A First Imaging Power Spectrum with the Hydrogen Epoch of Reionization Array Phase II

Dara Storer

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Miguel Morales, Chair

Marilena Loverde

Jason Detwiler

Program Authorized to Offer Degree:
Physics

University of Washington

Abstract

A First Imaging Power Spectrum with the Hydrogen Epoch of Reionization Array Phase II

Dara Storer

Chair of the Supervisory Committee:
Miguel Morales
Physics

The 21 cm signal from neutral hydrogen is a powerful probe for observing the Epoch of Reionization, a period in cosmological history when the first stars and galaxies were forming. Observing this epoch would give astronomers tremendous insight into the underlying physics driving structure formation in the early universe. However, the 21 cm signal is incredibly faint, and therefore measurements will require high precision instruments and analysis techniques.

This thesis works with data from the Hydrogen Epoch of Reionization Array, a radio interferometer designed to observe the 21 cm signal. We present efforts to identify and mitigate instrumental systematics, and explore their direct influence on our ability to achieve successful measurements. Through this, we present new and informative tools for data quality assessment in a many-element interferometer. This includes both statistical and visual data inspection techniques that allow observers to succinctly inspect extremely large volumes of data. Additionally, we make use of interferometric imaging techniques to pursue a power spectrum measurement. This work represents the first imaging-based power spectrum results with HERA phase II. We find a feature that inhibits our power spectrum measurement at some modes, but are noise limited at others. Our results indicate that with a deeper integration HERA has the potential to place scientifically interesting upper limits on the 21 cm signal at some power spectrum modes.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Number Page

GLOSSARY

21 CM SIGNAL: Emission from the hyperfine spin-flip transition of neutral hydrogen.

ANTENNA: A receiving element of the interferometer. In HERA's case this is used interchangeably with 'dish'.

AUTO-CORRELATION: The correlation between an antenna and itself.

BASELINE: A pair of antennas, specifically referring to the length and angle of separation between the antennas on the ground. Can be measured in meters or units of wavelengths.

BEAM: The instrumental response to an incoming electric field. See Section 6.3.2.

CALIBRATION: The process of correcting for instrumental contributions

CROSS-CORRELATION: The correlation between two different antennas.

DIGITAL NOISE: An observing mode in HERA where a digital noise source is used rather than observing the sky. See Figure 1.6 for details.

DISH: A receiving element of the interferometer. In HERA's case this is used interchangeably with 'antenna'.

EOR: Epoch of Reionization.

ϵPPSILON: Error Propagated Power Spectrum with Interleaved Observed Noise. Software package used for power spectrum calculation throughout this work.

FEM: Front End Module.

F-ENGINE: A module within the SNAP that performs the Fourier transform that takes the data from a time-series to a frequency spectrum.

FHD: Fast Holographic Deconvolution. Software package used for calibration and imaging throughout this work.

FLAG: Any process through which we remove data from all or part of our analysis. Can happen at the time, antenna, baseline, or frequency level.

GLEAM: Galactic and Extragalactic All-sky MWA Survey. A source catalog used for sky modeling throughout this work.

HEALPIX: Hierarchical Equal Area Iso Latitude pixelation. Software package for creating equal volume pixels on a curved sky.

HERA: Hydrogen Epoch of Reionization Array. Radio interferometer used throughout this work.

JD: Julian Date. A unit of time that increases by one for every calendar day. Frequently used to refer to nights of observation, as it does not wrap at midnight.

LOAD: FEM load data. An observing mode in HERA where a load source in the FEM is used instead of observing the sky.

NODE: A physical container within the HERA array where up to 24 antenna-polarization signal chains are processed.

PAM: Post-Amplification Module.

RFI: Radio Frequency Interference. Any non-astrophysical radio emission, typically human-generated.

SNAP: Smart Network ADC Processors

SSINS: Sky-Subtracted Incoherent Noise Spectra. Software package used for RFI identification and flagging throughout this work.

UV PLANE: The instrument measurement plane, defined in Equation 1.4.

VISIBILITY: The fundamental measurement of a radio interferometer. Samples the coherence function between the electric field received by two antennas. One complex number per baseline, time, frequency, and polarization.

VIVALDI FEED: HERA's antenna feed. See Fagnoni et al. (2021) for details.

ACKNOWLEDGMENTS

This work would have been impossible without the invaluable help from people in all corners of my world. Every ounce of support has been essential to this work, and I am beyond grateful. While I sadly cannot name each person (and furry friend) who has contributed along the way, I will name those who have been most essential to this work.

First and foremost, I thank my advisors Miguel Morales and Bryna Hazelton. To Miguel, thank you for being such a reliable support in both personal and scientific matters. Your unending enthusiasm, even in the face of perpetually problematic data, taught me to find roadblocks more exciting than frustrating. To Bryna, thank you for always being available to patiently explain things, even when it was far from the first time I'd asked. To both of you, thank you for leading your group with empathy and seeing the person before the scientist - it is a trait not often found in our field, and one whose value cannot be overstated.

Next, I would like to acknowledge several of the wonderful undergraduate and post-baccalaureate students I have had the pleasure of working with and mentoring. To Ellie McLaurin, for years of hard work understanding HERA's temporal structure. To Catalina Zamora, for detailed analysis of a HERA systematic earning the name 'Cat's Combs'. To Kiana Henny, for understanding the GLEAM catalog at low frequencies and informing some of my future analysis choices. To Nyla Budlong for painstakingly inspecting immense quantities of HERA data and doing it with a smile. Finally to Adele Myers, Imani Ware, Michael Reese, and Cyrus Goodman for embodying the spirit of inquisitive learning. To all those mentioned here, thank you for giving me the opportunity to grow in my role as a mentor, and for always being delightful to work with.

A special thank you to Brockton Stover, for tremendous work in the lab investigating the excess correlation feature, finding a solution, and working tirelessly at the HERA site to implement the repairs. Also to Josh Dillon, for acquainting me with HERA's analysis tools and mentoring me through the process of writing my first paper. Your patient and thoughtful support has been

invaluable.

Many thanks to the fellow graduate students in my lab: Pyxie Star, Nicel Mohamed-Hinds, and Eli Lilleskov. You are all wonderful, and have made my years at UW rich with both meaningful support and utterly meaningless chit-chat, both of which are vital to me. I also thank my thesis committee for the valuable feedback and support: Miguel Morales, Marilena Loverde, Jason Detwiler, Matt McQuinn, and Sarah Tuttle.

I would also like to thank all of the teachers and mentors who have guided me on the way to this point. Thank you to Rachel Webster, Ian Dell'Antonio, Wolfgang Losert, and Derek Richardson for advising me in research as an undergraduate and helping me discover my strengths and interests. A big thank you to Jonathan Pober, who advised my undergraduate thesis and was a valuable mentor both scientifically and personally. Finally, a most special thank you to William Pegram who taught my high school Earth sciences course so exceptionally that one simple lesson inspired the next 14 years of study in astronomy. Your dedication to your students is unmatched, and your passion for curiosity is a lesson I will always carry with me.

To my friends, family, and amazing partner - you mean the world to me, and your unending support is beyond words.

The large majority of this work took place in Seattle, Washington, which is on stolen Coast Salish land, specifically the ancestral land of the Duwamish, Suquamish, Stillaguamish, and Muckleshoot People.

DEDICATION

To my wonderful parents,

who have given me the world and shown me how to move through it.

Chapter 1

INTRODUCTION

This thesis focuses on the instrumental aspect of efforts to observe a period of the universe known as the Epoch of Reionization (EoR). Observing this epoch has the potential to inform our understanding of the early universe and how it evolved into what it is today. We will begin by providing some scientific background and motivating our eventual goal of observing the EoR. The majority of this thesis will explore instrumental considerations and systematics, and how they affect our ability to make successful measurements.

1.1 A Brief History of Cosmology

We will begin with a brief introduction to the history of the universe, also shown pictorially in Figure 1.1. We begin 13.8 billion years ago with the Big Bang, followed by a very rapid period of expansion known as inflation. Inflation lasted less than a nanosecond, after which the universe had already expanded by a factor of 10^{30}. At this point the universe is a hot ionized plasma of the most fundamental particles, and the minuscule quantum fluctuations that existed at the time of the Big Bang have been imprinted on the universe on cosmological scales. Within a few minutes after the Big Bang the universe had cooled enough for nucleosynthesis to occur, creating the earliest atomic nuclei. The universe continued slowly expanding and cooling for 380,000 years until it was sufficiently cool for nuclei and electrons to form the first atoms, an event known as recombination. Before recombination, photons were scattered by the dense plasma of charged particles and were unable to propagate. After recombination, photons were able to travel largely unimpeded through the cosmos for the first time. This sudden release of photons, known as the Cosmic Microwave Background (CMB), is still observable today (Kragh, 2023).

After recombination, the universe entered a period known as the Dark Ages, where the vast majority of luminous matter took the form of neutral hydrogen. During this time, the density fluctuations imprinted by inflation began to intensify, wherein the most dense regions grew denser

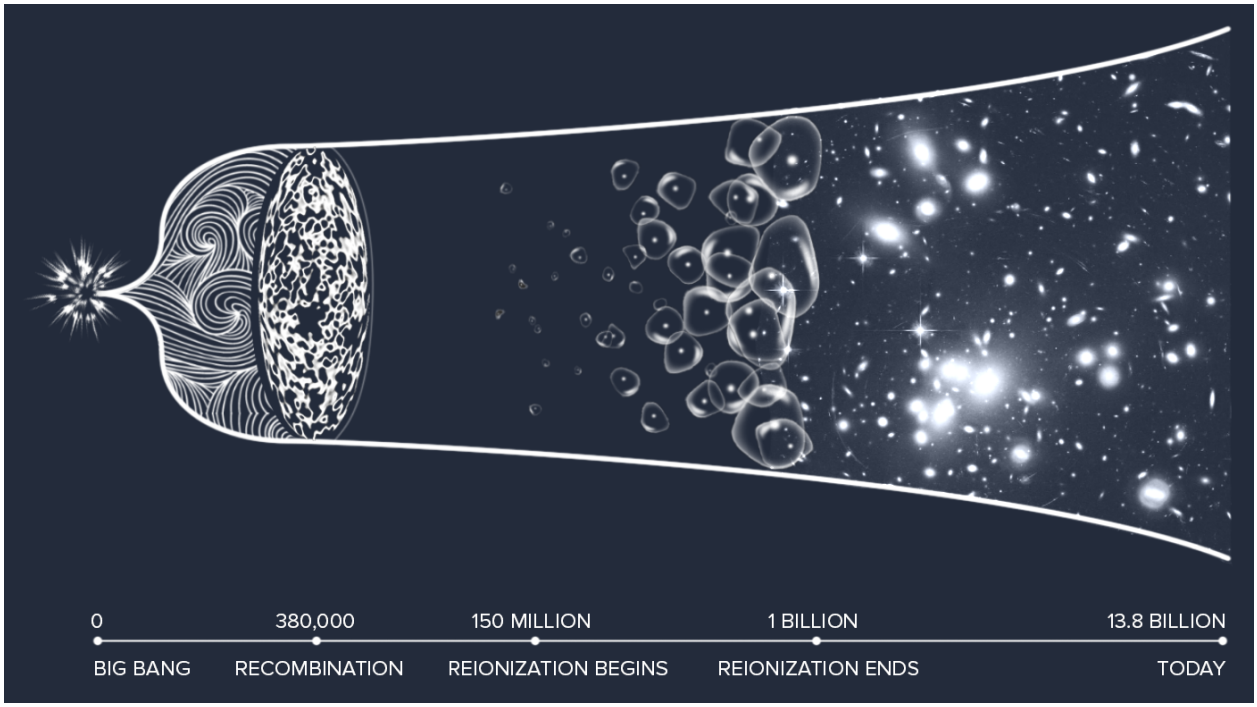


Figure 1.1: Graphic showing a simplified history of the universe showing several important epochs, including the Epoch of Reionization (EoR). The exact timing of the beginning and end of the EoR remains an active area of research. The timeline shown is not on a linear scale. Image credit: Aman Chokshi.

and the least dense regions further rarefied. Eventually, some regions became sufficiently dense for the first stars and galaxies to begin forming. As these luminous sources turn on they began emitting radiation that was sufficiently energetic to ionize the neutral hydrogen comprising the intergalactic medium (IGM). This period is known as the Epoch of Reionization (EoR).

Broadly speaking, the EoR refers to the time from when the first ionizing sources turned on in the early universe to when the universe became fully ionized as it is today. As depicted in Figure 1.1, reionization happened in 'bubbles' where the regions immediately surrounding luminous sources were ionized first and these bubbles continued growing and merging until the entire IGM was ionized. Today, there is very little neutral hydrogen in the IGM. The exact timing of the beginning and end of reionization remains an active area of research, but current estimates place the EoR between

150 million and 1.1 billion years after the Big Bang (Liu and Shaw, 2020; Zaroubi, 2012; Morales and Wyithe, 2010; Shimabukuro et al., 2021). Astronomers more commonly describe the timeline of the universe using the term redshift, which quantifies how much light has stretched since it was emitted, and directly corresponds to how long ago the light was emitted. In this framework, the redshift z of observed emission can be written as

$$z = \frac{\lambda_o}{\lambda_e} - 1 \tag{1.1}$$

where λ_o is the wavelength of the observed emission and λ_e is the wavelength at the time of emission. Light emitted today would therefore have redshift $z = 0$, and we can now describe the EoR as taking place sometime between redshifts $5.3 < z < 20$ (Bosman et al., 2022).

1.2 21 cm Cosmology: Why and How

There are many questions that remain unanswered about how exactly reionization happened. When did it end, and how long was it ongoing? What was the morphology of the ionizing bubbles? Which regions ionized first and why? While these questions directly pertain to the EoR itself, their answers have the potential to provide meaningful insight into the most fundamental questions of our universe. How did baryons and dark matter interact to form the structure of our modern universe? How did the first stars and galaxies form, and were they similar to the ones we still see today? Over the past decades theorists have formulated many different models that describe the timeline of reionization, the scale and morphology of the reionized bubbles, the dynamics of merging ionization fronts, and the uniformity of the inter-galactic medium (IGM) (Gnedin and Madau, 2022). Experimentalists have been able to constrain some of these models by confirming what we *don't* see. However, we still seek direct observation and mapping of the EoR.

One highly promising probe of the EoR is the 21 cm emission line coming from the hyperfine transition of neutral hydrogen. This transition arises due to the hyperfine interactions between a neutral hydrogen atom's proton and electron which splits their spin-alignment states into a singlet antisymmetric state and a triplet symmetric state. The energy difference between these two states corresponds to 1420 MHz, or a wavelength of 21 cm. This is a forbidden transition with a mean lifetime of 10^7 years. While the signal from this transition is incredibly rare, the early universe was

almost entirely comprised of neutral hydrogen, meaning that the total emission rate of the 21 cm signal was high enough for it to be detectable today. The signal is also emitted with a very precise energy (or wavelength), meaning that any modern observation of the signal can be very directly mapped to a particular redshift. To conceptualize this, we can consider any range of frequencies we observe as corresponding to a line-of-sight observation of the 21 cm signal, where lower frequencies correspond to a 21 cm signal that has been more redshifted and therefore was emitted both farther away and longer ago. Taking the EoR to happen sometime in redshift range $5.3 < z < 20$, we would expect to see 21 cm emission from the EoR arrive on Earth at wavelengths $1.3m < \lambda < 4.4m$, or a frequency range between 68MHz and 226MHz.

The important principle here is that by observing the 21 cm signal coming from neutral hydrogen, we can indirectly observe the evolution of ionized structures. As ionized hydrogen does not produce a 21 cm signal, we can consider a lack of signal to be an indication that the source was an ionized region. In this sense, we can consider the strength of the 21 cm signal at a particular redshift to be a measurement of how ionized a particular patch of the universe was at the time the signal was emitted. Again considering our observed frequency band to be equivalent to our line of sight, we therefore expect the strength of the 21 cm signal to fluctuate in frequency as our line of sight passes in and out of ionized bubbles. Attempts to map the strength of the 21 cm signal across the sky is a type of 3D imaging, otherwise known as tomography, where the redshifting of the 21 cm signal causes the line-of-sight spatial axis to map to a frequency axis. Despite the enormous volume of neutral hydrogen in the early universe, the 21 cm transition is sufficiently rare that the cosmological 21 cm signal remains very faint and extremely difficult to observe. While producing 3D tomographic maps of the signal is an eventual goal of the field, current experiments do not have the capability. Instead, we aim to perform a statistical measurement of the 21 cm signal's strength at different angular scales, called a power spectrum. We will describe the power spectrum measurement at length in Chapter 8.

In recent years, results from many different instruments have placed upper limits on the 21 cm power spectrum, including the Murchison Widefield Array (MWA) (Beardsley et al., 2016b; Barry et al., 2019b; Li et al., 2019; Kolopanis et al., 2019; Trott et al., 2020), the Low-Frequency Array (LOFAR) (van Haarlem et al., 2013a; Patil et al., 2017; Mertens et al., 2020), the Donald C. Backer Precision Array for Probing the EoR (PAPER) (Parsons et al., 2010a; Kolopanis et al., 2019),

the Giant Metrewave Radio Telescope (GMRT) (Pen et al., 2009; Paciga et al., 2013), the Long Wavelength Array (LWA) (Eastwood et al., 2019), and the Hydrogen Epoch of Reionization Array (HERA) (DeBoer et al., 2017; Berkhout et al., 2024; Abdurashidova et al., 2022). Despite two decades of work and progressively deeper power spectrum measurements producing better upper limits, a detection of the 21 cm angular power spectrum remains elusive. The power spectrum predicted by theory is quite faint, but all of the experiments listed are specifically designed to have sufficient sensitivity for a detection. However, our ability to detect the EoR signal remains limited by systematic errors both in the instruments and in our analysis techniques. In this thesis we will examine the source of systematics in HERA, outline efforts to mitigate those systematics, and present a first look at an imaging-based power spectrum measurement using data from HERA phase II.

