the raw data is lost in the final compressed data product. In order to deliver quality compressed data products, a critical function of the RTS is to control the pollution of bad data through the data reduction pipeline. Bad data needs to be identified and removed before performing data compression so that final products are not corrupted. The mitigation of the mixing of bad data is driven by the first form of an RTS. Ensuring the integrity of data compression can require internal consistency checks or the use of outside information.

As a concrete example, consider the form of data compression used by most radio arrays: synthesis imaging ([Thompson et al. 2001](#)). In this technique, individual baseline visibilities are combined in order to make an image of the sky at a given time and/or frequency. In order to construct the image of the sky, poorly behaving baselines must be removed, using either a model of the sky (outside information) or by comparison to the rest of the array (an internal consistency check). Next, the array must be calibrated. Again, a model of the sky and antenna beams can be used, or, depending on the configuration of the array, internal consistencies between redundant baselines can be used. Depending on the amount of confidence the analyst has on what the sky and beam models are vs. how redundant certain baselines should be, determines the method of the calibration. Additionally, how much does the analyst really need to know before performing this kind of data compression? If there is a high degree of confidence on baseline redundancy and only a marginal degree of confidence on the antenna beams, should the analyst start with baseline redundancy first, before incorporating knowledge of the sky, creating a hierarchy of confidence? We will discuss a more thorough example in [Section 3.3](#).

The previous example focussed on radio arrays because their properties make them strong contenders for an RTS. They naturally have a large number of degrees of freedom since cross correlating antennas in an array scales as the square of the number of antennas. Radio experiments with a large number of antennas include low frequency arrays, specifically for searching for the highly red shifted 21 cm transition of neutral hydrogen ([Pritchard & Loeb 2012](#)). The goals of these experiments focus on the detection of the Epoch of Reioniza-

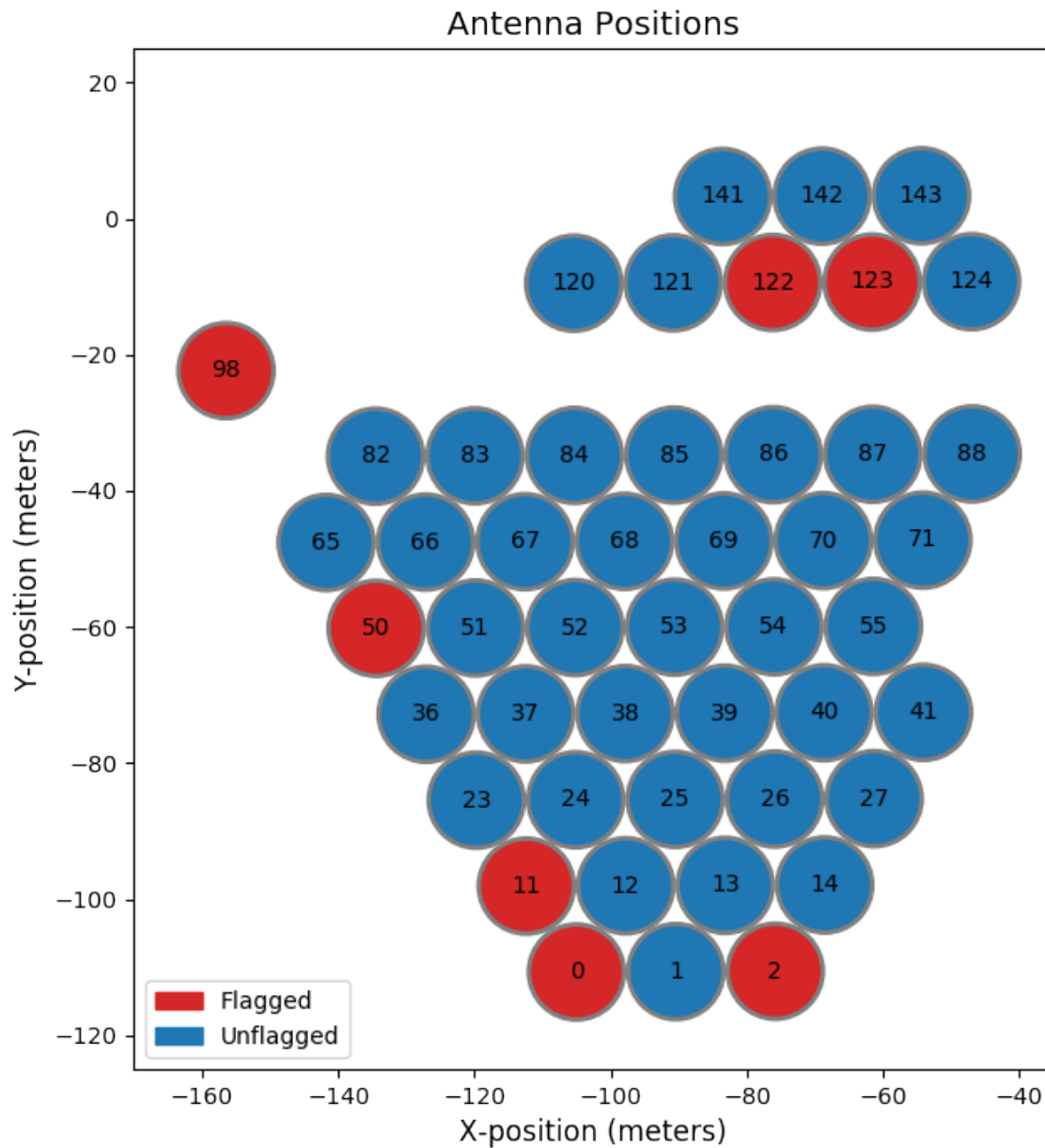


Figure 3.1: Antenna configuration of HERA-47 consisting of 40 nominally good antennas (blue) and 7 bad antennas that have been flagged (red) from the dataset. These 7 bad antennas were found as part of HERA’s RTS using the mean visibility and correlation metrics discussed in Section 3.3.2. This was the configuration of HERA on October 2, 2017 and for the data used in this Chapter.

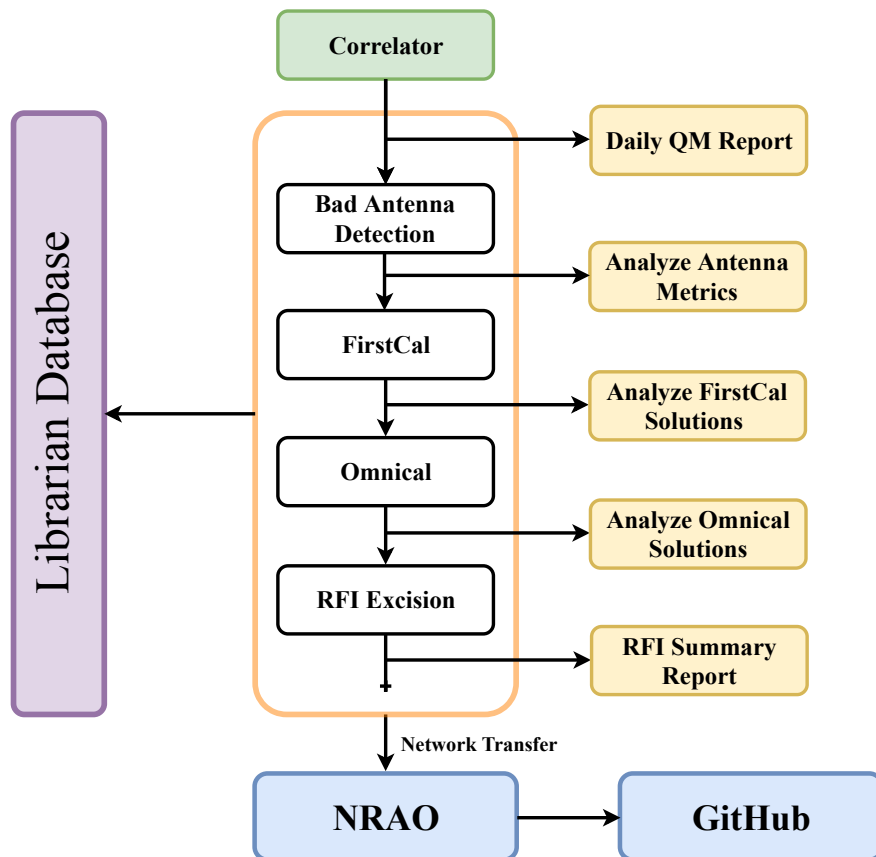


Figure 3.2: A schematic diagram of a real time system (RTS) for HERA. The RTS starts a night of observation by first ingesting the data and loading meta data into the Librarian database (section 3.3.1). An initial observing report is emailed and posted to internal Slack channels for quick inspections of autocorrelations and input signal levels. As this information is shared with the HERA collaboration, bad antenna detection metrics (red, section 3.3.2) are then applied to the raw input data. Redundant calibration algorithms (magenta, section 3.3.3) are then used to derive the per antenna complex gain solutions. Each redundant calibration stage produces their own metrics to check the performance of the calibration algorithms. Finally, a first pass at RFI flagging is performed using the techniques discussed in section 3.3.4. Metadata from the observation is stored in the on-site Librarian database and data products are copied to NRAO for further processing. Daily Jupyter notebooks containing visualizations of the metrics from each stage in the RTS are stored on GitHub.

tion (EoR) and include The Hydrogen Epoch or Reionization Array (HERA; [DeBoer et al. 2016](#)), the Low Frequency Array (LOFAR; [van Haarlem et al. 2013b](#)), the Murchison Wide-Field Array (MWA; [Tingay et al. 2013](#)), and the Precision Array for Probing the Epoch of Reionization (PAPER; [Parsons et al. 2010](#); [Ali et al. 2015](#)).

In this Chapter, we focus on the development of an RTS for the Hydrogen Epoch of Reionization Array (HERA), a low frequency, highly redundant non-steerable array located in the Karoo Desert, South Africa at the SKA Radio Reserve. Once fully constructed, HERA will consist of 350, 14 m diameter dishes arranged in a highly redundant close packed configuration ([DeBoer et al. 2016](#); [Dillon & Parsons 2016](#)). With full stokes correlation, HERA will produce ≈ 500 TB of raw visibility data for every night of observation. Given the practical constraints of on-site data storage and network bandwidth, an RTS is necessary to compress the data so that more than a few days of data can be recorded.

In addition to data volumes, HERA is also in the challenging position of taking science data as the array is being incrementally constructed. Keeping track of the state of the telescope and providing quick feedback about instrumental issues to on-site engineers for quick debugging is critical for maintaining the integrity of the data. Otherwise, faulty hardware could lead to corrupted and unusable datasets. In this way, HERA’s RTS will require low latency detection of hardware failures (the first form of an RTS) in order to eliminate data corruption during data reduction/compression (the second form of an RTS).

During the initial stages of HERA, data volumes are modest, amounting to ≈ 10 ’s of terabytes for the entire observing season. Therefore, the full RTS described above is not needed, and instead is used to characterize the instrument through performance metrics and antenna based calibrations. Since HERA is being incrementally constructed, it is critical to develop an initial RTS to learn about the instrument as quickly as possible. The data products derived from the first iteration of HERA’s RTS will be used for future iterations.

As a guiding principle, HERA’s RTS uses the least amount of information to derive its data products. As a new array, the a basic understanding of HERA’s performance is not yet known, therefore this philosophy underlies the entire system. Confidence in the current knowledge of the instrument is bootstrapped upon in the RTS. The fundamental starting point of for HERA are the antenna positions. Using antenna positions and derived baseline redundancies, the first data products are produced. Once confidence is built with these initial RTS outputs through pre-defined metrics, these data products can be incorporated back into the RTS for further processing. Using this methodology, HERA is iteratively characterizing itself.

In this Chapter, we will focus on the development of an RTS for highly redundant, non-steerable, low frequency radio arrays. Using HERA as a worked example, we will show how the use of redundancy discussed above can be used in an RTS and discuss the technical challenges of the experiment. Section 3.2 introduces the data set that will be used in this work. Section 3.3 will describe the necessary steps in HERA’s RTS. We will also describe future subsystems to include in the RTS. Finally, we’ll conclude in 3.4 with future prospects of an RTS system.

3.2 Observations

For the analysis presented in this Chapter, we use HERA data from two configurations. We use HERA data taken from a night of observation on October 15, 2017 (JD=2458042). The array consisted of 47 antennas during this observation as shown in figure 3.1. We flag antennas 0, 2, 11, 50, 98, 122, and 123 as they exhibit characteristics of being non-operational as discussed in section 3.3.2. The baseline visibilities span a frequency range from 100 MHz to 200 MHz with a channel width of 976.56 HZ. Each integration covers 10.7 seconds and the LST range for the entire data set spans 18:03:36 Hours to 06:58:48 Hours. We also use HERA data from January 1, 2018 (JD=2458140) in Figures 3.3 and 3.4. Antennas are not flagged in this data set. The first season of observations from HERA is currently undergoing an in-depth analysis.

3.3 A Real-Time System

In this section we discuss an approach to an RTS for non-steerable low frequency redundant arrays. We will use HERA as an example of such a system and describe the steps used in its RTS. We will present the key steps in HERA’s RTS pipeline. These stages and algorithms will provide examples of what an RTS can accomplish for a low frequency redundant radio array. The key stages of HERA’s RTS are shown in Figure 3.2. The first step is data ingestion, which involves storing and recording observations that can be tracked as data flows through the system. The next step involves detecting bad antennas. This is a critical step because it first gives us information about how the array is performing, which can then be relayed to on-site engineers. This also provides us with a clean data set to perform further processing. A per antenna relative calibration is performed using known information about the instrument as the third step. The final step in the RTS detects and flags radio frequency interference (RFI).

Since we are describing the first version of HERA’s RTS, we stop the real time analysis at RFI detection and flagging. Further stages of the RTS could include absolute calibration and baseline averaging for data compression. However, because the instrumental performance and models of the HERA sky had not been well characterized yet, these algorithms were not incorporated into the RTS. The data products from this version of the RTS were used to characterize baseline redundancy and determine models of the sky for future iterations of the RTS for HERA. We discuss absolute calibration at the end of this section.

3.3.1 Data Ingestion and Tracking

The HERA RTS is closely coupled to the “Librarian,” the project’s data management subsystem¹. The Librarian is the primary means by which the project stores, transfers, and queries its data products. Briefly, the Librarian system is composed of a loose confederation

¹<https://github.com/HERA-Team/librarian/>

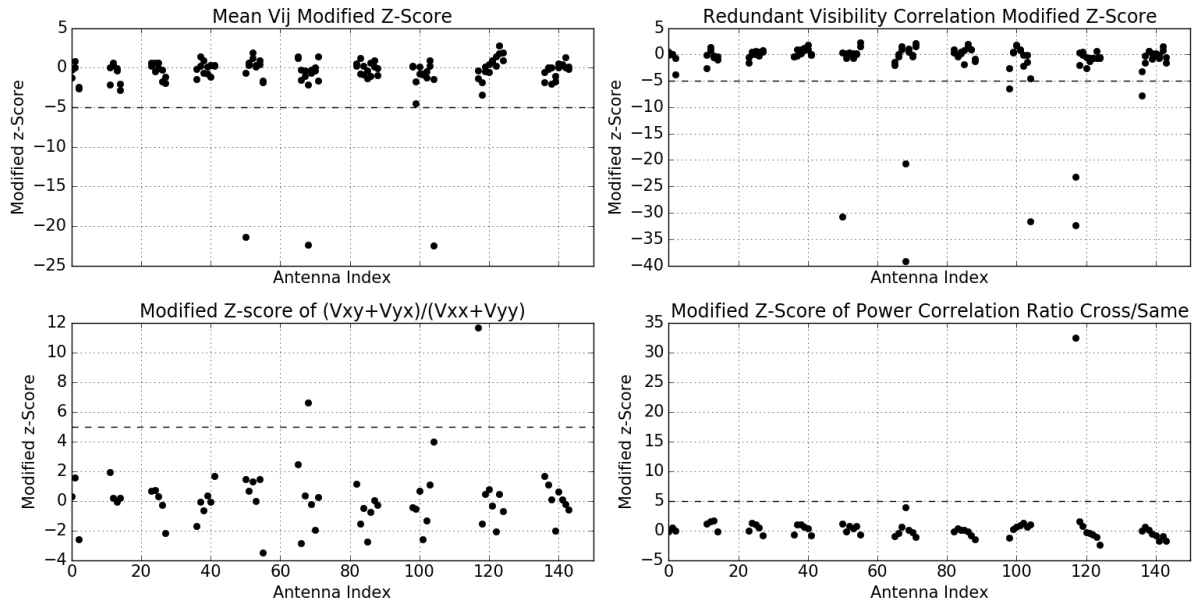


Figure 3.3: The modified Z-scores of each of the metrics described in section 3.3.2. The metrics are applied to a 10 minute window of data during observing (JD=2458140.35680). The mean visibility metric (top left) flags antennas 50, 68, and 104 since they are below our manually chosen threshold (dotted line) for flagging outliers. These antennas have very low amplitude compared to the others. However, the correlation metric (top right) additionally flags antennas 117 and 136, with antennas 98 and 68 on the border line for this integration. These antennas do not have amplitude correlations that are similar to others. In both cross polarization metrics in the bottom row indicate that the antenna 117 is actually cross polarized.

of data repositories that communicate over the Internet. Each repository includes data storage nodes, a database of file records and metadata, a Web interface for user interaction, and a Web-based JSON API for programmatic control. Three key design principles of the Librarian are:

- Files are uniquely identified by their names.
- Files are immutable.
- File metadata are decoupled from file data: each repository may contain zero, one, or many copies (“instances”) of each file.

Each repository operates independently, but can synchronize file metadata and file instances to other sites.

After each night of observing concludes, raw correlator data are automatically ingested into the on-site Librarian repository. (Bandwidth limitations on the correlator hardware precludes real-time ingestion.) Metadata are mostly derived from the file contents themselves, although some items are obtained from the HERA monitor-and-control (M&C) system. As the RTS operates, derived data products are also ingested into the Librarian, and a “standing order” causes these products to be automatically synchronized to a Librarian repository operating at NRAO. A standardized file naming convention makes it straightforward to associate RTS products with the raw data from which they were derived.

3.3.2 Bad Antenna Detection

The first step in an RTS after data ingestion, is to check the quality of raw data from the telescope. Data quality checks look for obvious errors such as antennas being dead (loss of power), errors in observing due to correlator hardware failures, and other telescope specific checks. In particular, for telescopes with a large number of antennas, keeping track of antennas that are failing versus performing well is a fundamental metric that needs to be monitored. For HERA, we check whether antennas have malfunctioned due to loss of power, analog hardware failures, correlator errors, and unexpected new errors that manifest themselves as being out of the ordinary. As shown in the RTS work-flow diagram in Figure 3.2, an initial quality check is the first step to further analysis.

Finding failing antennas is particularly critical because they touch every other antenna in the system via baseline visibilities. Without the excision of bad antennas, further analysis steps would produce invalid results. The fact that HERA is an incrementally built instrument, where new antennas are brought into the array continuously throughout observations or are being modified and adjusted daily, makes this a challenging effort. In order to detect antenna and correlator outages, HERA uses raw visibilities in the RTS pipeline to calculate numerous metrics that can be used to monitor and identify *bad antennas* and provide a path for relaying information quickly to rectify the issue with onsite engineers.

In the rest of this section we describe the handful of metrics that are implemented in HERA’s RTS to identify dead or cross polarized antennas. These metrics were first derived

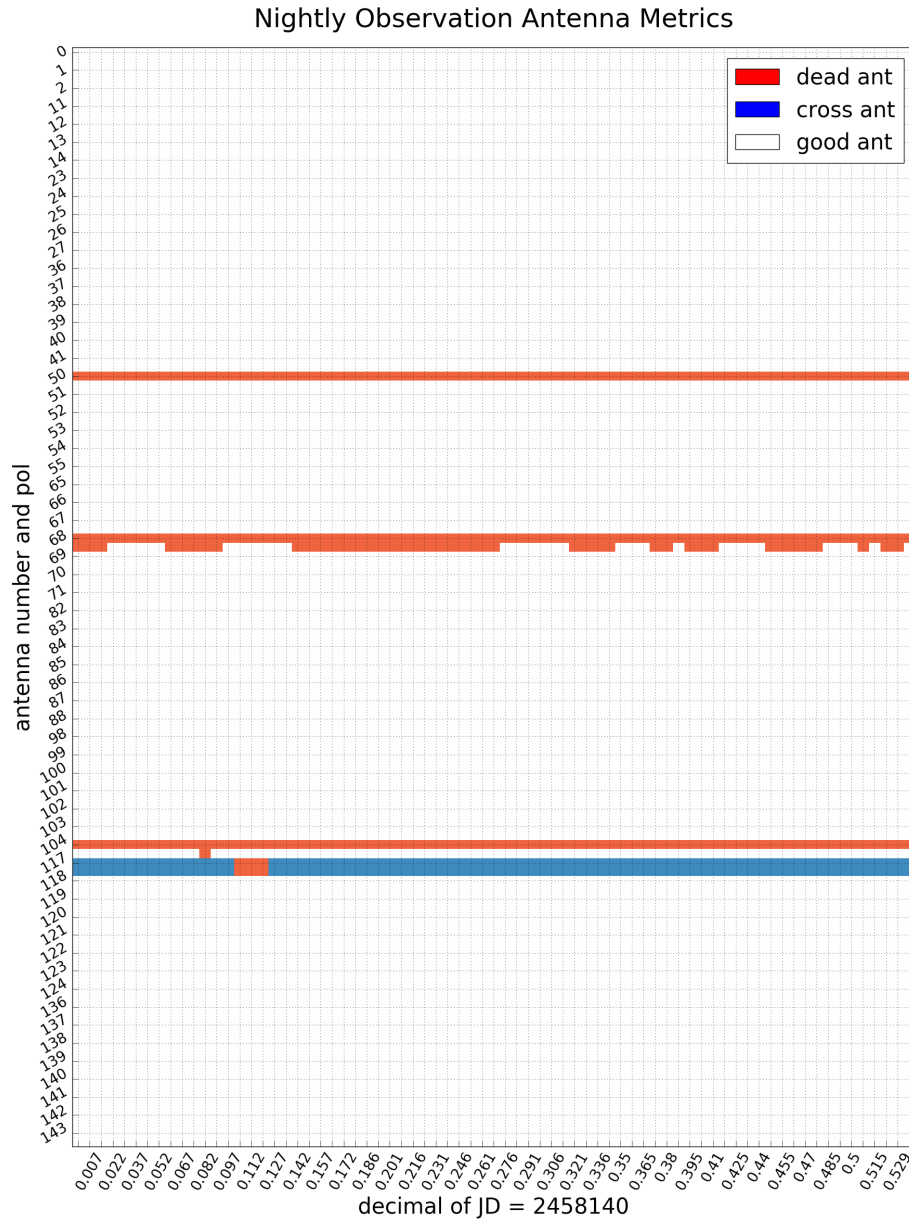


Figure 3.4: Antenna based metrics for a nights worth observation condensed into a single image. For a given observing night, HERA’s antenna metrics are calculated on a 10 minute cadence (x-axis) and for each antenna-polarization (y-axis; for readability, only antenna numbers are shown with each tick representing the east polarization and the space between ticks represents the north polarization of the previous tick) it is determined whether the antenna is a good antenna (green), a dead antenna (red) according to the mean visibility amplitude metric (section 3.3.2) or the correlation metric (section 3.3.2), or a cross polarized antenna (blue) according to the cross polarization metrics (section 3.3.2). For this particular night (JD=2458140) we find that antennas 50, 68, and 104 are bad antennas. Antenna 117 is a cross polarized antenna (however it does get labelled as a dead antenna for part of the night). These are corroborated by the per file metrics shown in figure 3.3.

from the analysis of the 128 antenna iteration of the PAPER experiment and as such, they can be adapted for use on any low frequency redundant array. We also show how these metrics perform on real HERA data from commissioning observations.

Mean Visibility Amplitude Metric

The most critical and likely failure that an antenna can have is that it has no (or low) signal, which is usually caused by a loss of power somewhere in the signal chain or a broken amplifier. This failure mode is characterized with unusually low signal amplitude from the antenna and so the visibilities associated with that antenna will have a lower than average amplitude. This leads to the definition of the mean visibility metric for antenna i :

$$M_i = \frac{\left(\sum_{j, i \neq j, \nu, t} |V_{ij}| \right)}{N - 1}, \quad (3.1)$$

where V_{ij} is the visibility between antennas i and j , N is the number of antennas in the array, and the sum is taken over all antennas j (s.t. $i \neq j$), times t , and frequencies ν . When M_i is compared across the array, it reveals antennas with anomalously low signal to noise ratios (SNR).

The Correlation Metric

In addition to looking at a total power metric, one can also appeal to array redundancies for identifying low power antennas. In this metric, non-redundancies in SNR between nominally redundant baselines can be traced back to an offending antenna that have a significantly lowered the redundancy. Let us define the set $\{i_j\}$ as the set of redundant baseline visibilities redundant with v_{ij} . The correlation metric can then be defined as

$$C_i = \sum_j \sum_{v_{i,j}, v_{k,l} \in \{i_j\}} \frac{\text{Med}_\nu \left(\left| \frac{\sum_t v_{i,j} v_{k,l}^*}{N_{\text{times}}} \right| \right)}{\text{Med}_\nu \left(\frac{\sum_t |v_{i,j}|^2}{N_{\text{times}}} \right) * \text{Med}_\nu \left(\frac{\sum_t |v_{k,l}|^2}{N_{\text{times}}} \right)} \quad (3.2)$$

where $v_{i,j}$ and $v_{k,l}$ are two redundant baselines, the inner sum is taken over some time interval with N_{times} integrations, Med_ν is the median over frequency, the first outer sum is taken over all baselines of the same type as b_{ij} , and the final sum is over antennas j such that $i \neq j$. Since the visibilities are not calibrated, C_i is a notion of the aggregated amplitude correlation coefficient for antenna i . Therefore, amplitude redundancies provide a way of accessing which antennas are bad (exhibiting low amplitudes), by highlighting those antennas with low power.