1.3 Interferometry

A telescope's resolution on the sky scales proportionally with the observed wavelength but inversely with the diameter of the receiving element. At radio frequencies, particularly at the frequencies most useful to 21 cm science, the telescope size required to achieve sufficient resolution on the sky is infeasible from a manufacturing perspective. Therefore, astronomers turn to radio interferometers rather than a traditional single-element telescope design. Fundamentally, a radio interferometer is composed of a set (or array) of receiving elements, commonly called antennas, that measure the electric field coming from the sky. Figure 1.2 shows a simple diagram of a two element interferometer, which can be extrapolated to any N-element array of receiving elements. The fundamental measurement of an interferometer is the time-averaged correlation of the electric fields measured by every pair of antennas in the array. This product is called a visibility, and can be written as,

$$V_{ij} = \langle E_i(t) E_j^*(t) \rangle_t, \tag{1.2}$$

where V_{ij} is the visibility for antenna pair (i, j), and E_i is the electric field measured by antenna i. More specifically, E_i is the integral over the incoming electric field collected by the entire receiving area of antenna i. To be more explicit, we can describe this incoming electric field as the Fourier transform of the electric field coming from the sky:

$$E(\mathbf{r},t) = \int E(\theta,t)e^{2\pi i\theta\cdot\mathbf{r}/\lambda}d^2\theta \tag{1.3}$$

where θ is the reference position on the sky and r the position on the ground. We can also think of any antenna as sampling the electric field on the ground, which means that a visibility is sampling the spatial coherence function of the electric field on the ground. We will describe the visibility measurement in more detail in Section 6.1. We use the term baseline to refer to any pair of antennas that are correlated to create a visibility. Baselines are vectors referring to both the distance and orientation of separation between two antennas on the ground. As visibilities are recorded per-baseline rather than per-antenna, it is useful to redefine our spatial coordinate system. We introduce the uv plane, where u and v are spatial coordinates in the instrument plane, measured in units of wavelength, representing the displacement between two antennas. Then, a baseline can be written as a point on the uv plane:

$$\frac{\vec{b}}{\lambda} \equiv (u,v) \tag{1.4}$$

where $\vec{b}$ is a baseline vector and λ is the wavelength of incoming radiation measured in meters. In Chapter 7 we will provide more detail on visibility measurements and how we can use interferometers to create images and power spectrum measurements, which is the final goal of this thesis. The interested reader can also find more a comprehensive overview of radio interferometry in Thompson et al. (2017) and Taylor et al. (1999).

1.4 The Hydrogen Epoch of Reionization Array

The Hydrogen Epoch of Reionization Array (HERA) is a second-generation 21-cm experiment located in the Karoo Desert of South Africa. When completed HERA will consist of 350 dish antennas, although only 201 antennas were online at the time of the data presented for analysis in this thesis. HERA is designed with the aim of producing precision measurements of the 21 cm signal of neutral hydrogen coming from the Epoch of Reionization (EoR) (DeBoer et al., 2017). It observes at frequencies from 50MHz to 250MHz, which corresponds to the 21 cm signal at redshifts $5 < z < 27$. HERA has no moving parts and performs a drift scan observation of a $\sim 10^\circ$ radius patch around zenith. As shown in Figure 1.3, the antennas are configured as a

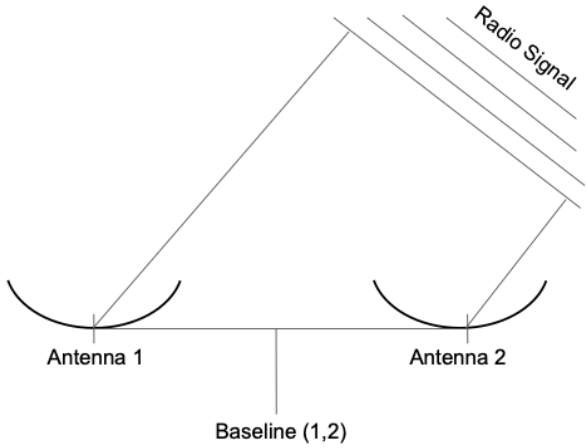


Figure 1.2: A simple diagram of a two element interferometer. Both elements measure the electric field, and the correlation of their measurements constitutes a visibility for that element pair, or baseline.

redundantly arranged compact core of 320 antennas surrounded by 30 outrigger antennas. The compact core and an outrigger antenna are also pictured in Figure 1.4. The core arrangement is designed to aid with redundant calibration techniques, outlined in Section 6.2.1, while the long separation between antennas given by the outriggers is designed to aid with imaging and sky-based calibration techniques, as discussed in Section 6.2.2. Each antenna is 14 meters in diameter and is constructed from a wire mesh formed into a dish shape by PVC pipe spines. This wire mesh reflects radio frequency emission just as a mirror would reflect optical light. The necessary precision of a telescope's reflecting surface scales inversely with the wavelength of the observed radiation, and therefore wire mesh in a radio telescope can have comparable precision to a glass mirror in an optical telescope.

The first phase of HERA used only a subset of antennas and re-purposed dipole antenna feeds from the pre-existing PAPER telescope. This thesis uses data exclusively from HERA phase II, which features an entirely new digital and analog signal chain and switches from a dipole to a Vivaldi antenna feed, shown in Figure 1.4. Vivaldi antenna feeds are suspended above each dish - these feeds are designed not to require a back panel and to minimize spectral structure in the

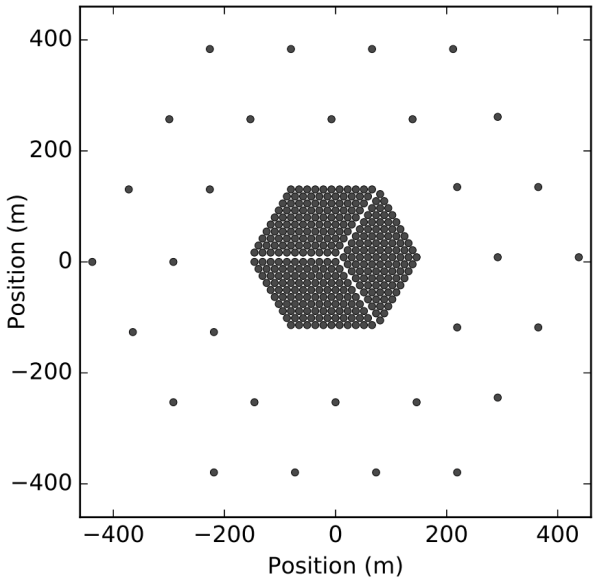


Figure 1.3: A simple diagram of the layout of HERA antennas. The 320 core antennas are separated into three sectors and are surrounded by 30 outrigger antennas.

signal. They are also chosen for their minimal wind load relative to other feed types, which is essential to maintaining a relatively stable feed position while suspended. Section 1.4.1 will outline the instrument's signal chain, and section 1.4.2 will describe HERA's data products and review some necessary terminology.

1.4.1 Signal Chain

This thesis will not go into extensive detail regarding HERA's design and signal chain, but some fundamentals are necessary context to understand many of the systematics outlined in Chapters 2 and 3. For a more detailed description of HERA's design and components, the reader is referred to DeBoer et al. (2017) and Berkhout et al. (2024). For the purposes of our analysis, we will focus on the basic elements shown in Figure 1.6. First, the electric field signal from the sky reflects up from the dish and is captured by the Vivaldi feed, which measures both an East and North polarization. Both polarizations of the signal are then fed into the Front End Module (FEM), which performs an initial amplification step. The FEMs also house a switch that allows observers to turn off the

Figure 1.4: HERA's compact core (top), outrigger antenna (middle) and Phase II Vivaldi feed (bottom).

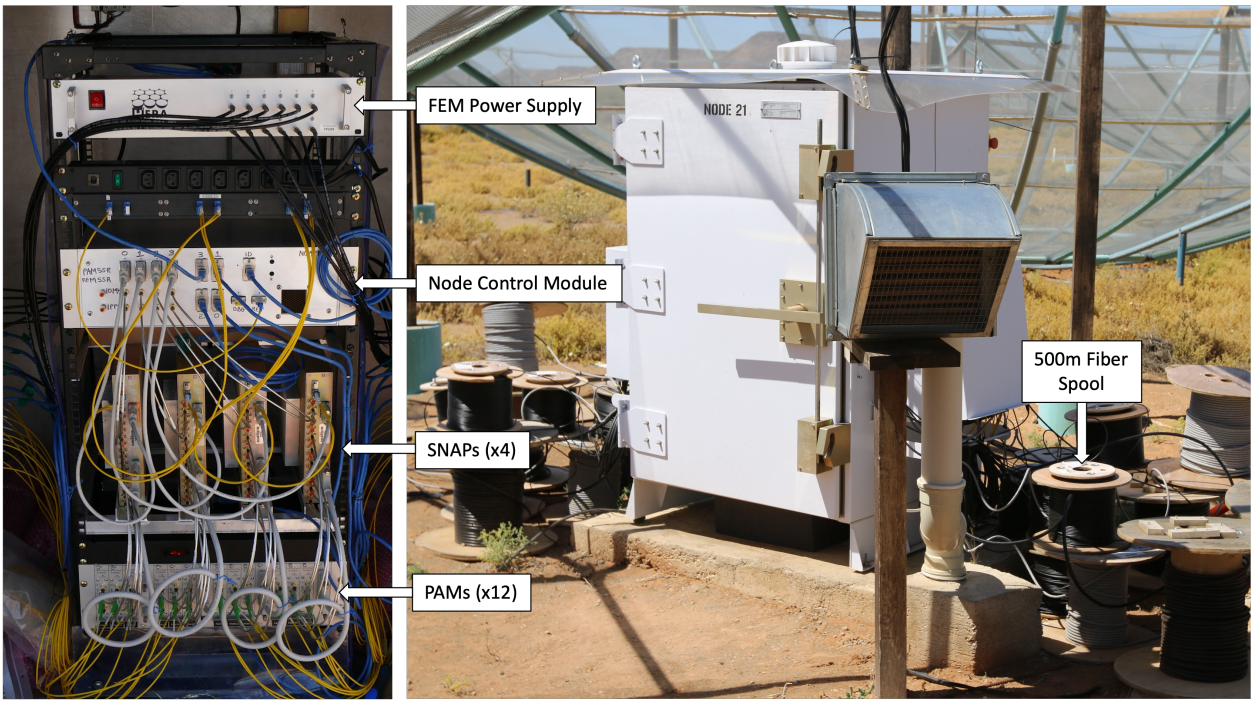


Figure 1.5: The inside and outside of a HERA node box. Starting at the top, the inside (a) shows the FEM power supply unit, node control module, four SNAPs, and twelve PAMs. The outside (b) shows the RFI-shielded node container and the 500m spools of optical cable that come from the antennas and into the node box.

incoming sky signal and instead observe from a 50 Ohm load input. As will be evident in chapters 3 and 4, this feature is essential to several diagnostic procedures.

The FEM then transmits the RF signals for each polarization using RF over Fiber (RFoF) on individual 500 meter long optical fibers to a node box, pictured in Figure 1.5. Each node box is RFI-shielded and contains processing modules for up to 12 antennas, or 24 antenna-polarizations, and node clocks are synchronized by a White Rabbit timing network (Moreira et al., 2009). The signals are passed through a patch panel on the exterior of the node box and then each fiber is fed into a Post-Amplification Module (PAM), which converts the signal back to electrical RF, performs an additional amplification step, and serves as an anti-aliasing filter for the analog-to-digital converter

(ADC) (Berkhout et al., 2024; Razavi-Ghods et al., 2017). Each PAM separately processes both polarizations of one antenna. The signals are then fed into the Smart Network ADC Processors (SNAPs), each of which process both polarizations of three antennas, for a total of four SNAPs in a node box. The SNAPs receive an analog signal over coaxial cable and then do analog-to-digital conversion with an interleaving approach to achieve a 500 Msps sampling rate per input. This high sampling rate is required to prevent aliasing for HERA's bandwidth, and therefore we cannot read out the data anywhere along the signal chain before this point. The signal is then channelized to 6252 channels with a 30kHz frequency resolution and transmitted from the SNAP to the Karoo Array Processor Building (KAPB), which is 10km from the array. As indicated in Figure 1.6, there is also a switch to a digital noise source in the SNAP just after the ADC, allowing observers to test the correlator system by only processing digital noise data rather than sky data. From here, GPU 'X-engines' cross-multiply the data in frequency sub-bands, do additional time and frequency averaging to achieve a workable data volume, and then transmit to a data catcher that receives the signal and writes the visibilities to disk. The entire signal chain from the output of the ADC until data is written to disk is considered part of the correlator. Figure 1.6 shows a simplified diagram of the signal chain that emphasizes the subsystems most relevant to this work.

1.4.2 Data Products + Terminology

In this section we will briefly describe HERA's data products and most relevant features, as well as introduce some important terminology. The primary data product that comes out of the HERA correlator is a complex visibility per baseline, time, polarization, and frequency. We can get the number of baselines N_{bl} in the array from the number of antennas N_a as

$$N_{bl} = \frac{(N_a)(N_a - 1)}{2}. \tag{1.5}$$

There are four possible polarization combinations for each baseline: North-North (NN), East-East (EE), East-North (EN), and North-East (NE). In HERA's case, after all averaging steps each visibility represents 9.8 seconds and 122kHz. For the completed HERA with 350 antennas we will record (61,075 baselines) x (1536 frequency channels) x (4 polarizations) x (1 real + 1 imaginary component) numbers for every time integration, for an average of 76.5 million numbers per second.

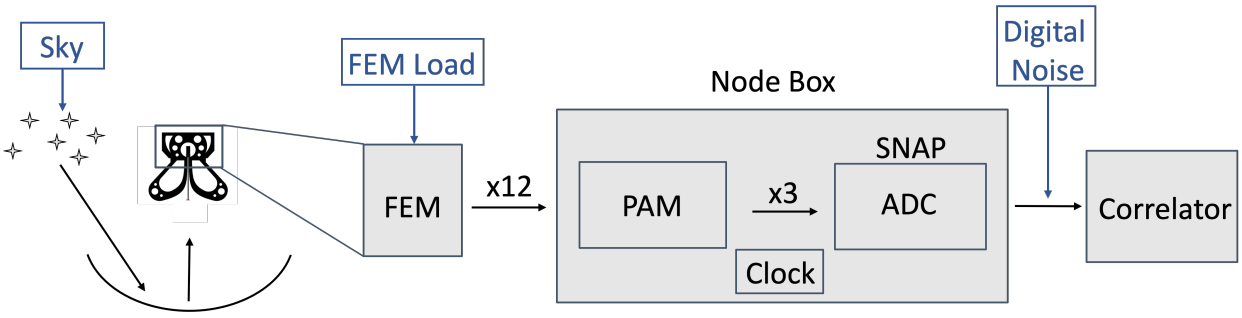


Figure 1.6: A simplified diagram of the HERA signal chain. Data from the sky is reflected up to the Vivaldi feed, passed into the FEM, then data for up to twelve antennas is passed to the node box which houses the PAM and the analog-to-digital converter (ADC), and finally is sent to the correlator. The light blue text indicates the four possible signal sources: sky, FEM load, and digital noise. This diagram represents the signal chain for just one polarization - in reality, each node and SNAP process 24 and 6 antenna-polarizations, respectively.

For the analysis set in this thesis, with 201 active antennas, this reduces to an average of 50.4 million numbers per second, or approximately 21.4 GB of data per hour of observation.

In this analysis we almost exclusively work with the same-polarization combinations NN and EE, and will often refer to them simply as North and East. The visibility that results from the correlation of an antenna with itself is called an auto-correlation, and while not useful for science results these are frequently used in the commissioning process. As described in the previous section, HERA is able to switch off the sky signal and write out data either from an analog load source housed in the FEM or from a digital noise source located just after the ADC. Additionally, input from the FEM can be turned off all-together, in which case the signal comes from the load of the amplifiers in the PAM. This closely resembles load from the FEM, but is slightly lower power. This setting has been useful in HERA diagnostics, but will not be discussed further for the purposes of this thesis. Going forward, we will use data from three of these observing modes: sky data, FEM load data[1], and digital noise data.

To acquaint the reader with the basic features of HERA data, Figure 1.7 shows the auto-

[1]The FEM load data has the same characteristics as a noise source, but we use the term load going forward to better differentiate between FEM load and digital noise data.

correlation spectra for two individual antennas as well as the cross-correlation spectrum of their composite baseline. Looking first at the auto-correlations we note that there is a relatively smooth band across the whole spectrum with intermittent bright spikes. These spikes are due to Radio Frequency Interference (RFI), which refers to any human-generated radio emission that is detected by the telescope. Two of the primary sources of RFI, FM radio and digital TV, are indicated in purple and grey, respectively. Most of the other spikes are due to known emitters that can be identified in South Africa's radio frequency allocation catalog and are relatively stable night-to-night. HERA's site was specifically chosen to be in a radio-quiet area far from any major population centers, but we nonetheless observe spurious RFI from non-stable sources such as personal electronic devices (phones, cameras, watches, etc.), reflections off of overhead airplanes, satellite emission, and other nearby instruments. The frequency range used for analysis in this thesis is 154.6MHz to 175.0MHz (or redshift $7.1 < z < 8.2$), indicated by the green band, and is chosen both for its relatively smooth and flat bandpass shape and for the lack of any major RFI emitters in that range.