Cross Polarization Metrics

The metrics described above look for antennas with abnormally low power and attribute this power to power loss or non functional amplifiers in the signal chain. Another reason for

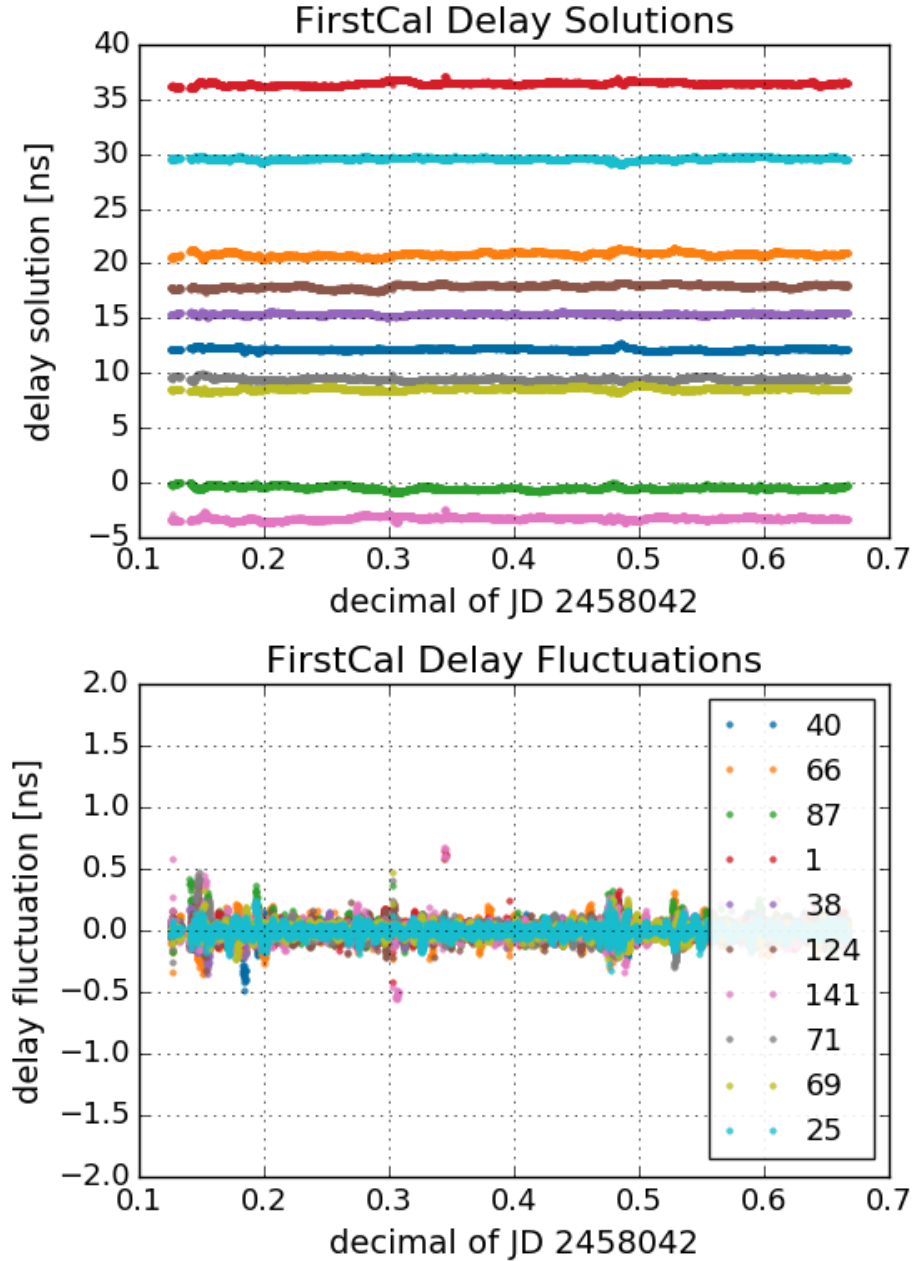


Figure 3.5: Per antenna delay solutions (top) and their fluctuation about the median delay (bottom) from a night of observation from HERA for calibration solutions derived from the **FIRSTCAL** algorithm. Only a quarter of the antennas and a single polarization are plotted here for clarity. The delays themselves appear to be fairly constant throughout the night with deviations below a nanosecond across the band. These delays redundantly calibrate the antennas to each other.

low visibility amplitudes is due to cross-polarized antennas, which occurs when the feed is rotated 90° with respect to the other feeds or when cable paths get swapped for polarizations along the signal chain. This type of error causes the linear polarization visibilities (EE or NN) to have a lower amplitude or correlation relative to the cross-polarizations (EN and NE) because cross-polarized visibilities have 10% the signal-to-noise (Dillon et al. 2017). Therefore, using the mean amplitude visibility and correlation metric (equations 3.1 and 3.2) on EE or NN visibilities individually may indicate bad antennas, but it may be the case that the antennas are actually cross-polarized. Metrics used to detect such a problem can be based on both the mean visibility and correlation metric.

Using equation 3.1, we define the mean visibility cross-polarization metric as

$$P_i = (M_i^{NE} + M_i^{EN}) / (M_i^{NN} + M_i^{EE}), \quad (3.3)$$

where M_i is the mean visibility metric defined in equation 3.1 and are calculated for all the polarization pairs. If P_i is larger than some threshold, then antenna i may be cross-polarized. This method is effective when applied to long baselines, but for the shortest baselines in HERA, large-scale astrophysical polarization (e.g. Lenc et al. 2016) is observed in all the instrumental visibilities (since, for example, the NN visibility is equal to the sum of the Stokes I and Q visibilities; c.f. Moore et al. (2013)). This brings P_i closer to unity for any antenna i .

Similar to equation 3.3, we can also define the same metric except with C_i instead of M_i :

$$PC_i = (C_i^{NE} + C_i^{EN}) / (C_i^{NN} + C_i^{EE}), \quad (3.4)$$

where C_i is defined in equation 3.2 and are calculated for all the polarization pairs. In the limit of low Faraday Rotation Measures along the line-of-sight (as found in the Southern Hemisphere by Bernardi et al. 2013), and bright linearly polarized foregrounds this metric suffers from the same limitations as the mean visibility cross polarization metric.

Application of metrics to HERA

Figure 3.3 shows the applications of the metrics described in section 3.3.2 to 10 minutes of HERA data from the RTS pipeline. Rather than using the metrics directly, we use the modified Z-score. Traditionally, the Z-score is defined as the number of standard deviations from the mean and is a standard method to detect outliers. However, since our metrics are not necessarily distributed normally we use the modified Z-score defined as

$$\bar{Z}_i = 0.6745 \frac{|x_i - \tilde{x}|}{\text{MAD}}, \quad (3.5)$$

where x_i is some data point (in our case this would be the metric value), $\tilde{x}$ is the median of $\{x_i\}$, and MAD is the median absolute deviation. The proportionality constant makes the modified Z-score equivalent to the standard Z-score for Normally distributed data. The authors in (Iglewicz & Hoaglin 1993) recommend that modified Z-scores with an absolute

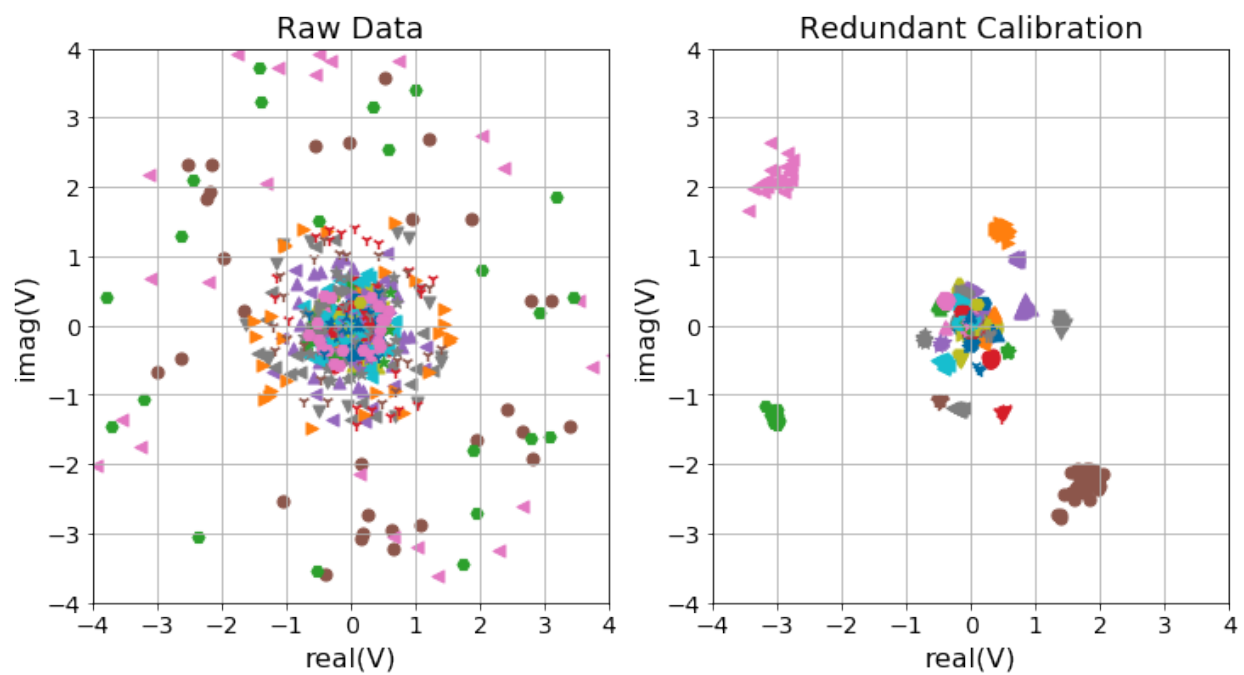


Figure 3.6: Complex visibilities for all baselines before (left) and after (right) redundant calibration when the galactic center is nearly at zenith. Each color and symbol pair denotes a different redundant baseline spacing and orientation. Before redundant calibration visibilities are uncalibrated and therefore have random phases with respect to one another and therefore their distribution is random in the real-imaginary plane. However, after calibration clusters of redundant baselines emerge and baselines that should be redundant begin to agree.

value greater than 3.5 should be flagged as outliers. In our analysis, we take the cutoff to be $Z_i = 5$ to account for greater variation in our metrics. Additionally, when determining if an antenna should be flagged or not, we average the Z-scores for the mean amplitude visibility metric and the correlation metric together before applying the cutoff threshold.

The top row of figure 3.3 show the modified Z-scores for the mean amplitude visibility metric and correlation metric defined in equation 3.1 and 3.2, respectively. We find that antennas 50, 68, and 104 are bad antennas according to the mean visibility metric and antennas 50, 68, 104, 117, and 136 are bad antennas with 68 and 98 being on the borderline of the 5 sigma threshold. Even though antennas 68 and 98 are on the border line, because we average the metrics together prior to applying the $Z_i = 5$ cutoff, only antennas 50, 68 and 104 are considered dead antennas.

Additionally, the cross polarized metrics in the bottom row of Figure 3.3, indicate that antenna 117 is actually cross-polarized and not a dead antenna. Even though the correlation metric (top right panel in Figure 3.3) determines that antenna 117 is a bad antenna, this is actually a false positive. The correlation metric is susceptible to the false positives for antennas that are actually cross-polarized, which follows from the fact visibilities associated with the cross polarized antenna will have a low amplitude due to the low correlation between polarizations. The cross-polarized metrics are necessary to determine whether antennas determined to be dead from the correlation metric are actually bad or cross-polarized.

Figure 3.4 shows an image of an entire night of condensed antenna metrics. The binary flags from applying the antenna metrics over every observation (where an observation for HERA is 60, 10.7 second integrations) Antenna-polarizations are plotted against the observations (where an observation for HERA data consists 60, 10.7 second integrations) for the night. For each observation, an indication of whether an antenna is good (green), dead (red), or cross-polarized (blue) is plotted. For this particular night, JD=2458140, we find that antennas 50e (East polarization), 68e, and 104e are flagged as dead antennas for the entire night. However, antennas 68n (North polarization), and 104n are flagged for only parts of the night, indicating that they are on cusp of failing. Figure 3.4 provide a condensed way of looking at all the antennas for a given nights of observation. HERA produces these plots after every nights observations which is then shared through emails to the collaboration and posts to Slack channels. In this way, problems of dead and cross-polarized antennas can be seen broadly within the collaboration and engineers on site can investigate and rectify the situation, if possible.

3.3.3 Redundant Calibration

After an initial flagging of bad antennas using the metrics described in section 3.3.2, or some other metrics, antenna calibration can begin. Calibration is a necessary step to remove instrumental and atmospheric effects from measurements to ensure that only true signal is being observed. For interferometers, effects that need to be calibrated out of visibility measurements include antenna based electronic delays, electronic bandpass structure, and the ionosphere. Each of these effects can be frequency and time dependent in nature.

HERA’s real-time calibration scheme is built upon two related philosophies: traceability and the cumulative solving of solutions. The traceability of calibration solutions is necessary so that unwanted features in calibration solutions can be traced back to a specific cause, possibly originating from a specific baseline, antenna, or at a specific time and frequency. These features can occur for many different reasons, including failure of the calibration algorithm due to unexpected inputs and thus introducing structure to the calibration solutions, or a mis-calibration at specific times and frequencies due to radio frequency interference (RFI) or other causes. Tracing back specific features in calibration solutions, and therefore calibrated data, to the offending piece of data allows the determination of bad data and the excision of it.

The second philosophy is that of cumulative solutions. Cumulative calibration solutions are those which have been built upon from previous calibration solutions. In this philosophy, initial calibration solutions are first solved for by using the least amount of information possible, for example only using theoretical antenna positions. The initial solutions are derived for every unit of data that exists (per antenna, time integration, and frequency channel). These solutions are kept independent as long as possible. Once confidence is built in the initial solutions, outside information can slowly be brought in to refine the solutions. For example, we can then begin to average solutions in frequency given some spectral coherence scale defined by the science objective. This hierarchy of solutions provide natural checkpoints for deriving and refining calibration solutions.

Generically, calibration of a non-steerable radio array with nothing known about the instrument, like HERA, is a non-trivial process. It depends on a number of factors such as where the array is built, the synthesized beam of the instrument, and the sources in the sky at the declination of the array. Traditionally, self-calibration is used to determine antenna based complex gain calibration solutions using the imaged source itself. This is a path forward for an instrument with good imaging capabilities. However, for redundantly spaced arrays which are not built for producing images, self-calibration can be challenging. Instead, redundancy provides an axis to derive first order calibration solutions by applying the constraint that redundant baselines of the same length and orientation should measure the same thing.

Redundant calibration techniques rely on a grid configuration of the array to derive a per antenna complex gain solution by comparing redundant measurements (Dillon & Parsons 2016). Redundancy based calibration algorithms are information-light, requiring only antenna positions (and in most cases, just the theoretical antenna positions) to derive gain solutions. This makes it an ideal algorithm for an RTS pipeline since antennas can be quickly calibrated relative to one another, leaving only a handful of degenerate parameters (Liu et al. 2010; Zheng et al. 2014, 2017a; Dillon et al. 2017) to be solved for using outside information.

For the rest of this section, we describe the redundant calibration algorithms used for HERA and present the calibration results. Redundant calibration for HERA is based on OMNICAL ², a package developed for the MitEOR (Zheng et al. 2014) experiment and is

²<https://github.com/HERA-Team/omnical>

based on the algorithms presented in (Liu et al. 2010). In addition to OMNICAL, a description of the FIRSTCAL algorithm is also presented, which is required to have a starting position for the OMNICAL algorithm to proceed. We finally end with a discussion of the application of redundant calibration to HERA data.

For a deeper discussion of redundant calibration, we point the reader to Wieringa (1992); Liu et al. (2010); Sievers (2017); Grobler et al. (2018). Redundant calibration has been used with a number of low frequency interferometers including PAPER (Parsons et al. 2014; Ali et al. 2015), MitEOR (Zheng et al. 2014), and MWA (Li et al. 2018).

Firstcal

As alluded to above, before using the full power of OMNICAL, a rough calibration is required. This arises due to phase ambiguities present in OMNICAL. As discussed in (Zheng et al. 2014), phase wraps between nominally redundant baselines can cause gain solutions to fall into false minima, causing spectral structure (Dillon et al. 2017). There are various ways to remedy this problem, most of which involve performing an initial rough calibration using global sky models (Zheng et al. 2017a) or imaging techniques. HERA performs the initial calibration by using an algorithm called *firstcal*, which uses redundancies present in the array to determine a per antenna delay that unwraps the redundant baselines with respect to one another.

Holding to the philosophy of using the least amount of information to derive the initial calibration solutions, *firstcal* derives a per antenna delay using ratios of redundant baselines. These per antenna delays are attributed to light travel times through cables of different lengths and analog electronics. Additionally, these delays are the dominant component of the phase structure present in antenna gain solutions. In order to see how FIRSTCAL works, consider an array with five antennas arranged on a linear grid with antenna numbers $\{1,2,3,4,5\}$. Each pair of antennas form a baseline visibility,

$$V_{ij}(\nu) = g_i(\nu)g_j(\nu)V_{ij}^{true}(\nu) + n_{ij}(\nu),$$

where $g_i(\nu)$ and $g_j(\nu)$ are frequency dependent complex gains for antennas i and j , V_{ij}^{true} is the true underlying visibility measured by the baseline formed between antennas i and j , and $n_{ij}(\nu)$ is the noise on the measurement, which is taken to be normally distributed. For our toy example, sorting visibilities into redundant groups we have the following set of redundant visibilities, $\{V_{12}, V_{23}, V_{34}, V_{45}\}, \{V_{13}, V_{24}, V_{35}\}, \{V_{14}, V_{15}\}$. Within a redundant group, we form ratios of visibilities, for example,

$$\frac{V_{12}}{V_{34}} = \frac{g_1 g_2^* V_{12}^{true}}{g_3 g_4^* V_{34}^{true}} = \frac{g_1 g_2^*}{g_3 g_4^*}, \quad (3.6)$$

where we have dropped the frequency dependence for clarity and used the fact that the true visibilities are equal for redundant baselines. We are left with a ratio of frequency dependent complex antenna gains. Up to first order, complex gains can be written as $g_i(\nu) = A_i(\nu)e^{-2\pi\tau_i\nu}$, where $A_i(\nu)$ is the frequency dependent gain amplitude and τ_i is the frequency

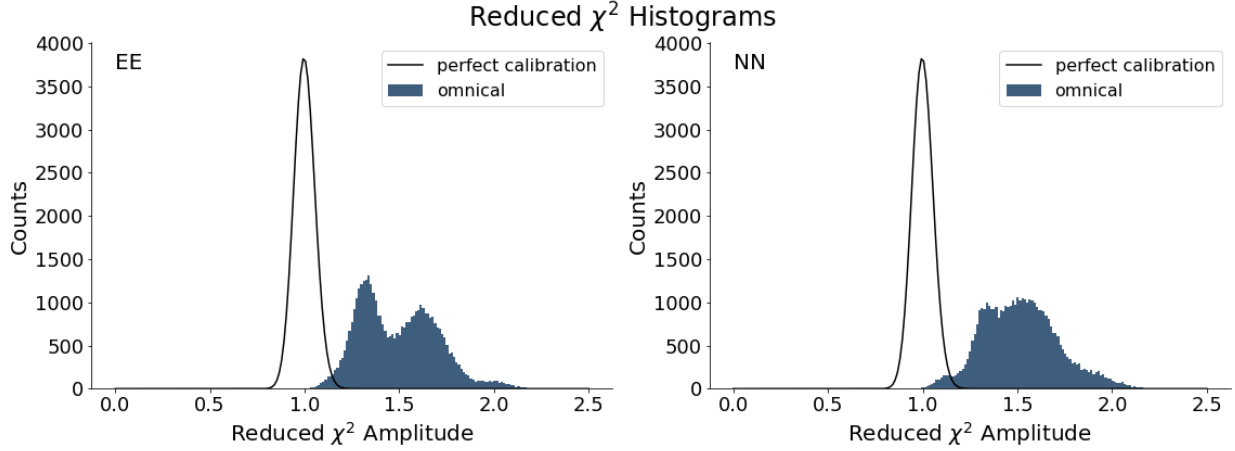


Figure 3.7: Histogram of the reduced χ^2 for the North (left) and East (right) polarizations for the OMNICAL run including all baseline separations. If OMNICAL perfectly calibrated the data we would end up with a χ^2 distribution (black) with mean of 1 and variance $2/\text{DoF}$, where DoF is the degree of freedom for the fit which is equal to 647 for this example. Both histograms show evidence for a bimodal distribution indicating that there are two groups for the deviation from redundancy. The mean of the distribution is 1.49 (1.51) and the variance is .042 (.0405) for the East and North polarizations. The means greater than 1 indicate that there is considerable non-redundancy. This can be attributable to differences in baselines, antenna beam variations, and non-redundancies in the dish and feed structures.

independent per antenna delay. Substituting into Equation 3.6 and looking at the phase component, we see that

$$\text{Arg}\left(\frac{V_{12}}{V_{34}}\right) = -2\pi\nu(\tau_1 - \tau_2 - \tau_3 + \tau_4). \quad (3.7)$$

Thus, ratios of redundant visibilities is equivalent to a difference in antenna delays. To determine the full sum on the right hand side of Equation 3.7, *firstcal* takes the Fourier transform of equation 3.6 and finds the peak delay (which is the sum of the antenna delays involved in the ratio) after normalizing the amplitude to remove effects of radio frequency interference (RFI). Once the peak delay is found, it is removed and a linear slope is fit to the residual phase, further refining the delay to sub pixel resolution of delays.

Going through this algorithmic process of finding the sum of the delays as described above for every ratio of redundant baselines (which goes as the quartic power of the number of antennas), we build up a system of equations and solve for the per antenna delay using linear least squares. For the toy five antenna array described above we end up with the

following system of equations:

$$\begin{bmatrix} 1 & -2 & 1 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 \\ 1 & -1 & 0 & -1 & 1 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 1 & -2 & 1 \\ 1 & -1 & -1 & 1 & 0 \\ 1 & 0 & -2 & 0 & 1 \\ 0 & 1 & -1 & -1 & 1 \\ 1 & -1 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \end{bmatrix} = Arg \begin{bmatrix} V_{12}/V_{23} \\ V_{12}/V_{34} \\ V_{12}/V_{45} \\ V_{23}/V_{34} \\ V_{23}/V_{45} \\ V_{34}/V_{45} \\ V_{13}/V_{24} \\ V_{13}/V_{35} \\ V_{24}/V_{35} \\ V_{14}/V_{25} \end{bmatrix} \quad (3.8)$$

This is a solvable system of equations, it is insensitive to 3 degenerate parameters (Dillon et al. 2017; Li et al. 2018). One of these degeneracies corresponds to an over all delay: a delay can be added to every antenna and not change the solution as the overall delay drops out in Equation 3.7. A phase gradient across the array in the x or y direction can also be added to each antenna which the ratios of visibilities would not be sensitive to. *Firstcal* makes no attempt to fix these degeneracies since they are a subset of *omnical* degeneracies and therefore unnecessary for performing *omnical*. Additionally, **FIRSTCAL** is performed one polarization at a time, so the relative x-y delay is also undetermined. Refer to (Dillon et al. 2017) for a detailed discussion on degeneracies with *firstcal*.

It is important to note that **FIRSTCAL** only solves for a per antenna delay and does not try to solve for the relative frequency dependent amplitude of the antenna gains. Even though possible within the **FIRSTCAL** formalism, we make no attempt to fit the amplitudes as to minimize the number of free parameters we are solving for. **FIRSTCAL** gets the phase of the complex gains close enough to values such that redundant baselines look similar in phase, **OMNICAL** then takes over to refine this solution in both phase and amplitude.

The **FIRSTCAL** algorithm is highly susceptible to bad data. If a dead antenna found it's way into the input data, redundancy would be broken between redundant baselines involving the offending antenna and those without. This manifests itself as determining wildly noise delays as a function of time for all antennas. The use of a good antenna detection algorithm is necessary for **FIRSTCAL** to produce high quality delays.

Omnical

Once **FIRSTCAL** has derived a per antenna delay, all antennas are now relatively calibrated to one another. Specifically, redundant baselines agree in phase to within a few radians. Using **FIRSTCAL** gains³ as an initial guess of the complex gain solution, we can further refine the calibration solutions using **OMNICAL**. **OMNICAL** independently solves for antenna gains per

³**FIRSTCAL** gains for a given antenna, i , are defined as $g_i = e^{-2\pi i \nu \tau_i}$, where τ_i is the derived delay solution in seconds and ν is frequency in Hz

time and frequency. For a thorough explanation of OMNICAL we refer the reader to Liu et al. (2010); Zheng et al. (2014). In the rest of this section we discuss the outputs of OMNICAL.

For every frequency and time, OMNICAL solves for g_i , g_j , and y_{i-j} that minimize

$$\chi^2(\nu, t) = \sum_{ij, i \neq j} |v_{ij} - g_i g_j^* y_{i-j}|, \quad (3.9)$$

where $v_{ij}(\nu, t)$ is the measured visibility, $g_i(\nu, t)$ and $g_j(\nu, t)$ are gains for antenna i and j , respectively, and $y_{i-j}(\nu, t)$ is the true correlation between voltage signals measured from antennas i and j . Equation 3.9 is a measure of the deviations from redundancy. To determine whether these deviations are due to noise or some other systematics, Zheng et al. (2014) defines the chi-squared per degree of freedom as follows:

$$\chi^2 = \frac{1}{DoF} \sum_{ij, i \neq j} \frac{|v_{ij} - g_i g_j^* y_{i-j}|}{\sigma^2}, \quad (3.10)$$

where the sum is taken over all baseline pairs, DoF is the degrees of freedom in the system of equations, and σ^2 is the per baseline thermal noise variance. In traditional χ^2 analysis, the DoF is equal to the number of observations minus the number fitted parameters. For the set of equations constructed in OMNICAL, the number of observations is equal to the number of baselines, N_{bls} , and the number of fitted parameters is the number of gains ($N_{antennas}$) and the number of unique redundant baseline groups, N_{red_groups} . This produces a DoF equal to $N_{bls} - (N_{antennas} + N_{red_groups})$. If the chi-squared per degree of freedom distribution has a mean of 1 and a variance of $2/DoF$ (Abramowitz & Stegun 1964), deviations from redundancy are noise like. However, if the distribution has a mean greater than one, this indicates that there is a systematic non redundancy that is greater than the noise.

Application of Redundant Calibration to HERA

In this section we present redundant calibration results from the RTS and discuss the performance of the HERA array. Figure 3.6 shows uncalibrated raw HERA visibilities (left) and visibilities after the application of per antenna complex gain solutions (right) derived from OMNICAL when Sagittarius A is nearly at zenith at 149 MHz. Each symbol and color pair indicates a set of baselines that should be redundant with one another. The increased clustering after the application of OMNICAL gain solutions indicates that redundant baselines are nearly measuring the same sky. We find that for long baselines⁴ the deviations from perfect redundancy is $\sim 5\%$ and increases to $\sim 20\%$ for the the shortest baselines, generally. Since positions of the antennas are known to ten centimeters (DeBoer et al. 2016), the observed deviations from redundancy are most likely attributed to variations in antenna

⁴At the observing frequencies for HERA, 100-200MHz, diffuse emission dominates the emission on large scales (de Oliveira-Costa et al. 2008b). HERA's long baselines generally see less diffuse emission and are more sensitive to point source emission, therefore they have lower amplitudes. This is in contrast to short baselines, which see more diffuse emission and have a larger amplitudes.