Figure 1.8 shows a sample auto-correlation when HERA is observing FEM load data. There are three bright spikes that are due to emission from the instrument's clock system, but otherwise we don't see RFI spikes as we are not observing emission coming from the sky. There are some small spikes visible here which are due to a systematic that will be discussed in Section 3.2.3. Lastly, it is worth noting that while the FEM load source does not have a frequency response the spectrum in Figure 1.8 has a slope with frequency. This is because we apply a 'whitening' filter to HERA data to flatten the spectrum of sky data, and when applied to the already flat spectrum of noise data it actually introduces a shape.

1.5 Barriers to Measurement

Before continuing it is worthwhile to outline several of the inherent challenges associated with radio interferometry and 21 cm cosmology. To start, there are several assumptions typical of interferometric analysis techniques that may not always be satisfied in practice. First, one often assumes that we can ignore the effects of sky curvature, called a flat-sky approximation. In practice this assumption holds as long as the observational field of view is small enough, which is not always

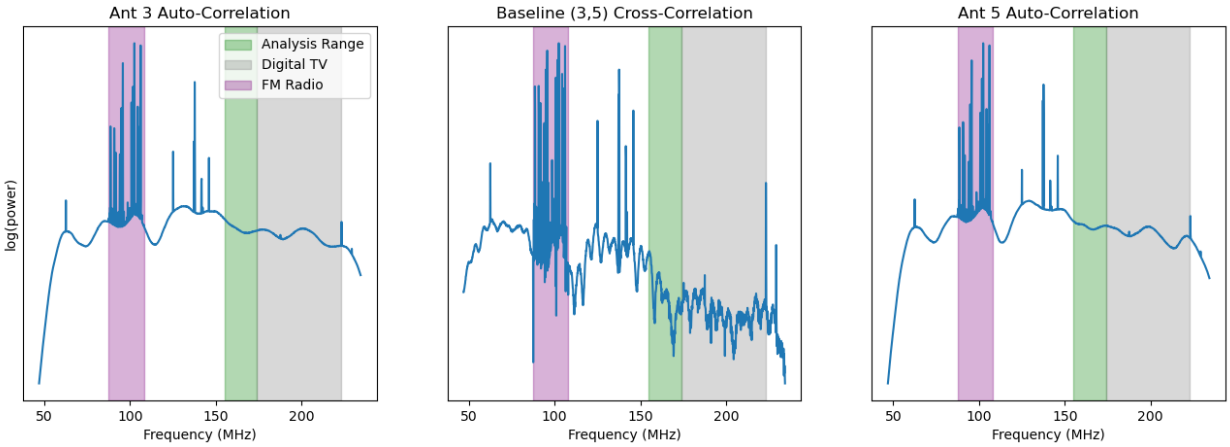


Figure 1.7: Sample auto-correlation and cross-correlation spectra from HERA sky data showing some key features. The left and right panel show individual auto-correlations, and the middle panel shows the cross-correlation between those two antennas. The bright spikes are Radio Frequency Interference (RFI). The FM radio band is shown in purple, digital TV in grey, and the analysis band used for this thesis is shown in green. All data shown here is the North polarization.

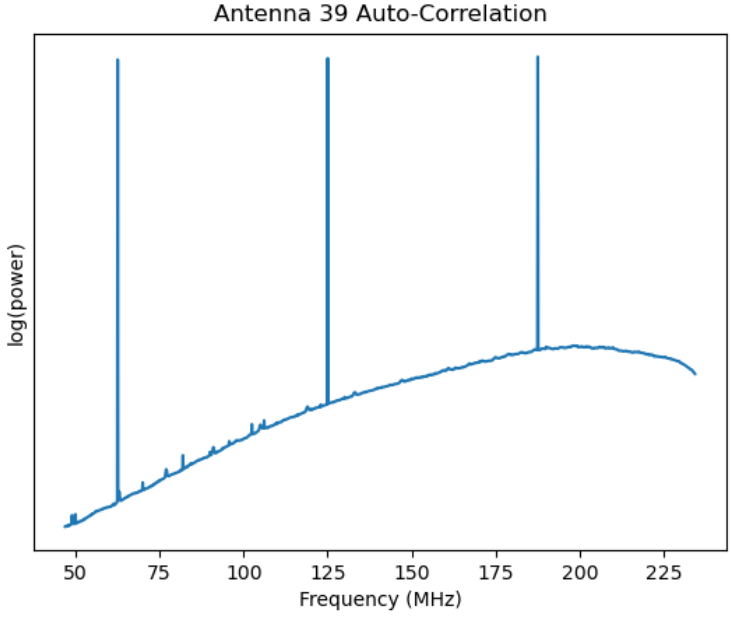


Figure 1.8: Sample auto-correlation spectrum from HERA FEM load data. The three bright lines are due to emission from HERA's clock system.

the case, and so more advanced analysis techniques may be necessary.[2] Second, we assume that all elements in our interferometer lie within a flat plane that is parallel to the sky plane. Commissioners of HERA took great care to place antennas as close to the same height as possible, but there are still discrepancies that need to be accounted for with care during analysis (Berkhout et al., 2024). Lastly, we assume that the incoming radiation is propagating through a non-refractive medium. In practice, the ionosphere introduces phase delays that cause positional displacements in observed sources (Jordan et al., 2017; Morales and Matejek, 2009). The intensity of ionospheric distortions varies based on location on the Earth and weather, and therefore the significance of this effect is highly variable.

When attempting to detect the 21 cm signal, which is incredibly faint, we also face substantial challenge in avoiding other sources of radio emission. Non-cosmological radio emission from the sky, predominantly synchrotron emission from our own and distant galaxies, is commonly referred to as foreground emission. At our observational frequencies foreground sources are approximately five orders of magnitude brighter than the 21 cm signal (Santos et al., 2005; Bernardi et al., 2010), and therefore can easily obscure our desired signal. There are two primary methods of mitigating foreground emission. First, we can use models of known sources on the sky and directly subtract them from our data. This requires precision knowledge of our sky model and of our instrumental response, but is nonetheless effective and will be discussed further in Chapter 5. Second, we can utilize the fact that foreground sources are generally quite spectrally smooth, whereas the 21 cm signal is highly spectrally structured. This allows us to use Fourier decomposition to separate our desired signal from the foregrounds.

In addition to unwanted astronomical radio emission, we also observe high levels of human-generated RFI. This is concerning for several reasons: it adds artificial power to our measurements, it is unstable in time and therefore unpredictable, and it introduces substantial spectral structure to our data. This last point is particularly concerning because such highly spectral structure is not avoidable in the way foreground sources are. RFI is primarily mitigated through data flagging, or removing contaminated data before analysis. This requires highly sophisticated flagging algorithms, as even very low level RFI remaining in the data can cause significant issues with power spectrum

[2]This is part of the original motivation behind the development of the analysis software package `FHD` that we use in this thesis, which fully accounts for a curved sky (Sullivan et al., 2012)

measurements (Wilensky et al., 2019).

Lastly, all instruments introduce some level of systematic and chromatic effects to a signal. Despite a careful design and high precision components no instrument can be constructed to perfection, and there will always be systematic effects. Unlike the previous barriers that are common across the field, every instrument has a unique set of systematics that requires a unique set of analysis and mitigation tools. Our ability to perform a power spectrum measurement relies heavily on minimizing systematic effects and collecting high volumes of high quality data. The primary purpose of this thesis will be to examine the various types of systematics HERA experiences, how we can identify and characterize them, and either resolve them or attempt to mitigate them through analysis techniques.

1.6 Thesis Outline

Throughout this thesis we will examine the instrumental character of HERA, identify data quality concerns, and pursue an imaging based power spectrum result. We will start in Chapter 2 by outlining new methods and statistical tools for performing data quality inspection and detailing how they are implemented as part of an automated data processing pipeline. In Chapter 3 we will overview several of the systematics we identified in HERA during the commissioning process and discuss how those systematics are dealt with either instrumentally or through analysis. Next, in Chapter 4 we will provide a detailed discussion of one particular systematic, excess correlation, that impeded analysis prospects and describe how a resolution was achieved. In Chapter 5 we will describe our data selection procedure and present the final analysis data set. Then, Chapter 6 will review the fundamentals of interferometric calibration and provide results validating the success of our calibration scheme. Chapter 7 will then briefly describe the process of interferometric imaging and present images of HERA data. In Chapter 8 we will detail our power spectrum calculation and present a set of imaging-based power spectrum results. Finally, we will conclude by summarizing our findings and their implications for the HERA telescope.

Chapter 2

DATA QUALITY INSPECTION

Portions of this chapter are from Storer et al. (2022).

As discussed in Chapter 1, the 21 cm fluctuation signal is four to five orders of magnitude fainter than the bright radio foregrounds. Attempts to measure the power spectrum using radio interferometers like HERA must therefore be executed with high sensitivity and precision analysis techniques in order to realistically achieve a detection (Liu and Shaw, 2020). Reaching sufficient sensitivity requires an interferometer with a large number of antennas observing for months, which introduces a high level of complexity to the system. The need for high sensitivity and precision results in thousands of interconnected subsystems that must be commissioned by a relatively small number of people poses a significant challenge. Additionally, due to the faintness of the signal, low level systematics that might be deemed negligible in other astronomical applications can have the potential to leak into the power spectrum and obscure the 21 cm signal. Therefore, systematics must either be resolved, methodically avoided, or directly removed in order to achieve sufficiently clean data. Some examples of contaminants common in these types of interferometers include adverse primary beam effects (Beardsley et al., 2016b; Ewall-Wice et al., 2016; Fagnoni et al., 2020; Joseph et al., 2019; Chokshi et al., 2021), internal reflections (Ewall-Wice et al., 2016; Beardsley et al., 2016a; Kern et al., 2019, 2020b,a), RFI (Wilensky et al., 2020; Whitler et al., 2019), and any analog or digital systematics resulting from the specific design and configuration of the array and its component electronics (Benkevitch et al., 2016; de Gasperin et al., 2019; Star, 2020).

While there are some systematics we can avoid using clever analysis techniques (see Kern et al. (2020b) for example), we manage most systematics by directly removing the affected antennas from the raw data. This requires us to identify and flag any data exhibiting a known malfunction, and develop methodologies for catching new or previously unidentified systematic effects. While the primary goal of flagging data is to produce the cleanest possible data for analysis, it has the added benefit of providing information regarding the scope and character of prevalent issues to the

commissioning team, which is essential to our ultimate goal of finding and resolving the sources of the problems. Section 2.1 describes the set of quantitative metrics used to automatically flag malfunctioning antennas, and Section 2.2 will outline the set of data visualization tools developed to help observers identify new system failures.

2.1 Antenna Metrics

While manual inspection of all data would likely be an effective approach to antenna flagging, for large-N interferometers like HERA the data volume poses a problem to this approach. When completed, HERA will have 350 individual dishes each with a dual-polarization signal chain including several analog and digital subcomponents. Even just for the 120 antennas included in the typical night of data, manual flagging would involve assessing 28,800 baselines, each of which has 1536 frequency bins and thousands of time integrations. Therefore, the hands-on time involved is neither practical nor reproducible, and so an automated approach is preferred.

Our approach is to design a set of statistical metrics based on common failure modes of the HERA data. We also optimize the metrics to use a limited fraction of the data so they are usable in a real time pipeline. We break these metrics into two categories: cross-correlation metrics (per-antenna values calculated using all baselines), and auto-correlation metrics (per-antenna values calculated using only the auto-correlations). These two methods have complementary advantages. The cross-correlation metrics require a larger data volume, but give us insight into the performance of the whole array and all component subsystems, whereas the auto-correlation metrics are optimized to use a small amount of data, and help assess functionality of individual array components. The remainder of this section will outline how each of our metrics is designed to catch one or more known failure modes in the smallest amount of data possible and validate that the automation procedure flags these failures effectively. We also use tools such as simulated noise and comparisons with manual flags to aid in validating our procedure. While these metrics were designed based on HERA data, it is worthwhile to note that both the approach and the metrics themselves are applicable to any large interferometric array.

The data set used for the remainder of this section was collected on September 29, 2020 (JD 2459122) when there were 105 antennas online, and is a different data set than the one used for the primary data analysis in this thesis. For the remainder of this thesis these two data sets will

be referred to as the commissioning and analysis data sets, respectively. It is important to note that the features and systematics discussed in this section are persistent in the most recent data and therefore still pertain to our analysis data set. Figure 2.1 shows the layout of the array at the time of the commissioning data set and illustrates the node architecture overlain with antenna cataloging developed in this paper. These flags were produced using almost ten hours of data from this night. The high fraction of malfunctioning antennas was partly attributable to limited site access due to the COVID-19 pandemic.

Flagging of misbehaving antennas is necessary in preventing them from impacting calibration, imaging, or power spectrum calculation steps. Here we define a misbehavior to be any feature which makes an antenna unusual when compared to others. In practical terms, the pathologies of antenna malfunction are not limited to the signal chain at the antenna, but could manifest anywhere in the system up to the output of the correlator. Depending on where along the signal chain the pathology lies, we might see evidence of it in either the auto-correlations, the cross-correlations, or both. For example, if an antenna's timing was out of sync with another's, its auto-correlations might look fine, but its cross-correlations would highlight this systematic. In particular, as an interferometric array grows in size, it is vital to track the health of the entire array, not just the auto-correlations or the cross-correlations in isolation. Section 2.1.1 will outline the cross-correlation metrics used on HERA data, and Section 2.1.2 will outline the auto-correlation metrics.

2.1.1 Cross-Correlation Metrics

Our most generalized metric for assessing antenna function tests how well antennas correlate with each other. There are many reasons antennas might not correlate: one of the gain stages might be broken, cables might be hooked up incorrectly, or might not be phase-aligned with other functional antennas. Assessment of cross-correlations in uncalibrated data is challenging because the correlations can vary widely depending on the baseline length and sky configuration. In particular, one must be able to tell the difference between baselines that include both the expected sky signal and noise versus baselines that include only noise. A metric which is robust against these and other

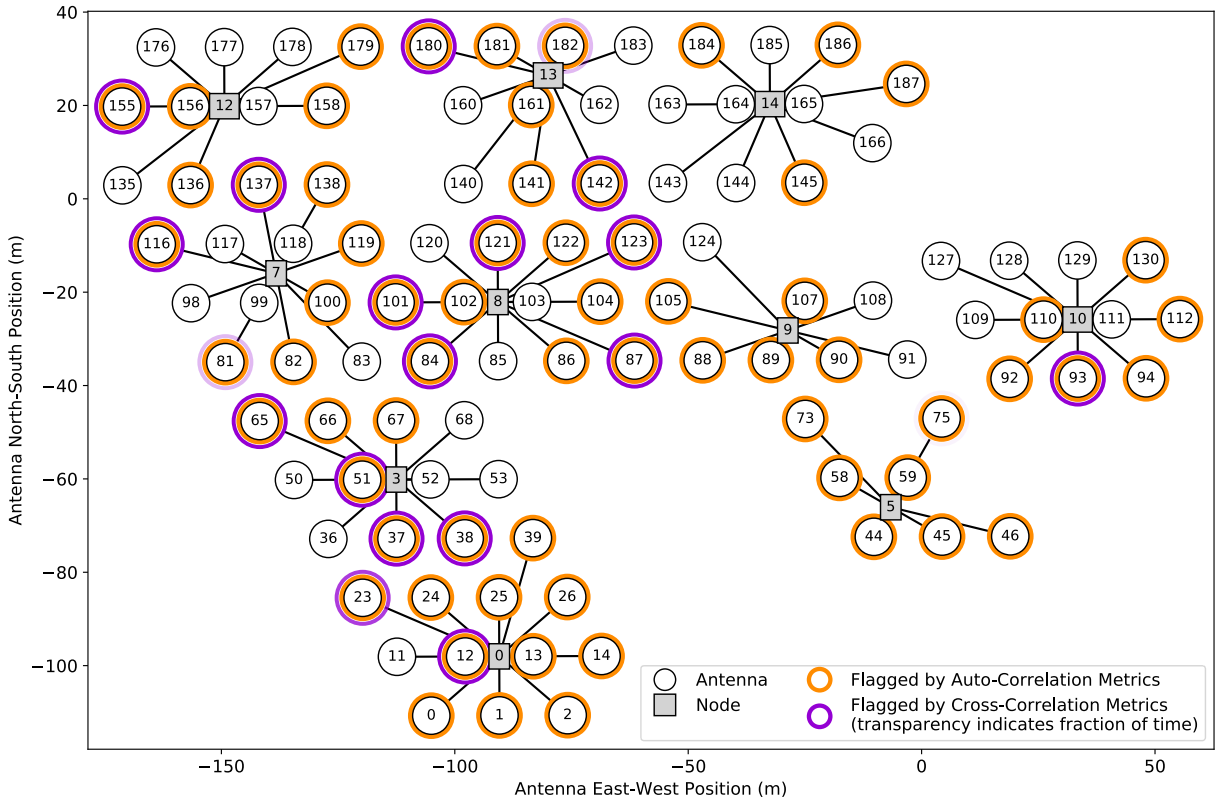


Figure 2.1: Array layout and antenna quality statuses on Sept 29, 2020 (JD 2459122) as determined by the algorithms laid out in Sections 2.1.1 and 2.1.2. In HERA, each antenna is connected to a node, which contains amplifiers, digitizers, and the F-engine (which Fourier transforms the data from a time-series to a frequency spectrum). Node connections are denoted here by solid black lines. Most of the elements are in the Southwest sector of the split-hexagonal array configuration, with a few in the Northwest and East sectors (Dillon and Parsons, 2016; DeBoer et al., 2017). Only actively instrumented antennas are drawn; many more dishes had been built by this point.

challenges is the normalized and averaged per-frequency correlation matrix C_{ij}:

$$C_{ij} \equiv \left\langle \frac{V_{ij}^{\text{even}} V_{ij}^{\text{odd}*}}{\left|V_{ij}^{\text{even}}\right| \left|V_{ij}^{\text{odd}}\right|} \right\rangle_{t,\nu} \tag{2.1}$$

where $\langle\rangle_{t,\nu}$ represents an average over time and frequency, V_{ij}^{even} and V_{ij}^{odd} are pairs of measurements of the same sky with independent noise, and i and j are antenna indices, such that ij represents an

individual baseline.[1] This holds for any correlator outputs separated by timescales short enough that the sky will not rotate appreciably, such that we can assume time adjacent visibilities are observing the same sky signal but with independent noise realizations. In HERA's case we are able to utilize our specific correlator output to construct even and odd visibilities that are interleaved on a 100 ms timescale. To explain this, we digress briefly into the output of the HERA correlator. In its last stage of operation, antenna voltage spectra are cross-multiplied and accumulated over very fine time intervals. These visibilities can be averaged over the full 9.6 second integration before being written to disk. However, in order to improve our estimate of noise and to aid in the estimation of power spectra without a thermal noise bias, we split these 96 spectra into two interleaved groups, even and odd, and sum them independently before writing them to disk. Thus, each is essentially 4.8 seconds of integrated sensitivity, spread over 9.6 seconds of observation. [2]

Division by the visibility amplitude in Equation 2.1 minimizes the impact of very bright RFI that might differ between even and odd visibilities and dominate the statistics. We experimented with alternative statistics like a maximum and a median to compress across time and frequency but found that with the normalized correlation a simple average was sufficiently robust.