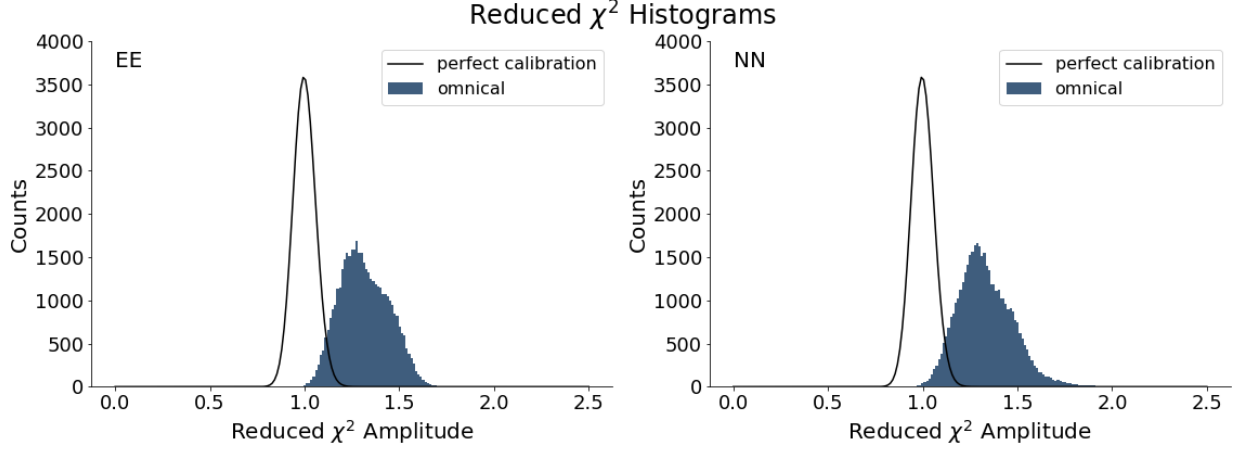


Figure 3.8: Histogram of the reduced χ^2 for East (left) and North (right) polarizations from the OMNICAL run including all baseline separations except the shortest baselines (less than 15 meters). For this iteration of OMNICAL, DoF=567. We see that the bi-modality in the distributions seen in Figure 3.7 are no longer seen here indicating that the shortest baselines have more non-redundancy than the long baselines. However, the non long baselines (greater than 15 meters) still show evidence for non redundancy as indicated by the mean of the distributions being 1.31 (1.33) for the EE (NN) polarization. The variance for the EE (NN) histogram is .0017 (0.021).

beams. Antenna beam variations can occur for a number of reasons, including the precision with which HERA feeds can be placed at the focal point of the dish and the degree to which the dish structure itself can be kept smooth. The antenna beam variations are highlighted most prominently on the shortest baselines since emission from large scales dominate. This potentially highlights the discrepancies in the sidelobes of the beam. The longer baselines do not have as much evidence for non-redundancy, possibly due to the decreased signal to noise ratio (SNR) as the sky is not dominated by point sources for HERA.

To further assess the quality of calibration solutions from OMNICAL, we compute the reduced χ^2 as defined in Equation 3.10 using a noise variance model estimated from auto correlations. Auto correlations provide a good estimate of the noise because they contain the total noise from both the signal chain (receiver noise) and the sky (sky noise). For the baseline formed between antennas i and j , the noise variance is estimated by averaging neighboring times and frequency bins in the autocorrelations of each antenna, i and j , and then multiplying the resultant filtered autocorrelations together. Dividing the result by the frequency channel bandwidth and integration time, which for the data used in this work is 97.65 kHz and 10.7 seconds, respectively, gives us an estimate of the noise variance per baseline, time, and frequency. Specifically, given the autocorrelation visibility for antennas i and j as V_{ii} and V_{jj} , the noise variance estimate is given by

$$\sigma^2(\nu_n, t_m) = \frac{1}{Bt} \frac{V_{ii}(\nu_{n+1}, t_{m+1}) + V_{ii}(\nu_n, t_m)}{2} \times \frac{V_{jj}(\nu_{n+1}, t_{m+1}) + V_{jj}(\nu_n, t_m)}{2}, \quad (3.11)$$

where ν_n and t_m label the n -th and m -th frequency and time bin of the autocorrelation visibilities.

In addition to an estimate of the noise variance, the number of degrees of freedom for the least squares fit is also required to calculate the reduced χ^2 . As discussed in [Zheng et al. \(2014\)](#), the degree of freedom is given by the number of measurements less the number of parameters. In the application of OMNICAL performed in this work, since each time and frequency channel is fitted independently of one another, the number of measurements is the total number of baselines (i.e. the number of cross correlations) and the number of parameters are given by the number of per antenna gains and the number of unique redundant baselines. The number of degrees of freedom is then $\text{DoF} = N_{bls} - (N_{antennas} + N_{redundant-groups}) = 647$ for the array described in [Figure 3.1](#).

[Figure 3.7](#) shows a histogram of the reduced chi-square calculated using [Equation 3.10](#) and [3.11](#) for an observation of HERA data taken on julian date 2458042 using all of the baselines. The mean of the reduced χ^2 is 1.49 (1.51) and has a variance of .042 (.0405) for EE (NN) polarizations. As discussed previously, for a perfect calibration, the reduced χ^2 distribution should have a mean of 1 and variance of $2/\text{DoF} = .0031$. A mean larger than 1 implies that there is considerable non-redundancy. Factors such as misplacement of antennas and per antenna beam variation contribute to this non-redundancy. This non-redundancy is paralleled in [Figure 3.6](#). Empirically, we see that the reduced χ^2 has a bimodal distribution indicating a difference in the behaviors of either certain times, frequencies, or baselines.

In an effort to find the cause of the bimodal nature of the reduced χ^2 distribution in [Figure 3.7](#), we excluded the shortest baselines before performing OMNICAL. Specifically, we excluded baselines that are redundant with baselines (12, 13), (1, 12), and (1, 13) in [Figure 3.1](#). Since the total number of measurements and parameters are reduced, the DoF has changed to 567. [Figure 3.8](#) shows the resultant reduced χ^2 histogram. The mean of the reduced χ^2 using only long baselines (> 14 m) is 1.31 (1.33) and the variance is .017 (.021) for the EE (NN) polarization. The resulting distribution does not empirically look bimodal, indicating that there is a difference in the redundancies of the short and long baselines of HERA. These discrepancies can be attributed to the same reasons the deviations from redundancies for long and short baselines in [Figure 3.6](#). Even though the distribution does not empirically look bimodal, there is still evidence for considerable deviations from redundancy indicated by the mean and variance.

[Figure 3.9](#) shows another slice of the reduced χ^2 as a function of eight different contiguous frequency ranges, each spanning ~ 10 MHz of bandwidth. For all bands, the mean of the reduced χ^2 distribution is greater than 1 indicating that there is non-redundancy. However, for bands 1 (119-129 MHz) and 4 (148-158 MHz), the variance of the distribution is close to

the expected variance of $2/\text{DoF} = .0035$.

To quantify the degree to which nominally redundant baselines are not redundant, we use the fractional error of baselines from their true visibilities (y_{i-j}) as defined in equation 3.9. For each baseline, we compute the fractional error defined as

$$\left(\frac{\delta v}{v}\right)_{i,j} = \frac{|v_{i,j} - y_{i-j}|}{|y_{i-j}|}, \quad (3.12)$$

where $v_{i,j}$ is the OMNICAL calibrated visibility between antennas i and j , and y_{i-j} is the true visibility for baseline (i, j) solved for in OMNICAL. Since y_{i-j} are the true visibilities, they should be similar to the calibrated baselines. Therefore Equation 3.12 determines the degree to which a baseline is not-redundant with respect to the model. Figure 3.10 shows the fractional error for four chosen baseline lengths and orientations. We find that for the shortest baselines, fractional errors range from 25% to 50% and contain a lot of spectral structure introduced by the true visibilities. However, for longer baselines, we find less spectral structure and fractional errors hovering around 25% across the band. This indicates that short baselines show evidence for more non-redundancy than longer baselines.

3.3.4 RFI Excision

The stages after calibration in an RTS are left for further metrics and analysis. A number of directions can occur at this point including data compression, imaging, and further science analysis. For HERA, the last stage in the RTS is to detect and flag RFI. Prior to redundant calibration, RFI excision is not required since all frequency channels and time integrations are independently calibrated. Therefore, visibilities with RFI, are going to have poor calibration solutions, and will not effect neighboring solutions. However, there may be a situation where some baselines have RFI contaminated channels which then gets propagated to the per antenna gain solutions. One might argue that this calls for implementing RFI flagging prior to OMNICAL, but if a channel is dominated by RFI on a few baselines it is highly likely that all baselines have the same RFI, but at a lower amplitude and hence the gain solution for that channel is not trustworthy. The contaminated channel will need to be flagged prior to performing further analysis. For such channels, a higher chi-squared value will occur, indicating a bad fit. Therefore, RFI excision is left to after redundant calibration and uses outputs of OMNICAL, such as the gains and chi-squared, as inputs to the RFI flagging algorithm.

Numerous automatic RFI detection techniques exist in radio astronomy. Many of these methods operate on visibilities and use thresholding techniques. Depending on the morphology of RFI, highly specialized techniques can be used. For example, RFI can be localized to individual frequency channels or time integrations, and therefore edge detection of the visibilities can be used. Also, wide band RFI effecting many frequencies and times, localized thresholding can be used. In addition to automatic detection techniques, it is almost always necessary to manually flag to get low level RFI. Finding the highest fraction of RFI contaminated data without over flagging is the goal of RFI detection and excision.

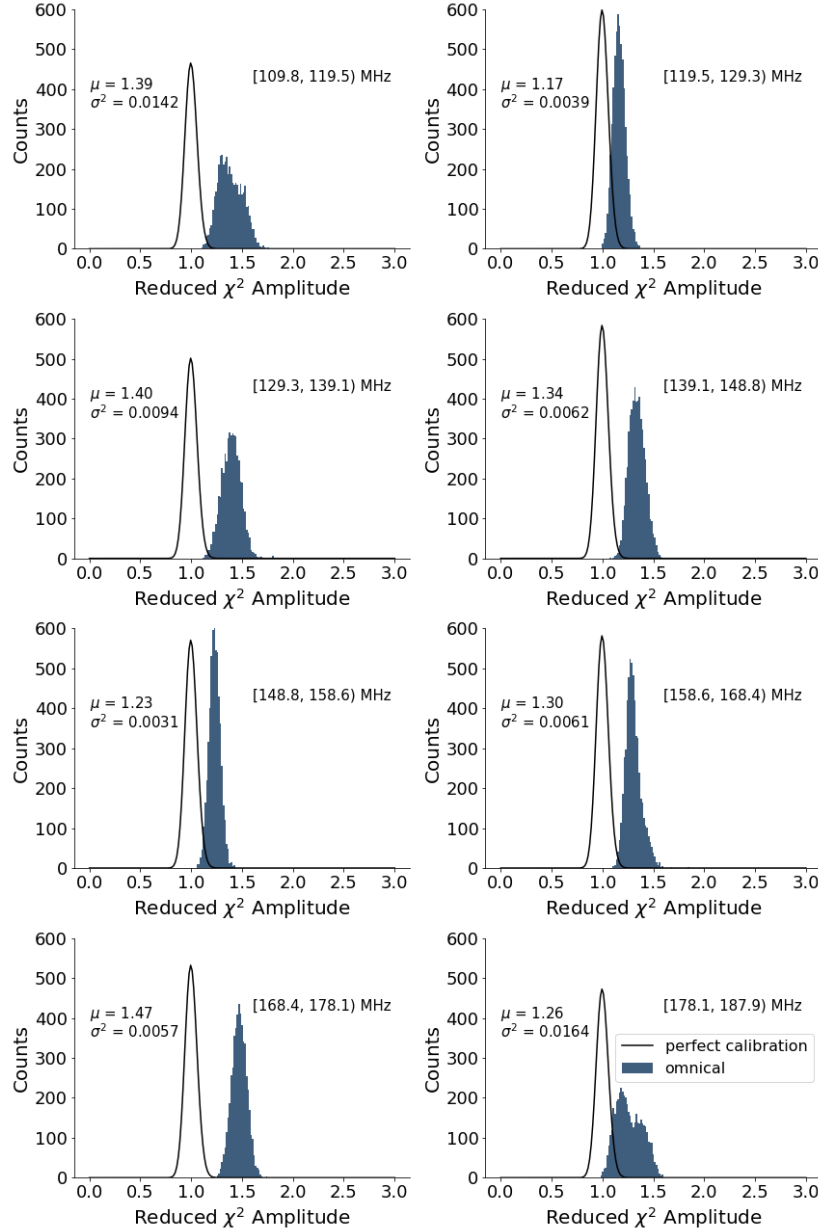


Figure 3.9: Histogram of the reduced χ^2 for East polarization from the OMNICAL run including all baseline separations except the shortest baselines (less than 15 meters) for 8 different frequency ranges (10 MHz per range) across the 100-200 MHz band HERA observes, starting from 109.8 MHz to 187.9 MHz. The means for each 10 MHz band is greater than 1, indicating that there is non-redundancy across the band. However, the variances in the middle frequency ranges tend to be closer to the expected value of $2/\text{DoF} = 0.0031$.

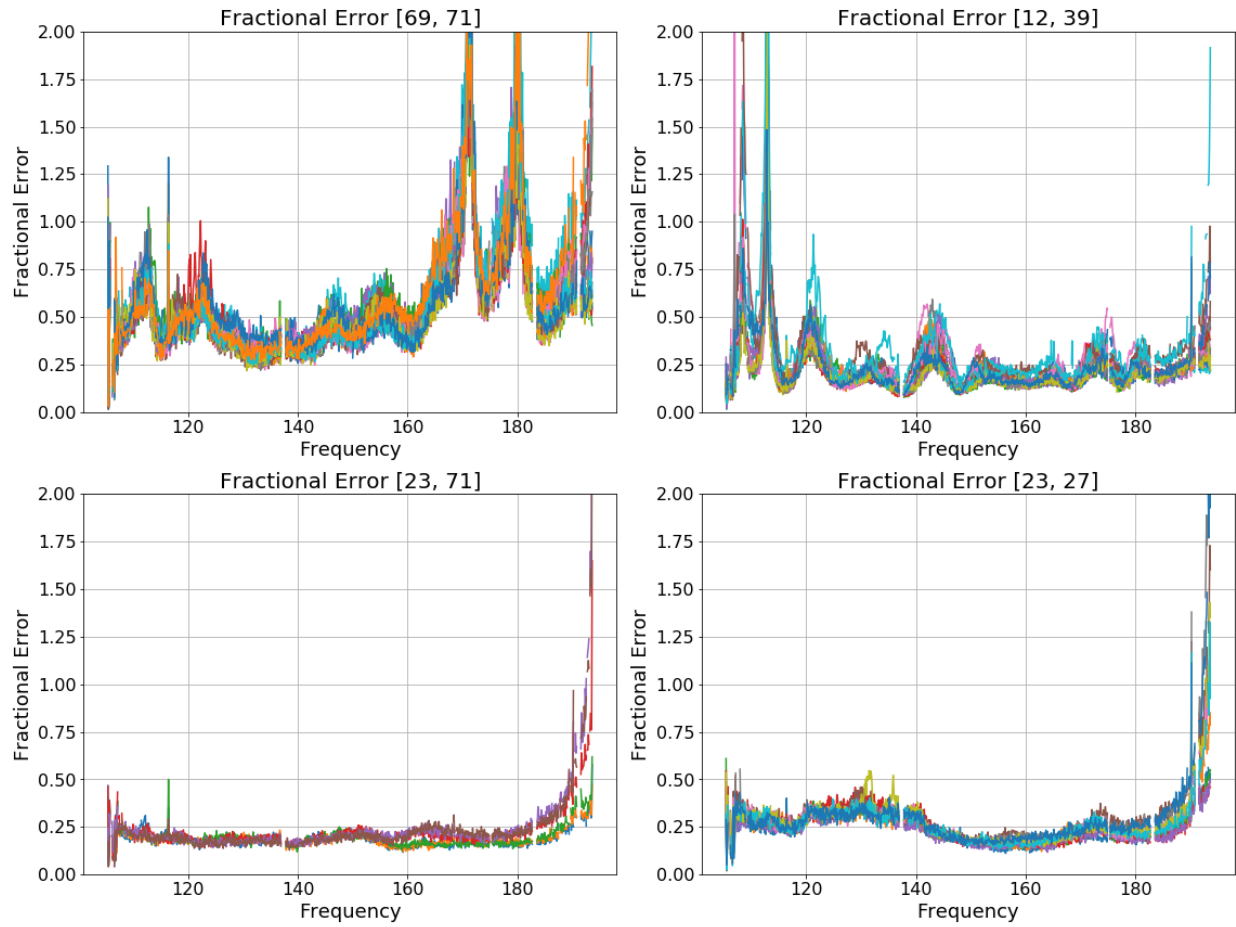


Figure 3.10: Fractional errors for four baseline lengths and orientations from the OMNICAL run without the shortest spacing baselines. Each curve in each subplot is a different instance of the baselines of length and orientation as indicated in the title (refer to Figure 3.1 for exact baselines). The shortest spacing baselines (14.7 meter baselines in the east-west and north-west directions) shown in the top two panels have the largest fractional errors, 25% to 50%, where the long baselines (bottom row) have the smallest errors hovering around 25%. Additionally, the structure in the fractional errors arises from the structure in the true visibilities solved for by OMNICAL.

In recent years with the popularization of machine learning, new and innovative ways of detecting RFI are being studied and implemented. Techniques such as K-Nearest Neighbors and Random Forest Classifiers have been used with up to 96% accuracy in detecting RFI (Mosiane et al. 2017). Using machine learning to detect RFI is an active area of research with exciting future prospects.

RFI excision on HERA

HERA’s RTS uses a simple, but effective, localized median filter technique to detect RFI on various data products to generate RFI flags. This technique is applied to raw visibilities, OMNICAL gain solutions, OMNICAL true visibilities, and the χ^2 as determined by OMNICAL (equation 3.9). Resultant flags are combined together to form a global RFI mask that is applied to each visibility. Using different data products to detect RFI is necessary to find different RFI events. Certain data products are better at identifying different types of RFI as seen in figure 3.11.

HERA’s RFI excision algorithm is applied to 2D images, or waterfalls. The algorithm begins by smoothing images using a 2D localized median filter with a 17x17 square kernel. A standard deviation matrix is constructed by finding the median absolute deviation of every pixel with respect to the smoothed image. Pixels that are 6 deviations above the median are flagged as RFI on a first pass through the image. Then a watershed algorithm starts by looking at neighbor pixels to already flagged pixels and flagging 2 sigma above the median. In this way, low level RFI from a central RFI burst are flagged.

Figure 3.11 shows the application of the watershed RFI algorithm to various data products produced throughout the RTS pipeline. As noted above, flagging on different data products is sensitive to different morphologies of RFI. The OMNICAL χ^2 ’s are sensitive to broad band RFI morphologies highlighted by the fact that the fraction of frequencies flagged in certain time bins is higher. Model visibilities and gain solutions also have sensitivity to broadband, but localized RFI. The flags from each of these data products are combined as the final RFI flag mask to be applied to every baseline.

3.3.5 Absolute Calibration

In an RTS pipeline as described in this Chapter, an absolute calibration is not included even though it is necessary for any scientific analysis of observations. In this context, an absolute calibration is one that can solve for the degenerate calibration parameters that OMNICAL can’t solve for, such as the overall flux scale of the array and the tip-tilt phase terms (Dillon et al. 2017). In this section we discuss different ways to absolutely calibrate the data as a post-processing step, as HERA does not currently have an absolute calibration step incorporated into its RTS pipeline.

The commonly adopted approach for complex gain calibration of radio interferometers involves observing one or a few bright point sources, ideally near zenith, over the course of an observation. Repeated measurements of a calibrator source can help pin down possible phase

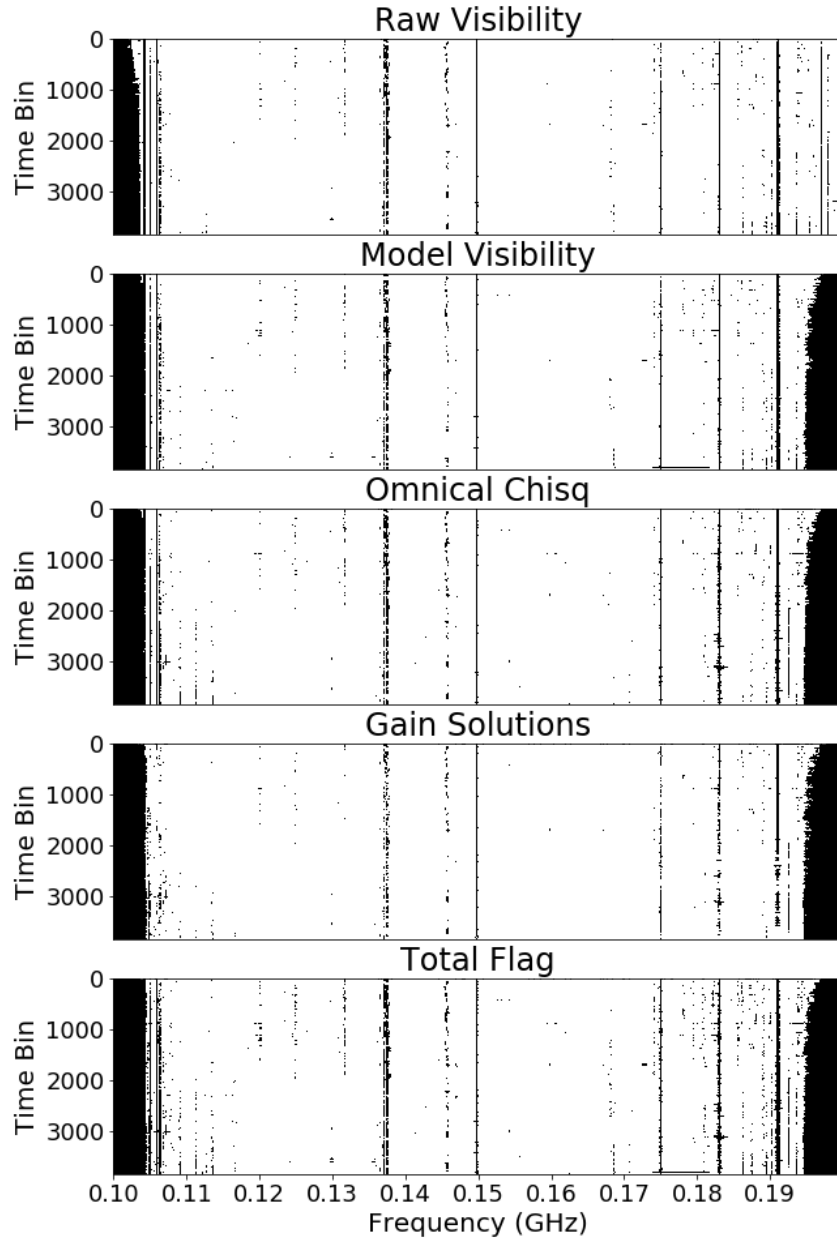


Figure 3.11: RFI flags generated from the watershed algorithm described in section 3.3.4 applied to the raw visibilities (first row), OMNICALvisibilities (second row), OMNICALchisq (third row), and OMNICALgain solutions (4th row) for an instance to the shortest east west baseline (12,13). The total flag (bottom row) shows the OR of the flags all the preceding flags. The application of the watershed algorithm on each data product picks out different characteristics of RFI.

drift of the calibration across the night. A bright source of known flux is generally observed once to set the overall flux scale of the array, which can also be used for complex bandpass calibration. This commonly adopted approach is complicated for the HERA experiment, as HERA is a drift scan array with a fairly compact primary beam ($\approx 10^\circ$ FWHM at 150 MHz). This means we are limited in the number of bright sources that transit our field of view, and cannot make repeated measurements of the same source over the course of an observation.

Depending on the stability of array, selecting one field per night for sky-based calibration may be enough to fill-in the remaining calibration parameters left over by redundant calibration. Indeed, for nominal observing conditions, we find per-antenna delay stability across the array to be on the order of ≤ 0.5 nanoseconds (Figure 3.5), which corresponds to a phase stability of ≤ 0.1 radians.

Another approach for absolute calibration would be to use a more sophisticated visibility simulation incorporating a point source sky catalogue and/or diffuse emission templates to perform absolute calibration at any and multiple times during the observation. Development of the Global Sky Model (GSM) (de Oliveira-Costa et al. 2008b; Zheng et al. 2017a) allows for quick simulations of the diffuse emission in the galaxy which dominates low frequency radio telescopes, and in particular, the shortest baselines. This was an approach used by (Zheng et al. 2014) for absolute calibration of the *Omniscope* low-frequency array. In addition, recent advances in low frequency point source cataloguing (Intema et al. 2017; Hurley-Walker et al. 2017) combined with a GSM template of diffuse emission would allow for a fairly realistic model of the short and medium length baselines in the HERA array.

An absolute calibration step in an RTS could involve either of these approaches. Automated (self) calibration and imaging scripts can be written and executed when reliable fields transit the field of view. The output images of such scripts would also serve as valuable metrics for assessing the instrumental performance. Compared to the calibration and metric steps described previously (antenna based metrics and redundant calibration) which are run on every single integration across the night, an absolute calibration step would be a considerably cheaper calculation: for a single field, one could select out a few integrations when the calibrators transit zenith and then transfer the calibration solutions from that field to nearby times. If one wanted to capture gain drift within a single observation, this could be repeated for a handful of fields, interpolating and then smoothing the solutions for the intermediary times to create a model of the gain solutions and their potential drift across the night. A crucial consideration for calibration of 21 cm arrays are the effects introduced by calibration errors. Model imperfections, whether it be from incomplete sky sources, incorrect beam models, etc., introduces calibration errors into the data that can swamp the faint cosmological signal (Barry et al. 2016; Ewall-Wice et al. 2017). This can be alleviated somewhat by using a self-calibration loop (i.e. calibrating once using an incomplete model and then using the calibrated data itself to refine the gain solutions), by smoothing the gains in frequency and time, or by enacting certain forms of data weighting in solving for the gain solutions. Work is currently underway testing the efficacy of these various absolute calibration approaches for implementation in future versions of the HERA RTS.

3.4 Conclusions

In this work we have presented a framework for building an RTS for highly redundant low frequency interferometers. RTS's are necessary for such an interferometer for a multiple reasons. First, in the era of big data, large amounts of data are beginning to be collected by next generation telescopes. Analyzing the data as fast and automatically decreases the time to science. For new telescopes in particular, fast characterization of the system is crucial to the performance issues can be addressed quickly with on-site engineers.

We have used HERA as an example of how to build an RTS for a highly redundant low frequency radio array. HERA's RTS consists of four main systems:

1. Data Ingestion: Raw data is received by the RTS and meta data derived from file contents are stored in a database.
2. Antenna Metrics: Bad antenna detection algorithms are employed detect antennas with low power inputs or are cross polarized. Antennas are flagged for the rest of the RTS.
3. Calibration: Redundant calibration is used to relatively calibrated HERA, exploiting the extensive redundancy. Redundant calibration is implemented with the OMNICAL algorithm and is run on every frequency channel and time integration independently. Once confidence in these independent solutions is established, use of outside information to smooth or compress data can be done.
4. RFI Detection and Flagging: After antennas are relatively calibrated to one another, RFI detection and flagging is implemented. The algorithm used by HERA focuses on finding 6 sigma deviations of amplitudes above the local median in time and frequency, followed by a watershed style algorithm. The algorithm is applied to raw data as well as data products from redundant calibration, including gains, the true visibilities, and the chi-squared goodness of fits.