Due to our chosen normalization, the correlation metric measures the phase correlation between visibilities, and is unaffected by overall amplitudes. If the phases are noise-like, the antennas will be uncorrelated and this value will average down to zero. If V_{ij}^{even} and V_{ij}^{odd} are strongly correlated, we expect this statistic to be near one. The normalization in Equation 2.1 is particularly useful in mitigating the effects of RFI and imperfect power equalization between antennas.

We can visualize the correlation matrix C_{ij} with each baseline pair ij as an individual pixel, such that the auto-correlations fall along the diagonal. A schematic of this visualization is shown in Figure 2.2. To emphasize any patterns related to electronic connectivity, antennas are organized by their node connection, and within that by their sub-node level electronic connections. Node

[1]The visibilities already have some inherent time averaging, but we include an additional time average in this notation to indicate that this metric can be calculated over any time-averaged interval. This could only include the time-averaging already inherent to the visibilities, or could include much longer averages across many visibility measurements.

[2]In HERA's case, the V^{even} and V^{odd} terms in Equation 2.1 reference these finely interleaved even and odd correlator outputs. However, one can use any set of time-adjacent output visibilities, including the full 9.6 second integrated ones. This is a much coarser interleave, but we find the results to be qualitatively the same in either case. For the remainder of this thesis, V^{even} and V^{odd} will refer to the finely interleaved case.

boundaries are denoted by light blue lines. While the nodal structure used here is specific to HERA, the principal of organizing by electronic connectivity is a generalizable technique for highlighting patterns that may be due to systematics in particular parts of the system. Additionally, plotting the matrices in this way allows us to assess the system health on an array-wide level and on an individual antenna level all in one plot, which is increasingly useful as the size of an array grows.

To study the performance of any single antenna it is useful to form a per-antenna cross-correlation metric C_i by averaging over all baselines that include a given antenna:

$$C_i \equiv \frac{1}{N_{\text{ants}} - 1} \sum_{j \neq i} C_{ij}. \tag{2.2}$$

where N_{ants} is the number of antennas. We calculate this metric separately for all four instrumental visibility polarizations: NN, EE, EN, NE. The panels below each matrix in Figure 2.2 show this per-antenna average correlation metric C_{ij} for each polarization.

Next, Figure 2.3 shows a visualization of C_{ij} for all four polarizations, using a representative subset of antennas for simplicity. Here the values have a bimodal distribution (most obvious in the East-East and North-North polarizations), where most antennas are either showing a consistently low metric value, or are close to the array average. This bimodality is also clear in the lower panels showing the per-antenna metric C_i. Here we see more clearly that there is a fairly stable array-level average metric value for each polarization, with a handful of antennas appearing as outliers. The dashed line in the lower panels shows the threshold that is used for antenna flagging, with the points below the threshold marked in red. Note that for this metric we only flag antennas if their metric is below the threshold in the North-North or East-East polarization. There are three primary features to note in Figure 2.3. First, we see that antennas 51 and 87 are lower than the array average in the North-North and East-East polarizations, but are higher than average in the other two polarizations. These points are marked in cyan in the lower panel. The reason for this pathology is that antennas 51 and 87 are cross-polarized, meaning that the cables carrying the East and North polarizations are swapped somewhere along the cable path - this will be discussed further in the next section. Second, we can see that the typical value of C_{ij} is higher in the East-East polarization than in the North-North polarization. This is because of the elevated signal-to-noise ratio observed in the East-East polarization due to contributions from the galactic plane and diffuse emission. Lastly, we observe that there appears to be a slight increase in the average metric power

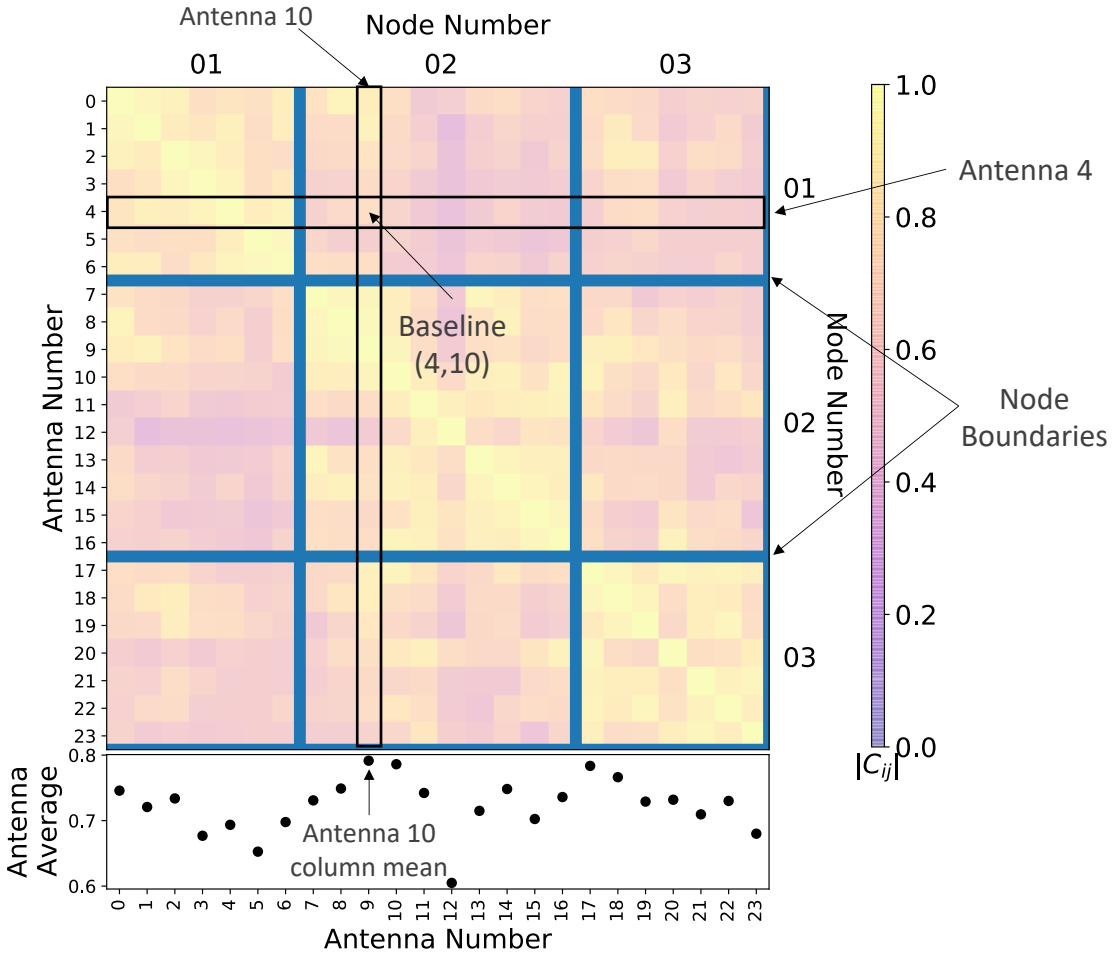


Figure 2.2: Schematic showing how we visually represent the matrix C_{ij} and the per-antenna metric C_i. Each pixel in the matrix represents an individual baseline ij, identified by the two antennas that pixel corresponds to. The light blue lines denote the node boundaries, and antennas within each node are additionally sorted by their sub-node level electronic connections. The panel below the matrix shows the per-antenna average, calculated as the column mean for each antenna. (Note that in practice this average is computed iteratively.)

for baselines within the same node compared to baselines with antennas in different nodes. Chapter 4 will thoroughly explore the nature of this node-based structure.

2.1.1.1 Identifying Cross-Polarized Antennas

As we have already seen in passing, the correlation metric C_{ij} clearly identifies cross-polarized antennas. Here, cross-polarized means that the physical cables carrying the East and North polarization measurements got swapped in the field. When things are hooked up correctly, we expect to see a stronger correlation between matching polarizations (i.e. EE[3] and NN), and a weaker correlation between different polarizations. Cross-polarized antennas have the opposite situation, with stronger correlation in EN and NE.

We identify this situation automatically with a cross-polarization metric formed from the difference between four polarization combinations in the per-antenna correlation metric:

$$C_i^{P_\parallel - P_\times} \equiv \frac{1}{N_{\text{ants}} - 1} \sum_{j \neq i} (C_{ij}^{P_\parallel} - C_{ij}^{P_\times}). \tag{2.3}$$

where $P_\parallel$ is either the EE or NN polarization, and $P_\times$ is either the NE or EN polarization.

We then calculate our cross-polarization metric as the maximum of the four combinations of same-polarization and opposite-polarization visibilities:

$$R_i = \max \left\{ C_i^{NN-NE}, C_i^{NN-EN}, C_i^{EE-NE}, C_i^{EE-EN} \right\} \tag{2.4}$$

We take the maximum because it's possible to get negative values for some of the $C_i^{P_\parallel - P_\times}$ when one polarization is dead and the other is not. However, when all four values are negative (i.e. a negative maximum), then the antenna is likely cross-polarized. In Figure 2.4 we show each of the four differences of C_{ij}. Two antennas, 51 and 87, show negative values in all four combinations, indicating swapped cables. Three other antennas—37, 38, and 101—show up negative in two polarizations, which indicate a single dead polarization, rather than a swap.

2.1.1.2 Identifying and Removing Antennas in Practice

Using our correlation metric C_i defined in Equation 2.2 and our cross-polarization statistic R_i defined in Equation 2.4 we can implement an iterative algorithm to flag and remove broken and

[3]HERA antennas, being fixed, are referred to by their cardinal directions. This avoids much confusion.

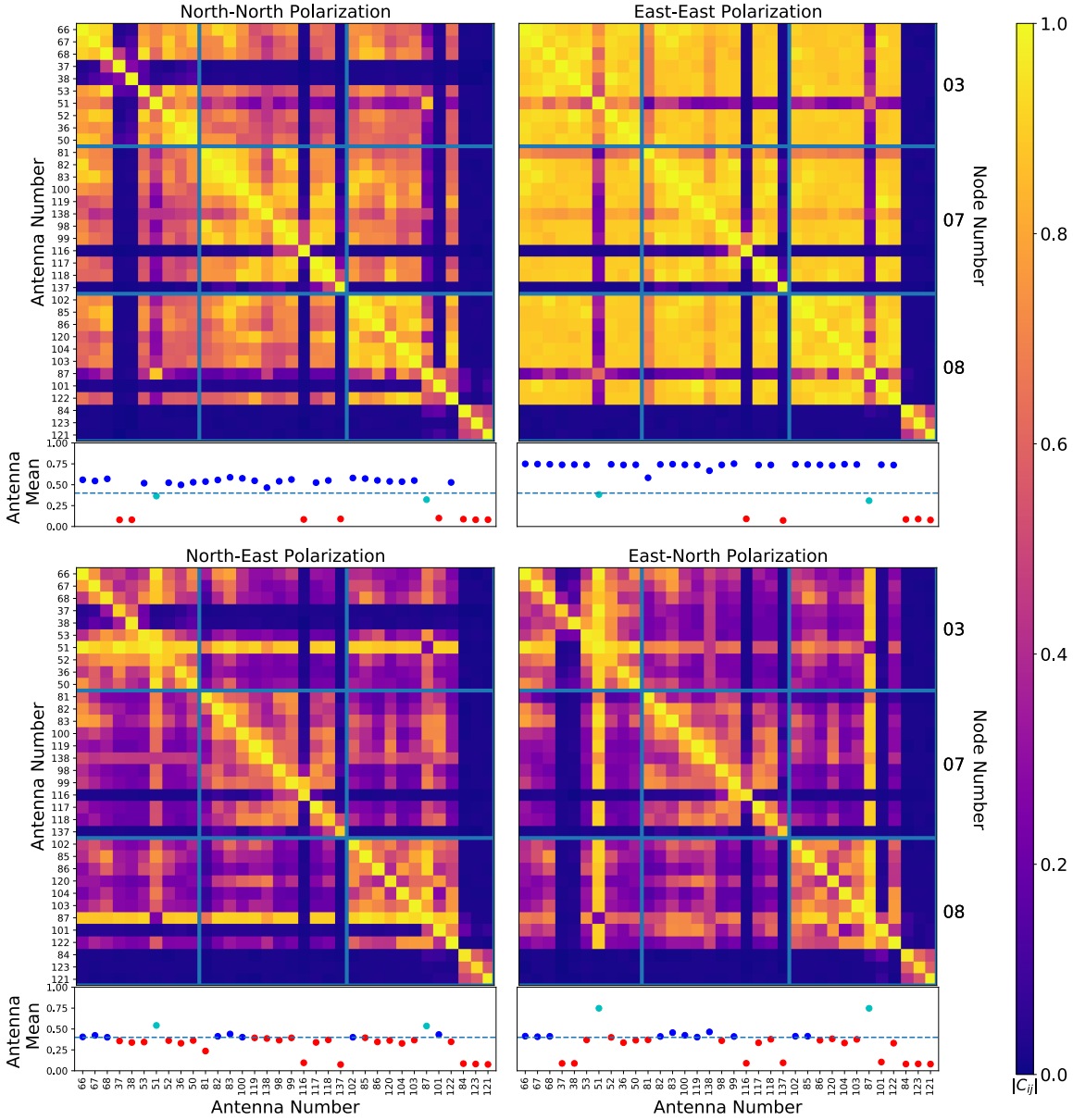


Figure 2.3: The correlation metric C_{ij} as calculated in Equation 2.1. Light blue lines denote the boundaries between nodes. The per-antenna average metric C_i as calculated in Equation 2.2 is plotted below each matrix. The dashed line indicates the flagging threshold, such that blue dots indicate unflagged antennas, red indicates flagged antennas, and cyan indicates antennas identified as being cross-polarized (see Section 2.1.1.1). Note that we do not use the North-East and East-North polarizations for antenna flagging.

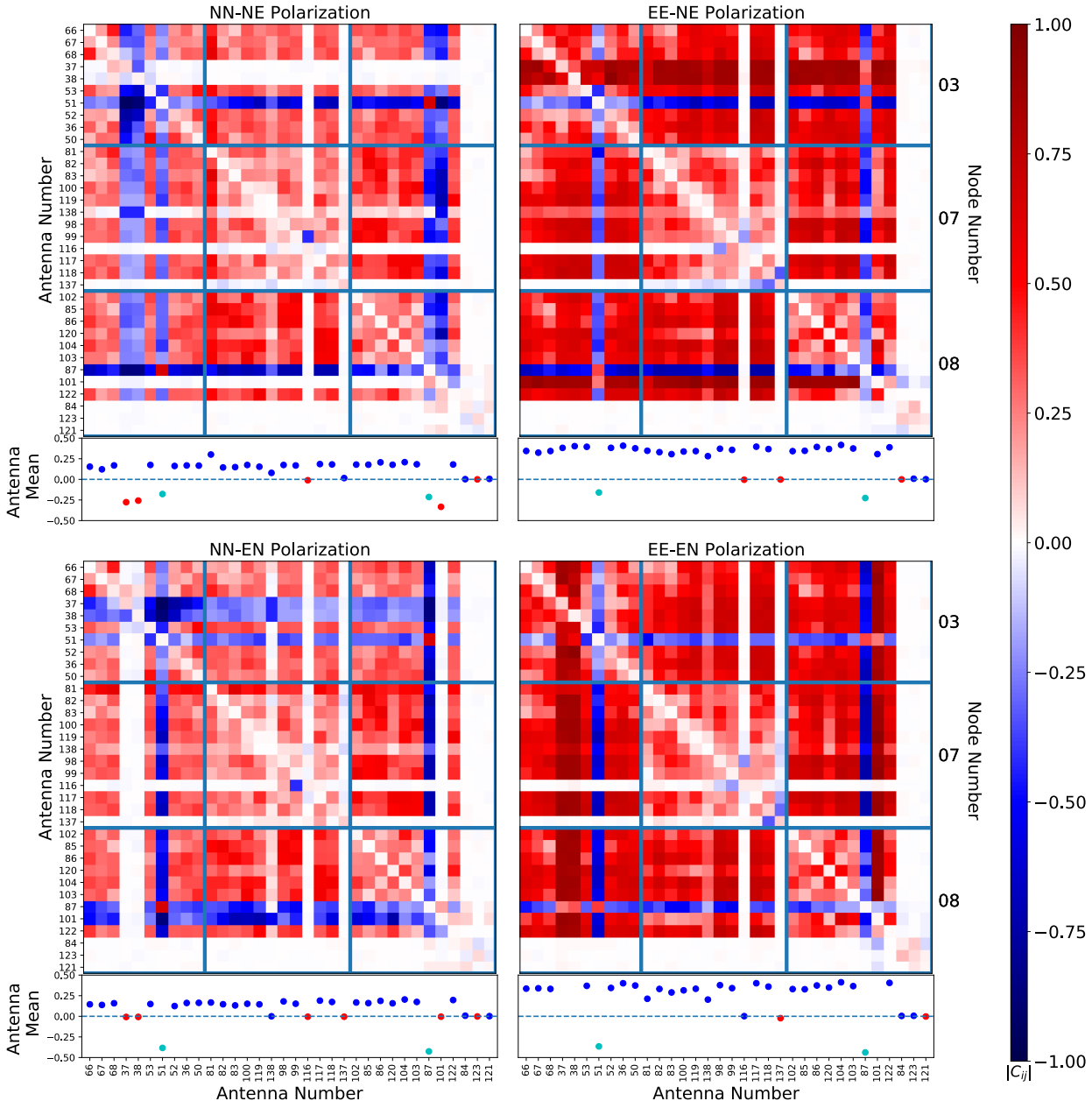


Figure 2.4: The cross-polarization metric defined in Equation 2.3. Any antenna with an average metric R_i that is negative in all four polarization combinations is deemed cross-polarized and marked in the lower panels in cyan. Antennas with a positive antenna mean are marked with blue dots, and those with a negative mean are marked with red dots. Here antennas 51 and 87 are cross-polarized.

cross-polarized antennas. In Figure 2.3 we clearly saw that dead antennas have a value of C_i very near zero. As a result, when we calculate C_i for functional antennas by averaging over all constituent baselines, the low correlation between a functional and a dead antenna will decrease the overall value of C_i for the functional antenna. In the case where only a couple of antennas are broken among the whole array this may be tolerable, but it is possible for this bias to cause functional antennas to look much worse than they are, and to potentially drop below the flagging threshold.