These four systems streamline data processing for HERA. As the first iteration of HERA's RTS, this system was mainly used to learn about the instrument. A number of lessons have been learned. First, bad antenna detection is a critical component to the RTS, since without it, calibration would fail. A number of other modes in antenna failures have been seen during observations including antennas turning into transmitters. Secondly, redundant calibration is a powerful tool, but does indicate a level of non-redundancy not initially expected. Current work characterizing the non-redundancy in HERA is in progress.

Future iterations of HERA's RTS will provide more robust and efficient algorithms. Firstly, the incorporation of an absolute calibration algorithm into the pipeline will be crucial for making on demand images of the sky at different frequencies and times. Images will provide another axis for characterizing and assessing array performance. Furthermore analysis from the initial seasons of HERA observations will provide further insight into metrics for the determination of bad antennas and RFI.

Looking forward, the development of an RTS in HERA is critical to perfect so that future telescopes can build off of it. In particular, future telescopes for detection EoR may include Fourier Transform Telescopes (FFT) telescopes ([Tegmark & Zaldarriaga 2009](#)) which output images of the sky by taking spatial Fourier Transforms of the sky.

Acknowledgments

A version of this Chapter is being prepared for publication and is reproduced here with the permission of all coauthors.

Chapter 4

Conclusions

4.1 Summary of Thesis

This thesis bridges the end of a first generation EoR experiment, PAPER, and the start of a revolutionary next generation EoR experiment, HERA. The techniques and lessons learned from PAPER are being used as guiding principles to build and analyze data from HERA. I covered the latest analysis techniques from PAPER, which uses foreground avoidance methods to access the 21 cm power spectrum in the first half of this thesis. For the second half, I switched to commissioning the early stages of HERA, where I discuss the use of real time systems to analyze incoming streaming data and the first observations with HERA.

In Chapter 2, I discussed the analysis of the PAPER-64 dataset. This work covered a number of novel analysis techniques for taking raw visibility data to an estimate of a power spectrum. The techniques highlighted include precision redundant calibration, fringe-rate filtering, and the use of optimal quadratic estimators with empirical covariances. The use of redundant calibration greatly increased the clustering of measurements relative to the previous calibration analysis of PAPER data. Fringe-rate filtering proved to be a powerful technique that can alleviate systematics associated with foregrounds on the sky and is a method for removing static phase offsets. Finally, the application of optimal quadratic estimators with empirically estimated covariances has proved useful for the estimation of the 21 cm power spectrum. The power of OQE comes from the effective down weighting of systematics described by the covariance matrix. Additionally, the flexibility of the framework to try different analysis methods has also had a positive impact.

Even though the estimation of signal loss through the power spectrum pipeline was underestimated and the upper limits estimated in Chapter 2 were off by 4 orders of magnitude, I presented a methodology for the estimation of the spin temperature given an upper limit on the 21 cm power spectrum. Using this framework, if an upper-limit of $(22.4 \text{ mK})^2$ was obtained, we can show that the spin temperature must be greater than 4 K.

In Chapter 3, I switched focus to commissioning HERA using a real time system (RTS). I began by broadly discussing what a real time system is and what they're used for. In particular, an RTS is an autonomous system that algorithmically processes data to produce

regular data products. With the increasing volume of data generated by modern telescopes, I argued that it is becoming necessary to use an RTS for data processing and data reduction. I then discussed how to develop an RTS for a new low frequency, highly redundant array for which no information is known, using HERA as the practical example. I presented the first version of HERA's RTS used during the first season of observing. This RTS focussed on the detection of bad antennas, redundant calibration, and RFI removal to get a first look at the performance characteristics of HERA. I then presented commissioning results from HERA's RTS during the first season, showing that levels of non-redundancy among nominally redundant baselines depends on the length of the baselines. I finished by discussing future directions for HERA's RTS, such as the incorporation of absolute calibration.

The first observing season of HERA ended on April 1, 2018. This season started with 47 antennas in October and progressed to ≈ 60 antennas by April. Currently, this first season is undergoing a deep analysis to start extensive characterization of HERA in terms of calibration, antenna, and power spectrum performance. These lessons will be incorporated into future iterations of HERA's RTS.

With the possible detection of the global signal from EDGES and the promising initial results from HERA, the future of 21 cm cosmology is bright. The design of HERA and the analysis techniques developed from PAPER have been key to furthering the field. Once HERA is at full capacity the detection and characterization of EoR will be imminent. Following an initial detection, 3D images of the evolution of EoR and the dark ages will be possible, delivering a method for cross-correlations with other experiments. This is an exciting time for 21 cm cosmology, with many prospects on the horizon.

Bibliography

- Abramowitz, M., & Stegun, I. A. 1964, Handbook of mathematical functions with formulas, graphs, and mathematical tables (Washington, U.S. Govt. Print. Off.)
- Acernese, F., Agathos, M., Agatsuma, K., et al. 2015, [Classical and Quantum Gravity](#), **32**, 024001
- Ali, S. S., Bharadwaj, S., & Chengalur, J. N. 2008, [MNRAS](#), **385**, 2166
- Ali, Z. S., Parsons, A. R., Zheng, H., et al. 2015, [ApJ](#), **809**, 61
- Anderson, L., Aubourg, É., Bailey, S., et al. 2014, [MNRAS](#), **441**, 24
- Asad, K. M. B., Koopmans, L. V. E., Jelić, V., et al. 2015, [MNRAS](#), **451**, 3709
- Barkana, R., & Loeb, A. 2007, [Reports on Progress in Physics](#), **70**, 627
- Barkana, R., Outmezguine, N. J., Redigolo, D., & Volansky, T. 2018, ArXiv e-prints, [arXiv:1803.03091 \[hep-ph\]](#)
- Barry, N., Hazelton, B., Sullivan, I., Morales, M. F., & Pober, J. C. 2016, [MNRAS](#), **461**, 3135
- Battaglia, N., Natarajan, A., Trac, H., Cen, R., & Loeb, A. 2013, [ApJ](#), **776**, 83
- Beardsley, A. P., Hazelton, B. J., Sullivan, I. S., et al. 2016, [The Astrophysical Journal](#), **833**, 102
- Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, [AJ](#), **132**, 1729
- Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, [ApJS](#), **208**, 20
- Bernardi, G., McQuinn, M., & Greenhill, L. J. 2015, [ApJ](#), **799**, 90
- Bernardi, G., de Bruyn, A. G., Brentjens, M. A., et al. 2009, [A&A](#), **500**, 965
- Bernardi, G., de Bruyn, A. G., Harker, G., et al. 2010, [A&A](#), **522**, A67
- Bernardi, G., Greenhill, L. J., Mitchell, D. A., et al. 2013, [ApJ](#), **771**, 105
- Bolton, J. S., Haehnelt, M. G., Warren, S. J., et al. 2011, [MNRAS](#), **416**, L70
- Bosman, S. E. I., & Becker, G. D. 2015, ArXiv e-prints: 1505.06880, [arXiv:1505.06880](#)
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2010, [ApJ](#), **709**, L133
- Bouwens, R. J., Bradley, L., Zitrin, A., et al. 2014, [ApJ](#), **795**, 126
- Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2006, [ApJ](#), **638**, 20
- . 2009a, [ApJ](#), **695**, 183
- . 2009b, [ApJ](#), **695**, 183
- Bowman, J. D., & Rogers, A. E. E. 2010, [Nature](#), **468**, 796
- Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., & Mahesh, N. 2018, [Nature](#), **555**, 67

- Burns, J. O., Lazio, J., Bale, S., et al. 2012, [Advances in Space Research](#), **49**, 433
- Chapman, E., Abdalla, F. B., Harker, G., et al. 2012, [MNRAS](#), **423**, 2518
- Chapman, E., Abdalla, F. B., Bobin, J., et al. 2013, [MNRAS](#), **429**, 165
- Clark, M., Plante, P. L., & Greenhill, L. 2013, [International Journal of High Performance Computing Applications](#), **27**, 178
- Clark, S., Dutta, B., Gao, Y., Ma, Y.-Z., & Strigari, L. E. 2018, ArXiv e-prints, [arXiv:1803.09390 \[astro-ph.HE\]](#)
- Costa, A. A., Landim, R. C. G., Wang, B., & Abdalla, E. 2018, ArXiv e-prints, [arXiv:1803.06944](#)
- Datta, A., Bowman, J. D., & Carilli, C. L. 2010, [ApJ](#), **724**, 526
- Datta, A., Bradley, R., Burns, J. O., et al. 2014, ArXiv e-prints, [arXiv:1409.0513](#)
- de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., et al. 2008a, [MNRAS](#), **388**, 247
- . 2008b, [MNRAS](#), **388**, 247
- DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2016, ArXiv e-prints, [arXiv:1606.07473 \[astro-ph.IM\]](#)
- . 2017, [PASP](#), **129**, 045001
- Dillon, J. S., Liu, A., & Tegmark, M. 2013, [Phys. Rev. D](#), **87**, 043005
- Dillon, J. S., & Parsons, A. R. 2016, [ApJ](#), **826**, 181
- Dillon, J. S., Liu, A., Williams, C. L., et al. 2014, [Phys. Rev. D](#), **89**, 023002
- Dillon, J. S., Neben, A. R., Hewitt, J. N., et al. 2015, ArXiv e-prints: 1506.01026, [arXiv:1506.01026](#)
- Dillon, J. S., Kohn, S. A., Parsons, A. R., et al. 2017, ArXiv e-prints, [arXiv:1712.07212 \[astro-ph.IM\]](#)
- Dodelson, S. 2003, *Modern cosmology* (San Diego, CA: Academic Press)
- Eastwood, M. W., Anderson, M. M., Monroe, R. M., et al. 2017, ArXiv e-prints, [arXiv:1711.00466 \[astro-ph.IM\]](#)
- Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, [ApJ](#), **633**, 560
- Ellingson, S. W., Taylor, G. B., Craig, J., et al. 2013, [IEEE Transactions on Antennas and Propagation](#), **61**, 2540
- Ellis, R. S., McLure, R. J., Dunlop, J. S., et al. 2013, [ApJ](#), **763**, L7
- Ewall-Wice, A., Chang, T.-C., Lazio, J., et al. 2018, ArXiv e-prints, [arXiv:1803.01815](#)
- Ewall-Wice, A., Dillon, J. S., Liu, A., & Hewitt, J. 2017, [MNRAS](#), **470**, 1849
- Ewall-Wice, A., Bradley, R., Deboer, D., et al. 2016, [ApJ](#), **831**, 196
- 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. 2014, [MNRAS](#), **445**, 213
- Fialkov, A., Barkana, R., & Visbal, E. 2014a, [Nature](#), **506**, 197
- . 2014b, [Nature](#), **506**, 197
- Fialkov, A., Cohen, A., Barkana, R., & Silk, J. 2017, [MNRAS](#), **464**, 3498
- Fialkov, A., & Loeb, A. 2016, [ApJ](#), **821**, 59
- Field, G. B. 1958, *Proc. IRE*, **46**, 240
- Furlanetto, S. R., McQuinn, M., & Hernquist, L. 2006a, [MNRAS](#), **365**, 115

- Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006b, [Phys. Rep.](#), **433**, 181
- Furlanetto, S. R., Oh, S. P., & Pierpaoli, E. 2006c, [Phys. Rev. D](#), **74**, 103502
- Furlanetto, S. R., Zaldarriaga, M., & Hernquist, L. 2006d, [MNRAS](#), **365**, 1012
- Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, [Space Sci. Rev.](#), **123**, 485
- George, E. M., Reichardt, C. L., Aird, K. A., et al. 2014, ArXiv e-prints: 1408.3161, [arXiv:1408.3161](#)
- Ghosh, A., Bharadwaj, S., Ali, S. S., & Chengalur, J. N. 2011, [MNRAS](#), **418**, 2584
- Gleser, L., Nusser, A., & Benson, A. J. 2008, [MNRAS](#), **391**, 383
- Greenhill, L. J., & Bernardi, G. 2012, ArXiv e-prints: 1201.1700, [arXiv:1201.1700 \[astro-ph.CO\]](#)
- Greig, B., & Mesinger, A. 2015, [MNRAS](#), **449**, 4246
- Grobler, T. L., Bernardi, G., Kenyon, J. S., Parsons, A. R., & Smirnov, O. M. 2018, [MNRAS](#), [arXiv:1801.03249 \[astro-ph.IM\]](#)
- Hall, G., ed. 1988, The ARRL Antenna Book, 15th Edition (American Radio Relay League, Newington, Connecticut, 1988.), 16
- 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
- Hickish, J., Abdurashidova, Z., Ali, Z., et al. 2016, [Journal of Astronomical Instrumentation](#), **5**, 1641001
- Hirata, C. M. 2006, [MNRAS](#), **367**, 259
- Högbom, J. A. 1974, [A&AS](#), **15**, 417
- Hu, W., & White, M. 1997, [ApJ](#), **479**, 568
- Hubble, E. 1929, [Proceedings of the National Academy of Science](#), **15**, 168
- Hurley-Walker, N., Callingham, J. R., Hancock, P. J., et al. 2017, [MNRAS](#), **464**, 1146
- Iglewicz, B., & Hoaglin, David C. (David Caster), . 1993, How to detect and handle outliers (Milwaukee, Wis. : ASQC Quality Press), includes bibliographical references (p. 73-78) and index
- Intema, H. T., Jagannathan, P., Mooley, K. P., & Frail, D. A. 2017, [A&A](#), **598**, A78
- Jacobs, D. C., Aguirre, J. E., Parsons, A. R., et al. 2011, [ApJ](#), **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
- Jarosik, N., Bennett, C. L., Dunkley, J., et al. 2011, [ApJS](#), **192**, 14
- Jelić, V., Zaroubi, S., Labropoulos, P., et al. 2010, [MNRAS](#), **409**, 1647
- . 2008, [MNRAS](#), **389**, 1319
- Jelić, V., de Bruyn, A. G., Mevius, M., et al. 2014, [A&A](#), **568**, A101
- Jensen, H., Datta, K. K., Mellema, G., et al. 2013, [MNRAS](#), **435**, 460
- Kern, N. S., Liu, A., Parsons, A. R., Mesinger, A., & Greig, B. 2017, [ApJ](#), **848**, 23
- Kohn, S., Chichura, P. M., Igarashi, A. S., et al. 2018, ArXiv e-prints, [arXiv:1802.04151 \[astro-ph.IM\]](#)
- Lenc, E., Gaensler, B. M., Sun, X. H., et al. 2016, [ApJ](#), **830**, 38
- Li, W., Pober, J. C., Hazelton, B. J., et al. 2018, ArXiv e-prints, [arXiv:1807.05312 \[astro-ph.IM\]](#)

- Lidz, A., Zahn, O., Furlanetto, S. R., et al. 2009, [ApJ](#), 690, 252
- Lidz, A., Zahn, O., McQuinn, M., Zaldarriaga, M., & Hernquist, L. 2008, [ApJ](#), 680, 962
- LIGO Scientific Collaboration, Aasi, J., Abbott, B. P., et al. 2015, [Classical and Quantum Gravity](#), 32, 074001
- Liu, A., & Parsons, A. R. 2016, [MNRAS](#), 457, 1864
- Liu, A., Parsons, A. R., & Trott, C. M. 2014a, [Phys. Rev. D](#), 90, 023018
- . 2014b, [Phys. Rev. D](#), 90, 023019
- Liu, A., Pritchard, J. R., Allison, R., et al. 2016, [Phys. Rev. D](#), 93, 043013
- Liu, A., & Tegmark, M. 2011, [Phys. Rev. D](#), 83, 103006
- . 2012, [MNRAS](#), 419, 3491
- Liu, A., Tegmark, M., Bowman, J., Hewitt, J., & Zaldarriaga, M. 2009a, [MNRAS](#), 398, 401
- Liu, A., Tegmark, M., Morrison, S., Lutomirski, A., & Zaldarriaga, M. 2010, [MNRAS](#), 408, 1029
- Liu, A., Tegmark, M., & Zaldarriaga, M. 2009b, [MNRAS](#), 394, 1575
- Liu, H., & Slatyer, T. R. 2018, ArXiv e-prints, [arXiv:1803.09739](#)
- Loeb, A., & Furlanetto, S. 2013, *The First Galaxies in the Universe*, Princeton Series in Astrophysics (Princeton University Press)
- Madau, P., Meiksin, A., & Rees, M. J. 1997, [ApJ](#), 475, 429
- Mao, Y., Tegmark, M., McQuinn, M., Zaldarriaga, M., & Zahn, O. 2008, [Phys. Rev. D](#), 78, 023529
- Marthi, V. R., & Chengalur, J. 2014, [MNRAS](#), 437, 524
- Masui, K. W., Switzer, E. R., Banavar, N., et al. 2013, [ApJ](#), 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
- Mirocha, J., & Furlanetto, S. R. 2018, ArXiv e-prints, [arXiv:1803.03272](#)
- Mirocha, J., Furlanetto, S. R., & Sun, G. 2017, [MNRAS](#), 464, 1365
- Mirocha, J., Mebane, R. H., Furlanetto, S. R., Singal, K., & Trinh, D. 2018, [MNRAS](#), [arXiv:1710.02530](#)
- Moore, D. C., Aguirre, J., Parsons, A., et al. 2015, ArXiv e-prints: 1502.05072, [arXiv:1502.05072](#)
- Moore, D. F., Aguirre, J. E., Parsons, A. R., Jacobs, D. C., & Pober, J. C. 2013, [ApJ](#), 769, 154
- Moore, D. F., Aguirre, J. E., Kohn, S. A., et al. 2017, [ApJ](#), 836, 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
- Mosiane, O., Oozeer, N., Aniyani, A., & Bassett, B. A. 2017, [IOP Conference Series: Materials Science and Engineering](#), 198, 012012
- Neben, A. R., Bradley, R. F., Hewitt, J. N., et al. 2016, [ApJ](#), 826, 199
- Newburgh, L. B., Bandura, K., Bucher, M. A., et al. 2016, in *Proc. SPIE*, Vol. 9906, Ground-

- based and Airborne Telescopes VI, 99065X
- Noorishad, P., Wijnholds, S. J., van Ardenne, A., & van der Hulst, J. M. 2012, *A&A*, **545**, A108
- Nunhokee, C. D., Bernardi, G., Kohn, S. A., et al. 2017, *ApJ*, **848**, 47
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2013, *ApJ*, **773**, 75
- Offringa, A. R. 2012, PhD thesis, University of Groningen
- Offringa, A. R., McKinley, B., Hurley-Walker, N., et al. 2014, *MNRAS*, **444**, 606
- Paciga, G., Chang, T.-C., Gupta, Y., et al. 2011, *MNRAS*, **413**, 1174
- Paciga, G., Albert, J. G., Bandura, K., et al. 2013, *MNRAS*, **433**, 639
- 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., Pober, J., McQuinn, M., Jacobs, D., & Aguirre, J. 2012a, *ApJ*, **753**, 81
- Parsons, A., Backer, D., Chang, C., et al. 2006, in *Signals, Systems and Computers*, 2006. ACSSC '06. Fortieth Asilomar Conference on, 2031
- Parsons, A., Backer, D., Siemion, A., et al. 2008, *PASP*, **120**, 1207
- Parsons, A. R., & Backer, D. C. 2009, *AJ*, **138**, 219
- Parsons, A. R., Liu, A., Ali, Z. S., & Cheng, C. 2015, ArXiv e-prints: 1503.05564, [arXiv:1503.05564 \[astro-ph.IM\]](#)
- Parsons, A. R., Pober, J. C., Aguirre, J. E., et al. 2012b, *ApJ*, **756**, 165
- Parsons, A. R., Backer, D. C., Foster, G. S., et al. 2010, *AJ*, **139**, 1468
- Parsons, A. R., Liu, A., Aguirre, J. E., et al. 2014, *ApJ*, **788**, 106
- Patil, A. H., Yatawatta, S., Zaroubi, S., et al. 2016, *MNRAS*, **463**, 4317
- Patil, A. H., Yatawatta, S., Koopmans, L. V. E., et al. 2017, *ApJ*, **838**, 65
- Patra, N., Subrahmanyam, R., Sethi, S., Udaya Shankar, N., & Raghunathan, A. 2015, *ApJ*, **801**, 138
- Patra, N., Parsons, A. R., DeBoer, D. R., et al. 2018, *Experimental Astronomy*, **45**, 177
- Penzias, A. A., & Wilson, R. W. 1965, *ApJ*, **142**, 419
- Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, *ApJ*, **517**, 565
- Peterson, J. B., Pen, U.-L., & Wu, X.-P. 2004, *Modern Physics Letters A*, **19**, 1001
- Petrovic, N., & Oh, S. P. 2011, *MNRAS*, **413**, 2103
- Philip, L., Abdurashidova, Z., Chiang, H. C., et al. 2018, ArXiv e-prints, [arXiv:1806.09531 \[astro-ph.IM\]](#)
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, *A&A*, **571**, A16
- . 2015, ArXiv e-prints: 1502.01589, [arXiv:1502.01589](#)
- Pober, J. C., Parsons, A. R., Jacobs, D. C., et al. 2012, *AJ*, **143**, 53
- Pober, J. C., Parsons, A. R., Aguirre, J. E., et al. 2013, *ApJ*, **768**, L36
- Pober, J. C., Liu, A., Dillon, J. S., et al. 2014, *ApJ*, **782**, 66
- Pober, J. C., Ali, Z. S., Parsons, A. R., et al. 2015, ArXiv e-prints: 1503.00045, [arXiv:1503.00045](#)
- Presley, M. E., Liu, A., & Parsons, A. R. 2015, *ApJ*, **809**, 18
- Pritchard, J. R., & Loeb, A. 2008, *Phys. Rev. D*, **78**, 103511
- . 2012, *Reports on Progress in Physics*, **75**, 086901

- Rhoads, J. E., Malhotra, S., Stern, D., et al. 2013, [ApJ](#), 773, 32
- Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, [AJ](#), 116, 1009
- Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, [ApJ](#), 802, L19
- Robertson, B. E., Furlanetto, S. R., Schneider, E., et al. 2013, [ApJ](#), 768, 71
- Rybicki, G. B., & Lightman, A. P. 1979, *Radiative processes in astrophysics* (New York, Wiley-Interscience, 1979. 393 p.)
- Santos, M. G., Cooray, A., & Knox, L. 2005, [ApJ](#), 625, 575
- Schenker, M. A., Robertson, B. E., Ellis, R. S., et al. 2013, [ApJ](#), 768, 196
- Shaw, J. R., Sigurdson, K., Pen, U.-L., Stebbins, A., & Sitwell, M. 2013, ArXiv e-prints, [arXiv:1302.0327 \[astro-ph.CO\]](#)
- Sievers, J. L. 2017, ArXiv e-prints, [arXiv:1701.01860 \[astro-ph.IM\]](#)
- Silva, M., Santos, M. G., Cooray, A., & Gong, Y. 2015, [ApJ](#), 806, 209
- Singh, S., Subrahmanyan, R., Shankar, N. U., et al. 2018a, [Experimental Astronomy](#), 45, 269
- Singh, S., Subrahmanyan, R., Udaya Shankar, N., et al. 2018b, [ApJ](#), 858, 54
- Skidmore, W., TMT International Science Development Teams, & Science Advisory Committee, T. 2015, [Research in Astronomy and Astrophysics](#), 15, 1945
- Smoot, G. F., Bennett, C. L., Kogut, A., et al. 1992, [ApJ](#), 396, L1
- 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., Chang, T.-C., Masui, K. W., Pen, U.-L., & Voytek, T. C. 2015, [ApJ](#), 815, 51
- Switzer, E. R., & Liu, A. 2014, [The Astrophysical Journal](#), 793, 102
- Taylor, G. B., Ellingson, S. W., Kassim, N. E., et al. 2012, [Journal of Astronomical Instrumentation](#), 1, 1250004
- Tegmark, M. 1997, [Phys. Rev. D](#), 55, 5895
- Tegmark, M., & Zaldarriaga, M. 2009, [Phys. Rev. D](#), 79, 083530
- The CHIME/FRB Collaboration, :, Amiri, M., et al. 2018, ArXiv e-prints, [arXiv:1803.11235 \[astro-ph.IM\]](#)
- Thompson, A. R., Moran, J. M., & Swenson, Jr., G. W. 2001, *Interferometry and Synthesis in Radio Astronomy*, 2nd Edition (New York, Wiley-Interscience, 2001. 692 p.)
- Thyagarajan, N., Parsons, A. R., DeBoer, D. R., et al. 2016, [ApJ](#), 825, 9
- Thyagarajan, N., Udaya Shankar, N., Subrahmanyan, R., et al. 2013, [ApJ](#), 776, 6
- Thyagarajan, N., Jacobs, D. C., Bowman, J. D., et al. 2015, [ApJ](#), 804, 14
- 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, [ApJ](#), 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. 2013a, [A&A](#), 556, A2
- . 2013b, [A&A](#), 556, A2
- Vedantham, H., Udaya Shankar, N., & Subrahmanyan, R. 2012, [ApJ](#), 745, 176
- Vedantham, H. K., Koopmans, L. V. E., de Bruyn, A. G., et al. 2014, [Monthly Notices of the Royal Astronomical Society](#), 437, 1056
- Vedantham, H. K., Koopmans, L. V. E., de Bruyn, A. G., et al. 2015, [MNRAS](#), 450, 2291

- Visbal, E., Barkana, R., Fialkov, A., Tselikhovich, D., & Hirata, C. M. 2012, [Nature](#), **487**, 70
- Voytek, T. C., Natarajan, A., Jáuregui García, J. M., Peterson, J. B., & López-Cruz, O. 2014, [ApJ](#), **782**, L9
- Vrbanec, D., Ciardi, B., Jelić, V., et al. 2016, [MNRAS](#), **457**, 666
- Wang, X., Tegmark, M., Santos, M. G., & Knox, L. 2006, [ApJ](#), **650**, 529
- Wieringa, M. H. 1992, [Experimental Astronomy](#), **2**, 203
- Wouthuysen, S. A. 1952, [AJ](#), **57**, 31
- Wu, X. 2009, in Bulletin of the American Astronomical Society, Vol. 41, American Astronomical Society Meeting Abstracts 213, 226.05
- Yatawatta, S., Zaroubi, S., de Bruyn, G., Koopmans, L., & Noordam, J. 2008, ArXiv:0810.5751, [arXiv:0810.5751](#)
- 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
- Zheng, H., Tegmark, M., Dillon, J. S., et al. 2017a, [MNRAS](#), **464**, 3486
- Zheng, Z.-Y., Wang, J., Rhoads, J., et al. 2017b, [ApJ](#), **842**, L22