To prevent the expected value of our metric from being biased by dead antennas, we implement an iterative metric calculation and flagging approach, outlined in Algorithm 2.5. First, we calculate C_i for all antennas and identify any that are completely dead (i.e. C_i=0) and remove them. Then, recalculate C_i and R_i for all antennas, identify and remove the worst antenna if it falls below the threshold. We continue with this recalculation and reassessment of the metrics until all remaining antennas are above the threshold in both metrics. Figure 2.6 shows a comparison between the values of C_i calculated by directly averaging C_{ij} for each antenna versus using the iterative algorithm. We see clearly from this figure that implementing an iterative approach brings our data into a truly bimodal realm where establishing a threshold is straightforward. Based on the observed values, we set an empirical threshold of $C_i = 0.4$, such that any antennas below that value will be flagged and removed. Note that the two antennas marked in cyan are both cross-polarized, so their value near the threshold is not worrisome. As noted in section 2.1.1.1, these points are flagged for having a maximum value of R_i below zero. This iterative approach to flagging is robust against varying proportions of broken antennas, which is essential for flagging during the commissioning phase of an array. While the iterative approach somewhat increases computation time, we find the trade-off to be worthwhile. Even with the iterative approach, flagging based on the cross-correlation metrics scales at worst with the number of visibilities, which we find reasonable. Our most computationally expensive step is simply reading in all of the data. In the case where this step becomes computationally prohibitive in a real-time pipeline, we may take advantage of the time-stability of the correlation metric and calculate antenna flags on a sparser time interval. As it is computationally infeasible to hold data for all baselines over the whole night in memory at once, we reserve time domain data quality assessments for auto-correlations only.

It is also relevant to note that we only use the North-North and East-East polarizations during

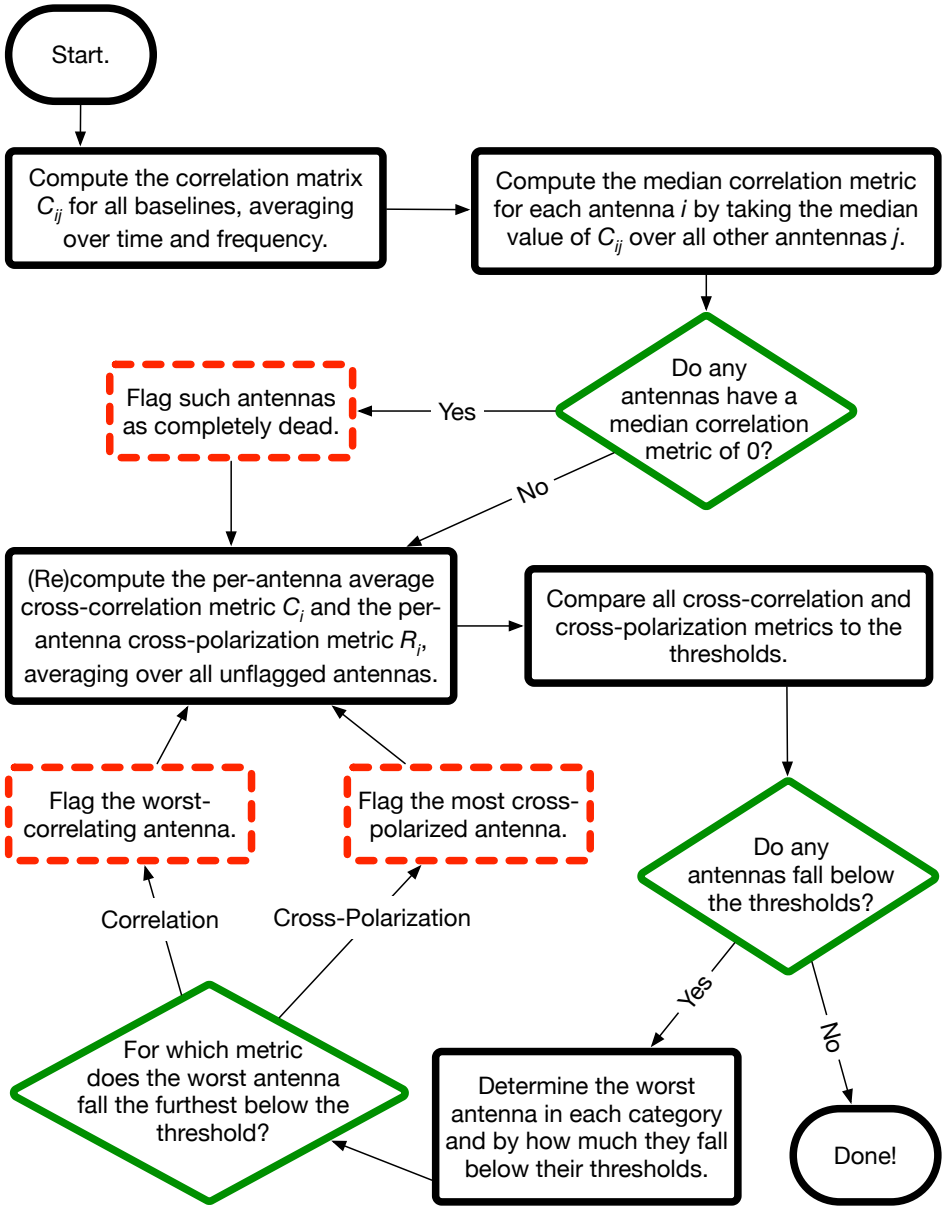


Figure 2.5: Pseudocode flowchart of the cross-correlation metrics flagging algorithm. Figure credit to Josh Dillon, Storer et al. (2022).

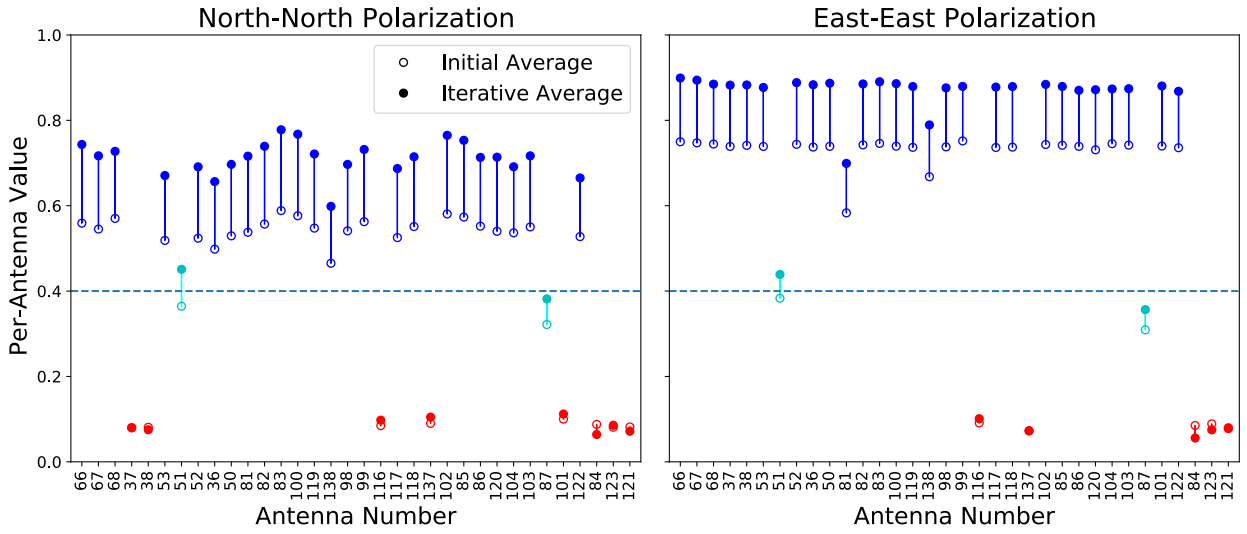


Figure 2.6: Comparison of the final value of C_i calculated for each antenna using a direct average over C_{ij} versus the iterative calculation outlined in Algorithm 2.5. We see that using the iterative algorithm helps create a clearer boundary between functional and nonfunctional antennas. Antennas marked in cyan are those that were flagged for being cross-polarized. This plot uses the same representative subset of antennas used in Figures 2.3 and 4.2.

antenna flagging. There are multiple reasons for this. First, we see in Figure 2.7 that the expected value of C_{ij} is higher in same-polarization correlations compared to cross-polarization correlations. The larger separation in expected value between functional and dead antennas leads to a more robust flagging threshold. Second, we have no evidence in HERA data to indicate the existence of systematics that appear only in correlations between different polarizations. Therefore, flagging only on the same-polarizations allows for clearer distinction between functional and broken antennas without missing any known failure modes of the system.

2.1.2 Auto-Correlation Metrics

While the correlation metrics provide essential checks on overall system health and can diagnose particular per-antenna failures, there are many ways an antenna fail that it will not make evident - Chapter 3 will provide several examples of these failures. Therefore, we must also implement

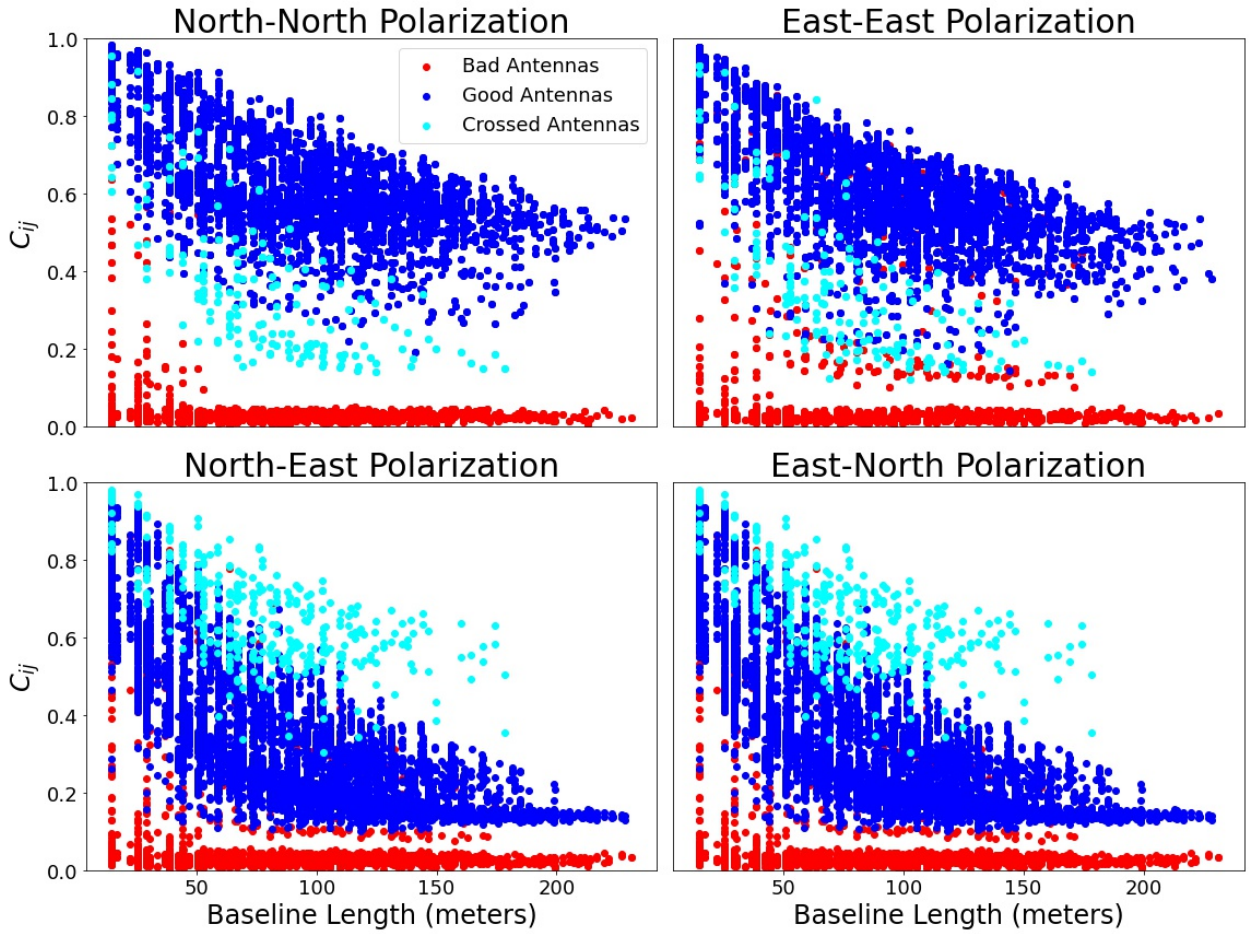


Figure 2.7: A scatter plot of the correlation metric C_{ij} versus baseline length for all four polarization combinations. The classification based on constituent antennas is indicated with color. We note that if an antenna fails in either polarization it is marked bad in all four panels.

systems for more closely inspecting antenna behavior, which we base almost exclusively on auto-correlation data. We choose to focus on auto-correlations V_{ii} for two reasons. The first is data volume. The number of auto-correlations scales with N_{ant} while the number of visibilities scales with N_{ant}^2—far too big to load into memory at once for a whole night of data. Second, because our goal is to identify malfunctioning antennas before calibration, we focus on auto-correlations because they are easier to compare without calibration.

During the commissioning of HERA we created two primary pipelines for catching per-antenna failures using the auto-correlations. The first is an automated system that runs in parallel with the correlation metrics pipeline. The per-antenna pipeline does outlier detection on the auto-correlation spectra and automatically flags antennas that do not match their peers. Due to a careful iterative approach, this system is robust even when a large fraction of antennas fail. This thesis will not discuss the auto-correlation metrics in detail, but an interested reader can find extensive discussion of this system in (Storer et al., 2022). The second is a suite of data inspection plots that is automatically produced every night, and will be discussed at length in the next section.

2.2 Nightly Data Inspection

In addition to the fully automated antenna flagging systems outlined above, we also saw need for a more generalized system that would allow observers to identify new and recurring systematics that would not be caught by the automated system. To address this issue, we implemented a suite of Jupyter Notebooks that are produced as part of the Real Time Processing (RTP) system that runs with every night of observation. These notebooks primarily utilize auto-correlation data to produce a wide variety of diagnostic plots and data visualizations, but they incorporate some information from cross-correlations as well. They are then inspected by nightly observers who can report any new issues to the commissioning team, and are also used by the analysis team to help select nights of observation with the highest quality data.

The set of plots included in the nightly notebooks was curated over the course of a year as observers developed a better understanding of how system and antenna failures were most likely to manifest. As of the time of writing, there are six total nightly notebooks with over fifty different plots and visualizations between them - not all of these are inspected every day, but they serve as essential tools for the analysis team to quickly identify the best nights of data and to more easily

track the source of any pathologies found during the analysis pipeline. The following subsections will give brief examples and explanations of several of the more frequently used plots that have been included since the inception of the nightly notebooks.

2.2.1 Sky Map

The sky map, shown in Figure 2.8, shows a map of the sky at radio frequencies (Remazeilles et al., 2015b) with an overlay indicating exactly what HERA observed that night. The brightest radio sources, identified from the GLEAM 4Jy Catalog (White et al., 2020) are shown as black dots, and the most observationally significant sources are named. The shaded band indicates the Full Width Half Max of the HERA beam. The full HERA beam extends much wider, but this band shows where the vast majority of our sensitivity lies. This map helps the analysis team quickly identify what parts of the sky were observed, where important sources were at any given time, and provide context for what to expect from that night of observation.

2.2.2 Antenna Map

The antenna map, shown in Figure 2.9 shows a map of all active antennas in the array. The location of the node boxes is indicated with a gold star, and antennas within the same node are colored accordingly. The antennas outlined in black are those that were flagged by the correlation metric described in Section 2.1.1. Antennas with full color are actively recording data, and those with faded colors are electronically connected but not actively recording. For example, we can see that the entirety of node five (green) is connected, but antennas 75 and 76 are not recording data on this night. This figure gives observers a quick understanding of how many antennas are online, what fraction of them are properly correlating, and if there are any major node failures. In the example shown, an observer would quickly note that nodes 5, 7, 14, and 20 are completely failing to correlate, which indicates a failure with that node's white rabbit clock system.

2.2.3 Auto-Correlation Slices

Figure 2.10 shows the auto-correlation spectra - power versus frequency - for a single observation time from the middle night. The North and East polarizations are plotted together for each antenna.

The antennas are organized such that each row includes all antennas within a given node. Within the rows, antennas are sorted by SNAP connection and within that by PAM connection. This sorting allows observers to quickly identify if particular nodes or subcomponents are failing and makes it much easier to spot patterns in systematics. The antenna number labels are colored based on the 'a-priori antenna status', which is a designation given to each antenna by the HERA site crew that indicates how well we expect an antenna to function based on its commissioning status. This plot only shows five nodes of antennas, but in practice we make this plot for all nodes in the array.

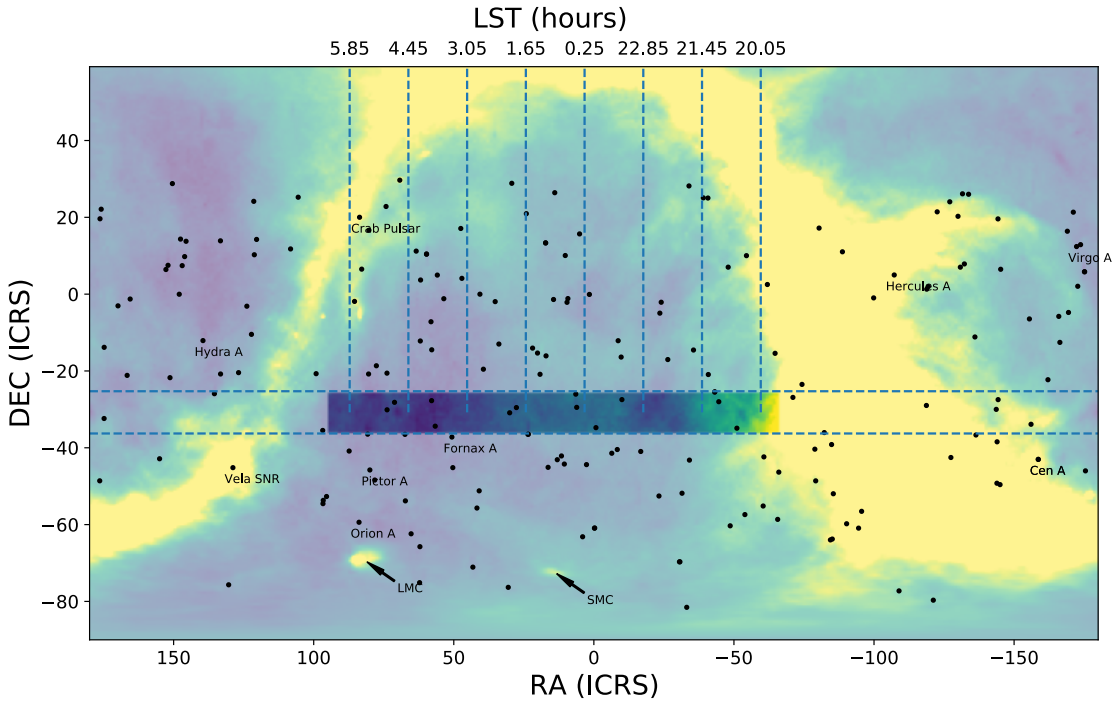


Figure 2.8: Map of the radio sky (Remazeilles et al., 2015a), with the HERA observation band for JD 2459122 shaded, based on a Full Width Half Max of 12 degrees. Individual sources shown are those included in the GLEAM 4Jy catalog White et al. (2020) with a flux greater than 10.9Jy.

2.2.4 Auto-Correlation Waterfalls

In addition to plotting auto-correlation spectra of a single observation time, which show a snapshot at a single time, we also want to visualize the spectra over the whole night. For this, we use a 'waterfall' plot, which plots time versus frequency and uses color to represent power. An example including a subset of five nodes is shown in Figure 2.11. These are called waterfall plots because time starts at the top of the plot and progresses to the bottom. These plots are organized the same as the slices, but allow us to see additional features that may evolve in time, or failures that might

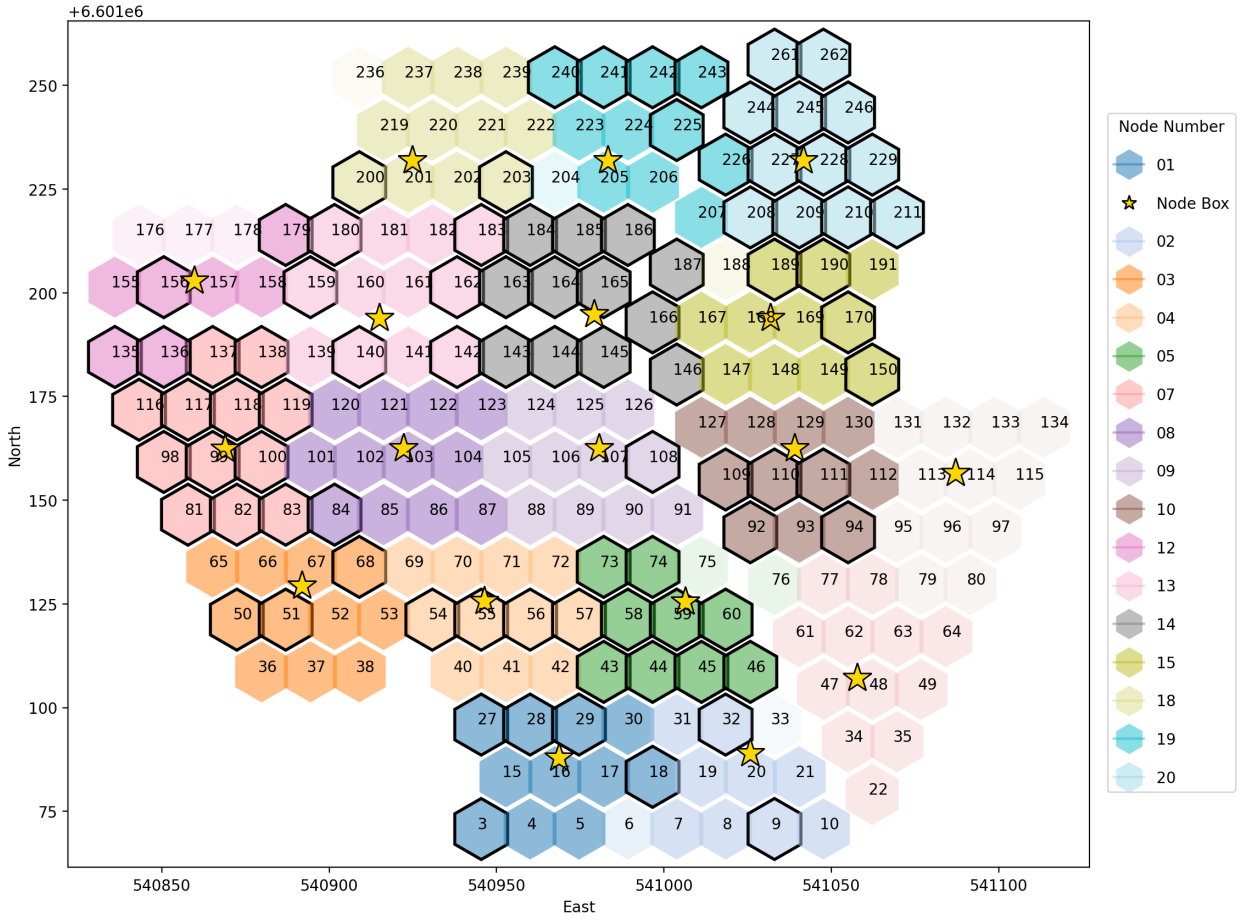


Figure 2.9: A map of all connected antennas, colored by node connection, and the location of node boxes. Antennas outlined in black have already been flagged by the cross-correlation metrics.

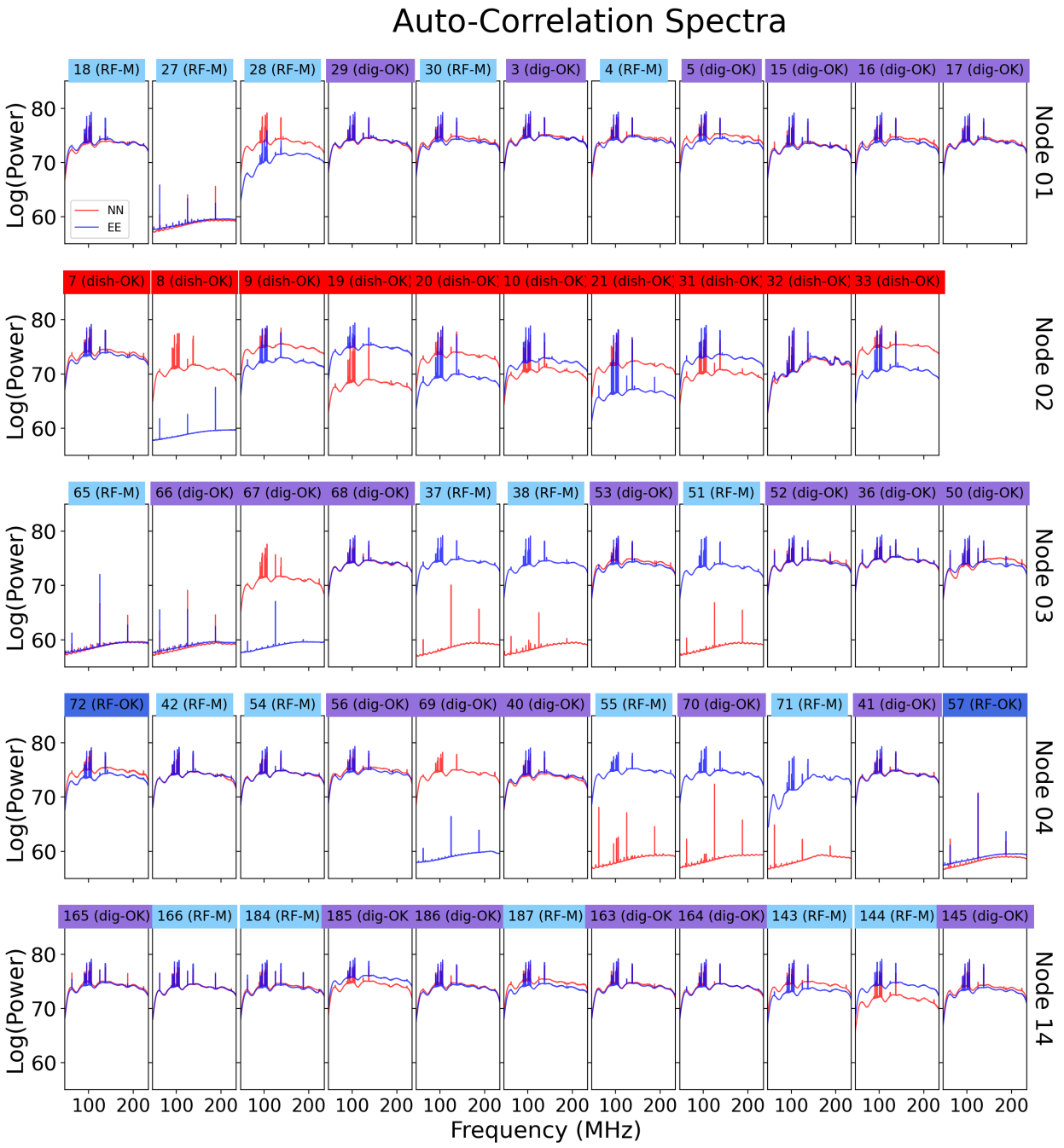


Figure 2.10: Single observation auto-correlation spectra slices, organized by electronic connection and labelled based on commissioning status.

have started part way through the night. For waterfalls we plot each polarization individually - here we are choosing to only show the North polarization. The antennas with dark blue waterfalls are those that were observing load rather than sky, which has much lower power and therefore falls off the colorbar.

2.2.5 Auto-Correlation Mean-Subtracted Waterfalls

Lastly, we have a plot designed to highlight any time instability in the data. Since the sky rotates relatively slowly and we expect our instrument to be stable in time, one would expect the data to evolve smoothly along the time axis. However, we have found that HERA data can show significant time instability and even discontinuities - henceforth, this will be referred to as 'temporal structure'. In order to highlight these features in the data, we create waterfall plots where the mean spectrum (time-averaged visibilities per frequency) is subtracted from the original waterfall. These are called mean-subtracted waterfalls, and are shown in Figure 2.12. There are several features here worth noting. First, one notices that while the majority of antennas evolve relatively smoothly in time, there are several that stand out as having spurious temporal structure. As all antennas are observing the same sky and *should* have the same instrumental response, we consider any temporal structure that is common across all (or most) antennas to be due to sky evolution, and any structure that is unique to a small number of antennas to be a systematic. Second, we see that RFI (which is known to fluctuate rapidly in time) manifests as highly variable vertical lines, with the FM band being particularly evident. As RFI remains consistently within the same frequency bands throughout the entire night we can easily separate this feature from broader temporal structure that spans a larger portion of the bandpass.

Antenna 50 (in node three) is an extreme case, but cases like antennas 45 and 32 show more subtle effects that we would consider spurious structure. There are ten antennas shown that appear to be mostly solid teal and have no time evolution at all - these are antennas that were observing on load rather than sky, and therefore were observing a stable noise source rather than a changing sky. While it is not visible on the color scale chosen here for sky data, these antennas do exhibit spurious temporal structure that must be attributed to instrumental effects. The development of these mean-subtracted waterfalls was our first window into HERA's issues with time stability, which

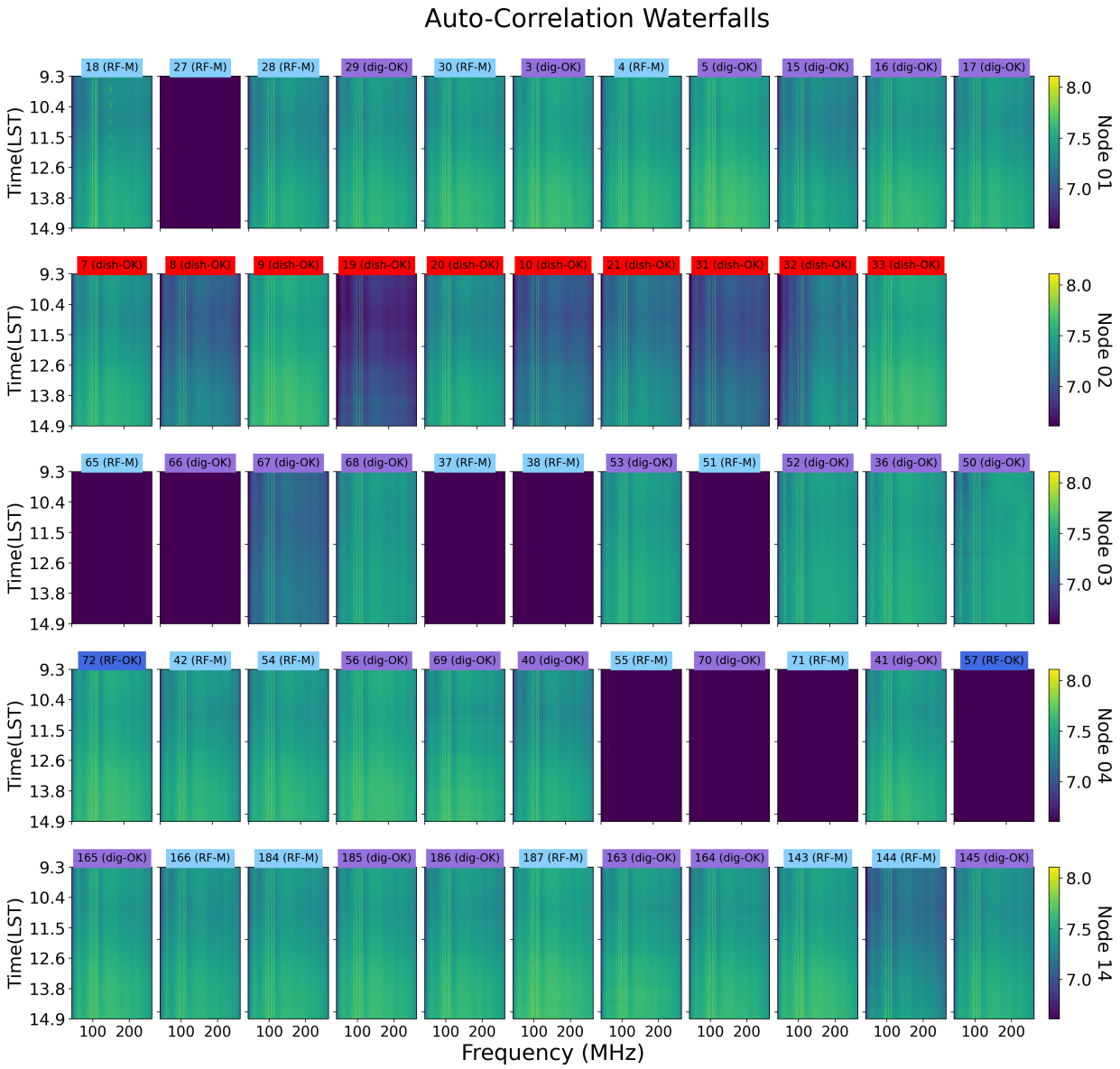


Figure 2.11: Waterfall plots showing the time evolution of auto-correlations for a subset of antennas in the array, organized by electronic connection and labelled based on commissioning status. Data shown is the North polarization.

is a major assumption held throughout the analysis pipeline and therefore must be addressed. We will continue to explore HERA's temporal structure in both sky and load data, along with other related systematics, in Chapters 3 and 4.

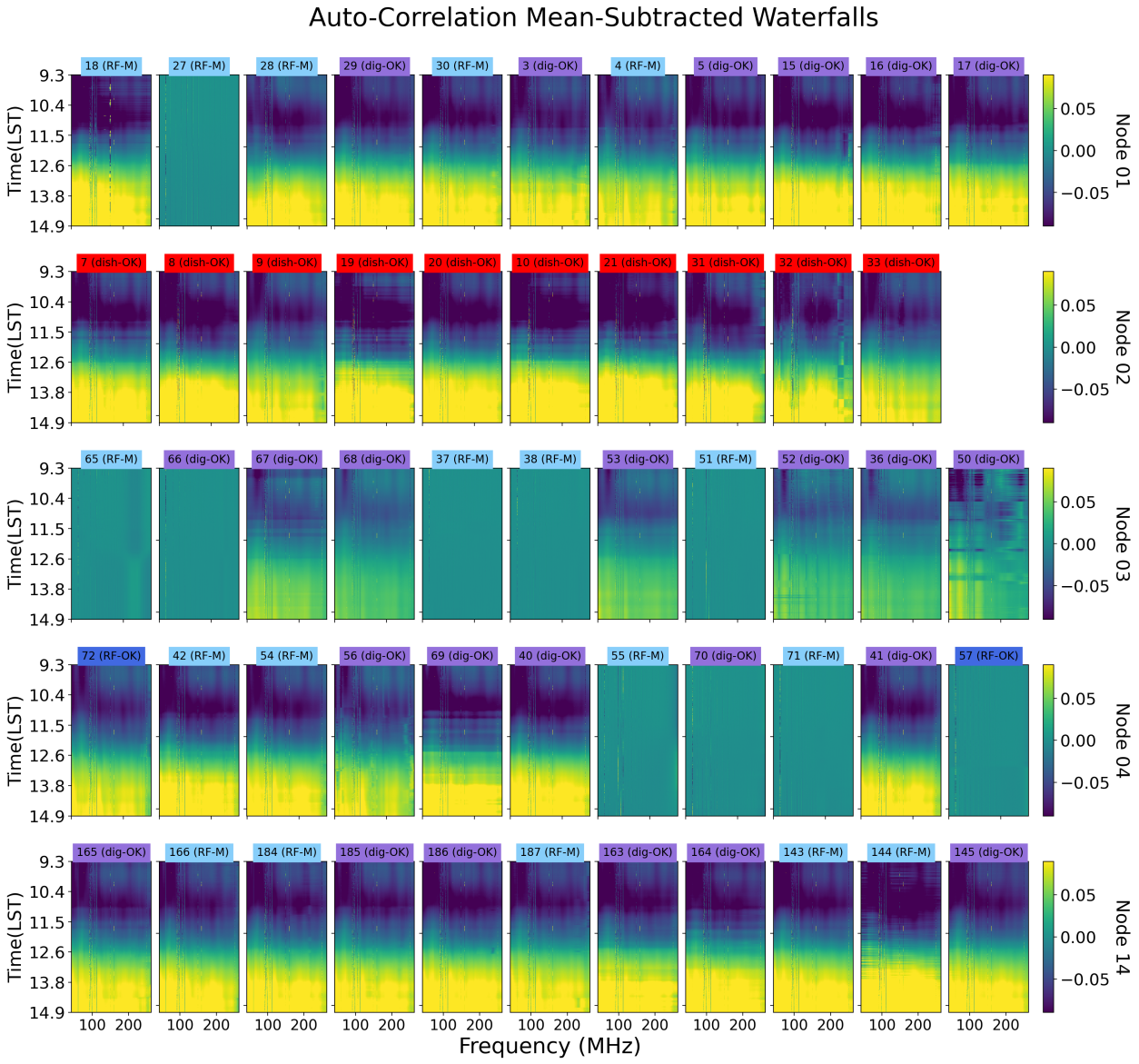


Figure 2.12: Waterfall plots showing the time evolution of auto-correlations for a subset of antennas in the array, organized by electronic connection and labelled based on commissioning status. Here the average spectrum across the time range shown has been subtracted from all data - we call this a mean-subtracted waterfall. The colorbar units are set to represent percent deviation from the mean.

Chapter 3

HERA COMMISSIONING AND DATA QUALITY

In this chapter we will overview some additional checks we impose on our data during the commissioning process and outline several examples of systematics that were uncovered during this process. Section 3.1 will start by reviewing how we validate HERA's astrometry, or its ability to accurately measure the position and brightness of sources on the sky. Then, Section 3.2 will provide several examples of systematics that were discovered during commissioning, many of which using the tools outlined in the previous chapter. We provide these examples to help the reader understand the character of our data and to provide context for analysis choices we will make later on.

3.1 Astrometry Validation Through Deconvolution

In order to successfully create power spectrum measurements using an imaging approach, which is the final goal of this thesis, we must confirm that HERA as an instrument is able to accurately measure sources on the sky. This is particularly important when performing sky-based calibration, as we depend heavily on sources in our observed images matching those in our sky model. In order to validate HERA's astrometry, we will use a process known as deconvolution. Here, deconvolution refers to an iterative process where one finds the brightest source in an image, subtracts it and creates a new image, and continues on until no new sources can be identified. The removed sources can then be used to create a catalog that includes the flux and position of all detected sources. In our case deconvolution is implemented using the software package `Fast Holographic Deconvolution FHD` - we won't go into detail on the implementation of deconvolution in this thesis, but the interested reader can find more discussion in Sullivan et al. (2012). Deconvolution can have many purposes, but the one that is useful to us is using the resulting source catalog to validate HERA's astrometry. We use the GaLactic and Extragalactic All-sky MWA survey (GLEAM) as a comparison catalog, and use a basic source matching algorithm to identify matches between GLEAM and HERA's deconvolution catalog. We hope to validate that HERA does not have bias in its measurements of

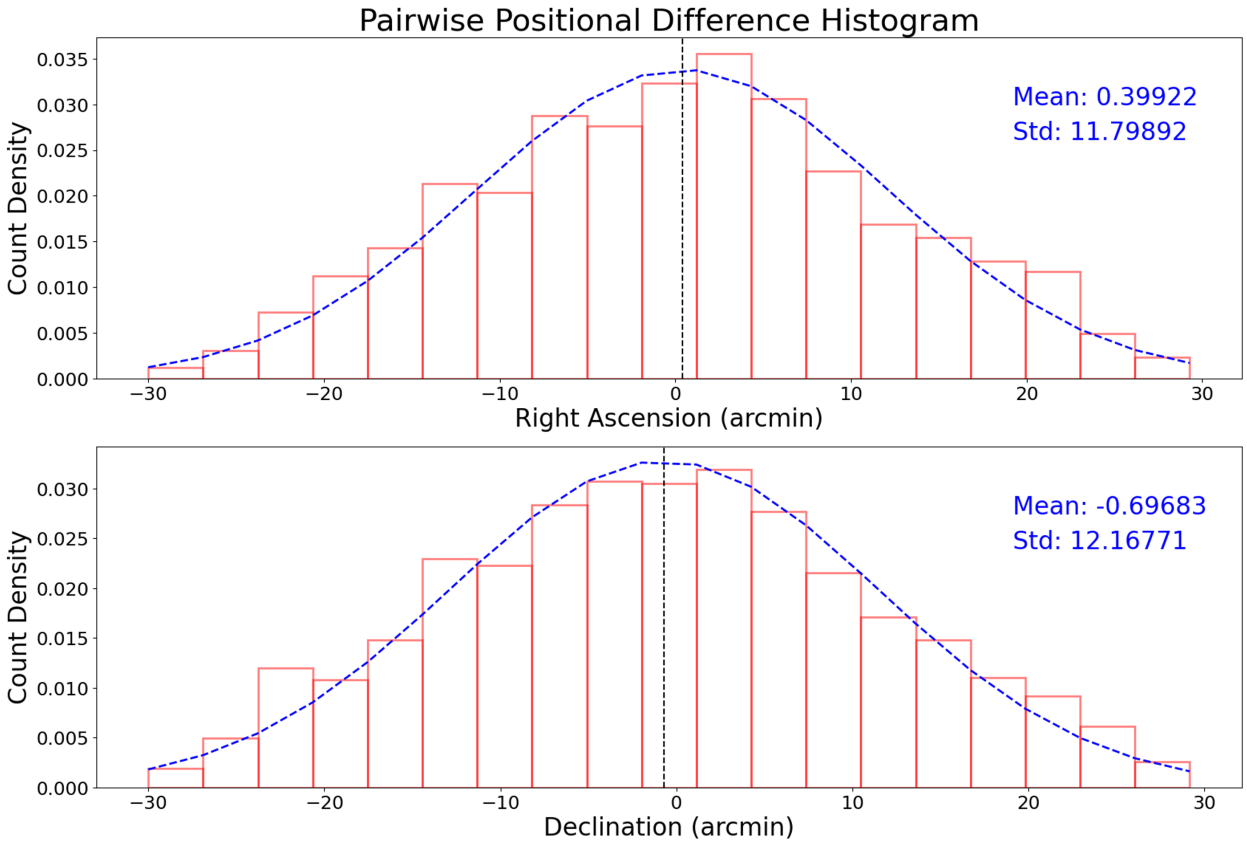


Figure 3.1: Histograms of the RA (top) and DEC (bottom) positional difference between sources in the GLEAM catalog and in HERA's deconvolution catalog. The vertical dashed line represents the distribution mean, and the blue dashed line shows a Gaussian fit to the distribution. The mean and standard deviation of the distribution are also printed.

source positions or brightness.

For every source HERA detects that we can confidently pair with a GLEAM source we calculate the difference in Right Ascension (RA), Declination (DEC), and flux between the two catalogs. Figure 3.1 shows a histogram of the positional difference in RA (top panel) and DEC (bottom panel), with a vertical dashed line representing the distribution mean and a blue dashed curve showing a Gaussian fit to the histogram. In both cases we see that the mean is very close to zero and the distributions have minimal skew, meaning that there is minimal bias in our positional

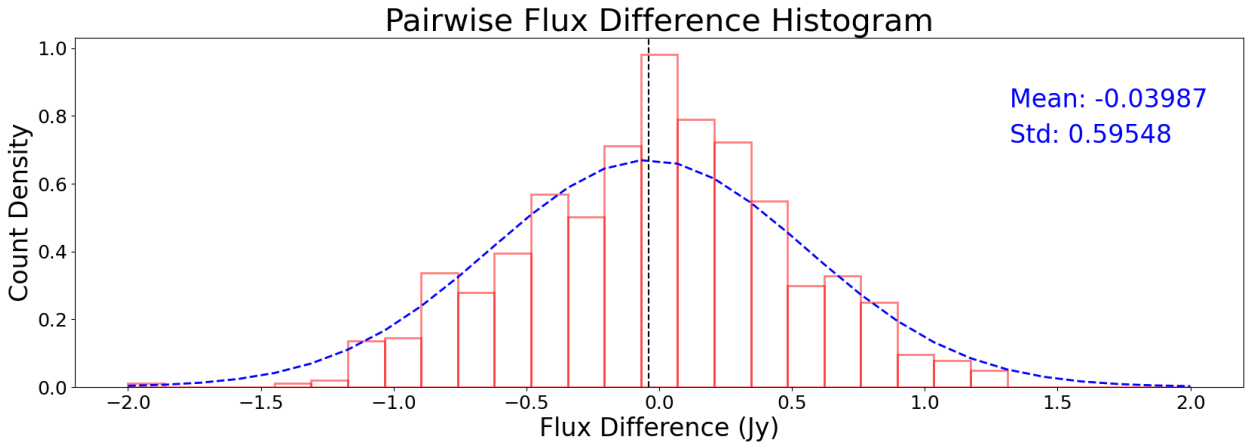


Figure 3.2: Histogram of the flux difference between sources in the GLEAM catalog and in HERA's deconvolution catalog, measured in Janskys. The vertical dashed line represents the distribution mean, and the blue dashed line shows a Gaussian fit to the distribution. The mean and standard deviation of the distribution are also printed.

measurements. The width of the histogram is somewhat larger than we would expect for HERA's full configuration, but given the imaging constraints with the limited number of baselines available for this deconvolution run we do not expect to see high precision matches. Rather, our focus is to identify whether there is bias in our source measurements.

Similarly, Figure 3.2 shows a histogram of the flux difference between matched sources in the GLEAM and `FHD` deconvolution catalogs. Again we see that the mean is very close to zero, the distribution has no notable skew, and there are no concerning outliers. This validation of our astrometry builds confidence that there are no major errors in HERA's measurements or our imaging pipeline as it pertains to source location and flux.

3.2 HERA Systematics

Throughout the commissioning of HERA the automated and manual data inspection systems outlined in the previous chapter have uncovered many features and systematics in HERA data that must be addressed before proceeding with analysis and pursuing a science result. When these systematics are uncovered our first goal is to discover the root cause and resolve the issue through

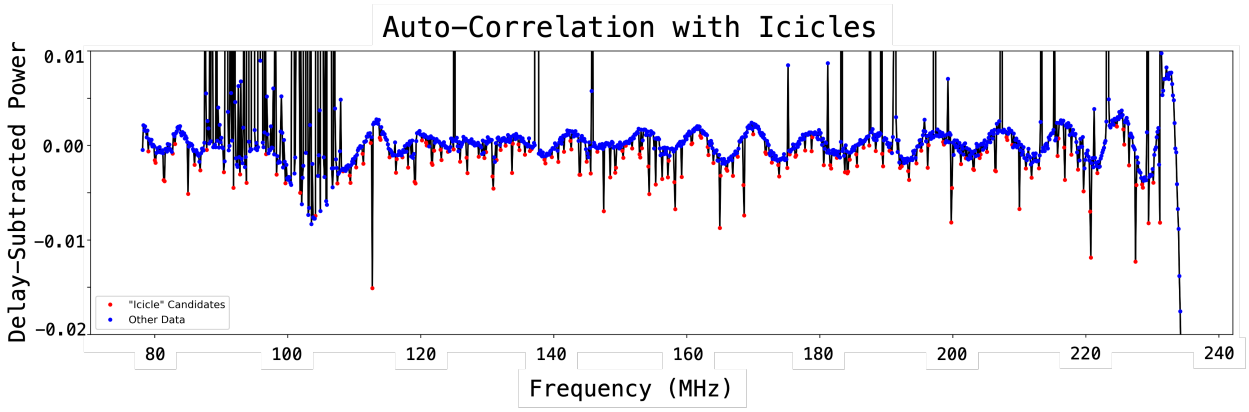


Figure 3.3: A sample auto-correlation spectrum showing the icicles systematic. Data points in blue represent the bandpass behaving as expected, and points in red indicate icicles. To highlight the icicle feature a clean delay-filtered spectrum has been subtracted from the raw auto-correlation.

either hardware or software repairs. If this route is unsuccessful, we may proceed either by choosing to omit the affected data from analysis or by developing techniques to mitigate the systematic through the analysis pipeline. The following subsections will each outline one of the more interesting systematics uncovered throughout the commissioning process and describe efforts to resolve them. Several of these systematics remain unresolved, and therefore will be handled via analysis and data selection techniques.

3.2.1 Negative Power Spikes (Icicles)

During the early commissioning phase of the current HERA system observers noticed negative 'spikes' in the auto-correlation spectra of all antennas, shown in Figure 3.3. The standard bandpass shape is shown in blue, and each negative spike is indicated by a black line with a red dot. As these spikes appeared to drip down from the primary bandpass they received the name 'icicles'. There is no mechanism for the sky to sharply lose power at some frequencies, and therefore it was clear that this is a systematic. Losing power in this way introduces a lot of frequency structure to the data and has the potential to cause artificial signal loss when we make power spectrum measurements, which is cause for great concern. By tracking how this systematic had evolved with

time the commissioning team was able to line up changes in the icicles with changes to the firmware in the SNAP. Through this process, we identified that certain changes to the handling of the Fast Fourier Transform (FFT) in the SNAP were causing this signal suppression. More specifically, there was an error in how data was being binned into frequency channels. The commissioning team made changes to the FPGA and FFT code accordingly and the icicles systematic has since been resolved. There are no icicles present in the data used for analysis in this thesis.

3.2.2 High Delay Ripple

The next systematic we will discuss is a high delay ripple we see on a subset of antennas in the array, usually at higher frequencies. Figure 3.4 shows a comparison of a normal auto-correlation spectrum and one with the high delay ripple. The top left panel shows a normal spectrum, with a blown up section shown in red, showing that the spectrum is in fact quite smooth. The top right figure shows the delay spectrum for this antenna, which is the result of doing a Fourier transform along the frequency axis of the original auto-correlation. In this case the delay spectrum looks as expected, with most of the power at small delays and falling off to noise at higher delays. The lower left panel shows a spectrum with the high delay ripple - in the full spectrum this just looks like a thickening of the line, but the blown up panel shows that it is actually a rapid ripple in the spectrum. The delay spectrum here shows strong spikes at a delay of 2700 nanoseconds, indicated by two dashed red lines. While the example shown is using auto-correlation data we also see the systematic in cross-correlation data, particularly between antennas that are within the same node. As already mentioned, our analysis pipeline relies heavily on our instrument having a smooth frequency response, and therefore this systematic has the potential to be highly disruptive.

Given the nature of this systematic we initially suspect that the cause could be a common-mode injection, meaning that the same spurious signal is injected on multiple inputs to the signal chain and eventually is correlated with itself. The commissioning team has run many tests at the HERA site aiming to identify the source of this systematic. In particular, we applied chokes to many places along the signal path in hopes of mitigating the systematic and thereby identifying the source, but all choke tests have shown no effect on the ripple. Unfortunately, this systematic remains present in the data, and is not feasible to avoid through analysis with an imaging approach (as is used

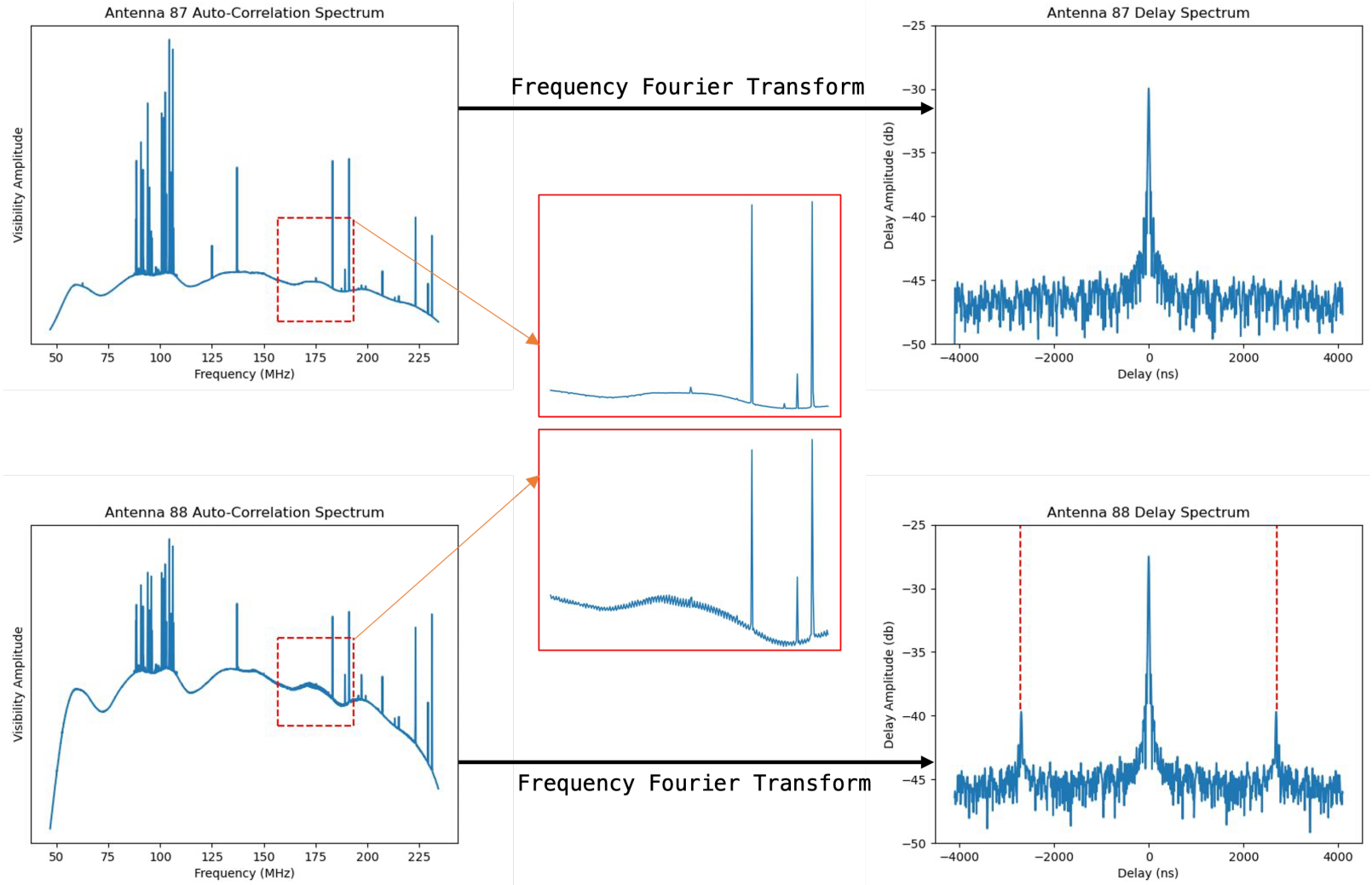


Figure 3.4: An example of one clean antenna (top row) and one with a high delay ripple (bottom row). The left column shows auto-correlations of sky data for each antenna. A subset of these spectra are shown enlarged within the red boxes. The right hand column shows the delay spectrum for each antenna, which is constructed by taking the frequency Fourier transform of the auto-correlation spectrum.

here), and therefore we mitigate the effects of this ripple by removing any affected antennas from our analysis data set. While this is effective at mitigating the systematic, it has the potential to significantly reduce our data volume and therefore have a negative impact on our power spectrum measurements.

3.2.3 Frequency Combs

When observing sky data we expect to see spikes in frequency due to RFI, but when observing load data we would anticipate a spectrum with no frequency structure beyond the expected noise. However, this is not the case. Figure 3.5 shows auto-correlation spectra for three sample antennas observing FEM load data. The first feature we note are the three brightest lines, all over-plotted in black. These lines are actually due to emission from the system's clock, and while somewhat irksome they are very predictable and therefore easy to consistently exclude from analysis. More concerning to our analysis are the set of lower amplitude spikes seen especially clearly at lower frequencies. Of particular concern is the observation that these spikes appear to be periodically

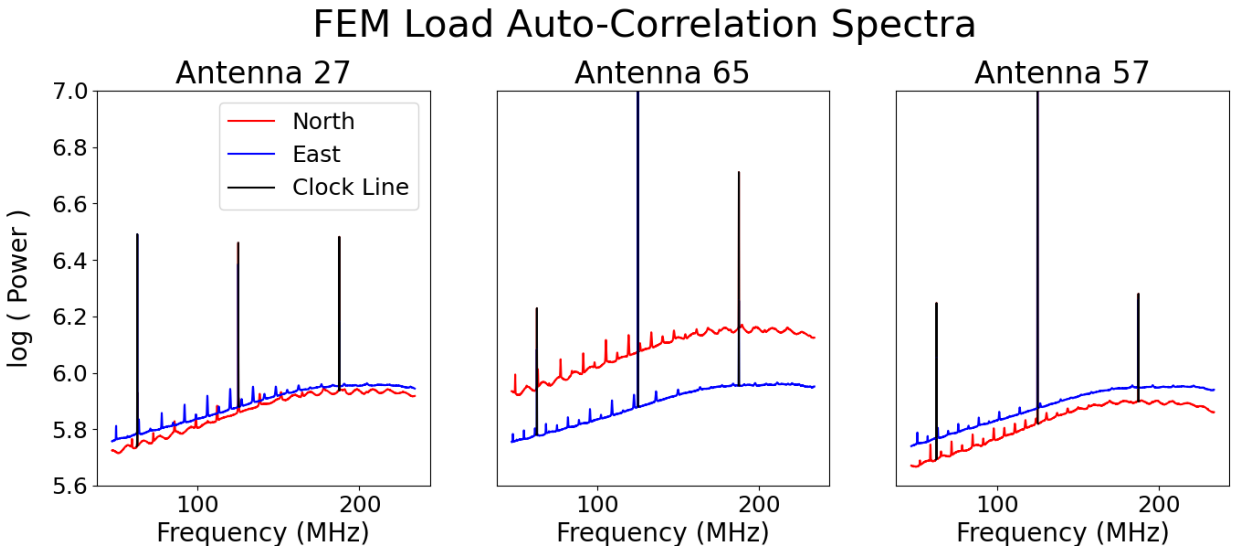


Figure 3.5: Three examples of auto-correlation spectra from FEM load data. The three black spikes in each spectrum are emission from HERA's internal clock, and are easily handled in the analysis pipeline. The lower level, seemingly periodic spikes are referred to as HERA's 'frequency comb'.

spaced in frequency, hence the name 'frequency comb'. In an attempt to diagnose the source of these combs, we implemented a peak detection algorithm to identify spikes and calculate their spacing. We found that spikes within each antenna-polarization are periodic, but the period of the combs does not match between different spectra. Even further, we find that the periodicity of the spikes on any given antenna-polarization slowly changes over the course of a night. This leads us to believe that the period may be affected by factors like temperature, but makes it much more difficult to pinpoint a component that could produce this structure. Given that HERA currently has other, higher amplitude systematics the commissioning team has prioritized other issues, and therefore this systematic has not yet been resolved.

3.2.4 Frequency 'Worms'

We define frequency 'worms' as any structure that is isolated to a narrow range of frequencies and exhibits spurious temporal and/or spectral structure. This feature originally earned its name for its tendency to wander back and forth across waterfall plots, thus making it 'worm-like', but the term's meaning has evolved to more often refer to features that remain centered over a single frequency bin. Figure 3.6 shows example worms from four different antennas all observing sky data. In this figure the left hand panel is a mean-subtracted waterfall plot with a white line laid over top indicating the approximate center frequency of the worm feature. The right hand panels are a single slice at this frequency bin, showing its evolution in time. The top left panel shows a 'clean' antenna, while the other three examples were selected to highlight the variability of the worm feature. For example, in the top right of Figure 3.6 we see the spectrum rapidly switching between two fairly binary states centered over a single frequency. From the right hand panel we can see that these transitions can happen as quickly as one integration to the next. However, in the bottom left of Figure 3.6 we see a very different character. In this case, we see a feature that is still centered over one frequency, but rather than alternating between states it seems to be intermittently spreading and thinning as the power level fluctuates. In the bottom right we see yet another character, where the feature manifests at some times as a sharp but somewhat persistent power change, and at other times the power is fluctuating so rapidly that it resembles noise.

Regardless of the exact character of a worm they introduce substantial spectral structure from

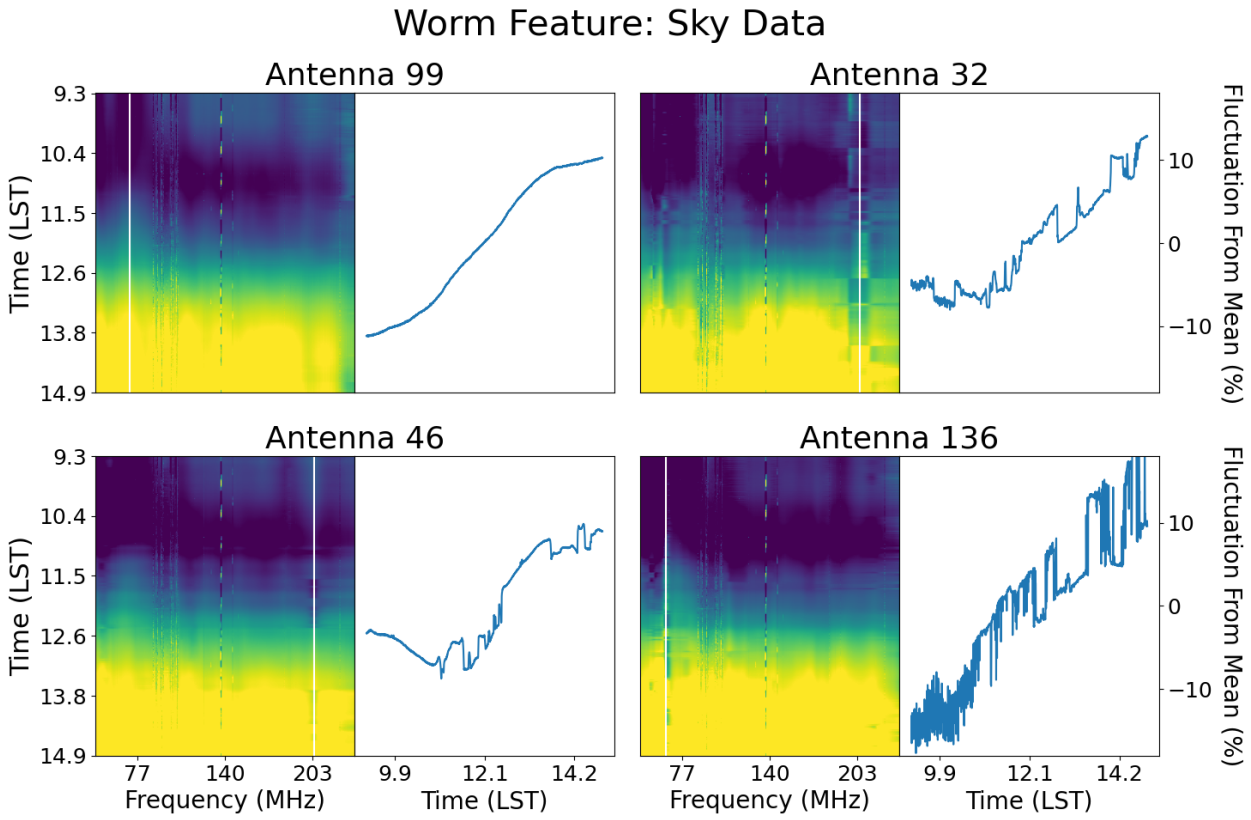


Figure 3.6: Data from four different antennas all observing sky data. The left hand panels are mean-subtracted waterfalls and the right hand panels are slices of those waterfalls at a frequency indicated by the white line that show the evolution of that frequency bin in time. The top left panel is an example of a clean antenna with no worm feature, while the other three show various characters of worms.

the instrument. In diagnosing this systematic the first consideration was whether worms were present in FEM load and digital noise data or only in sky data - knowing this helps us narrow down which part of the system might be introducing the effect. Figure 3.7 shows the same mean-subtracted auto-correlation waterfalls and slices but for FEM load data rather than sky data. Again, three antennas have been selected that show particularly interesting worms, and a clean antenna is also shown for reference. It is important to note that only ten to twenty percent of antennas exhibit worms on any given night. We clearly see that even on FEM load data there are

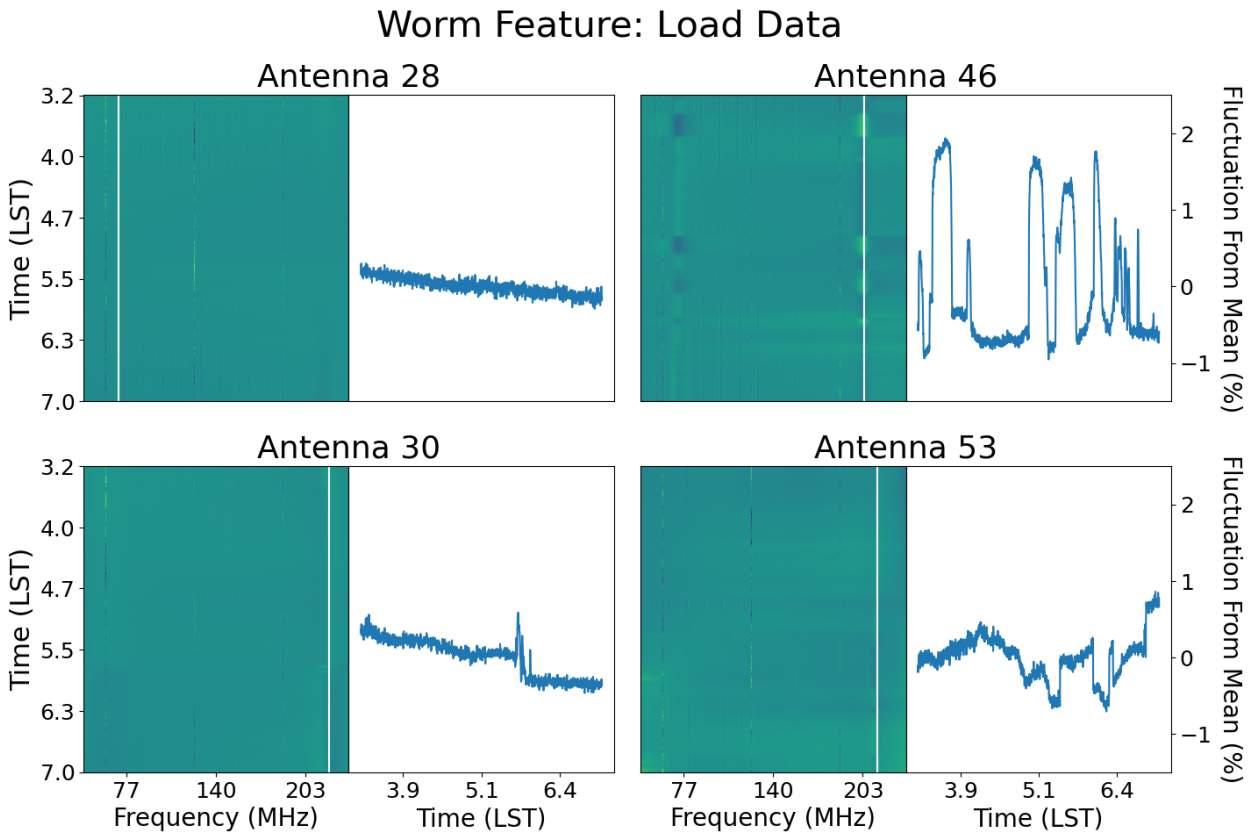


Figure 3.7: Data from four different antennas all observing load data. The left hand panels are mean-subtracted waterfalls and the right hand panels are slices of those waterfalls at a frequency indicated by the white line that show the evolution of that frequency bin in time. The top left panel is an example of a clean antenna with no worm feature, while the other three show various characters of worms.

substantial worms in the data, although it is an open question whether their character remains the same. The logical next question is to look for worms in digital noise data. Figure 3.8 shows the mean-subtracted waterfall and corresponding slice for digital noise data - note that only one antenna is shown because digital noise data is very consistent between inputs and different antennas are indistinguishable in this case. Here it is clear that there is no spurious structure in the digital noise data. At first glance this means that the worm structure must be introduced somewhere between the FEM load source and the digital noise source. Referring to Figure 1.6, this makes the PAM

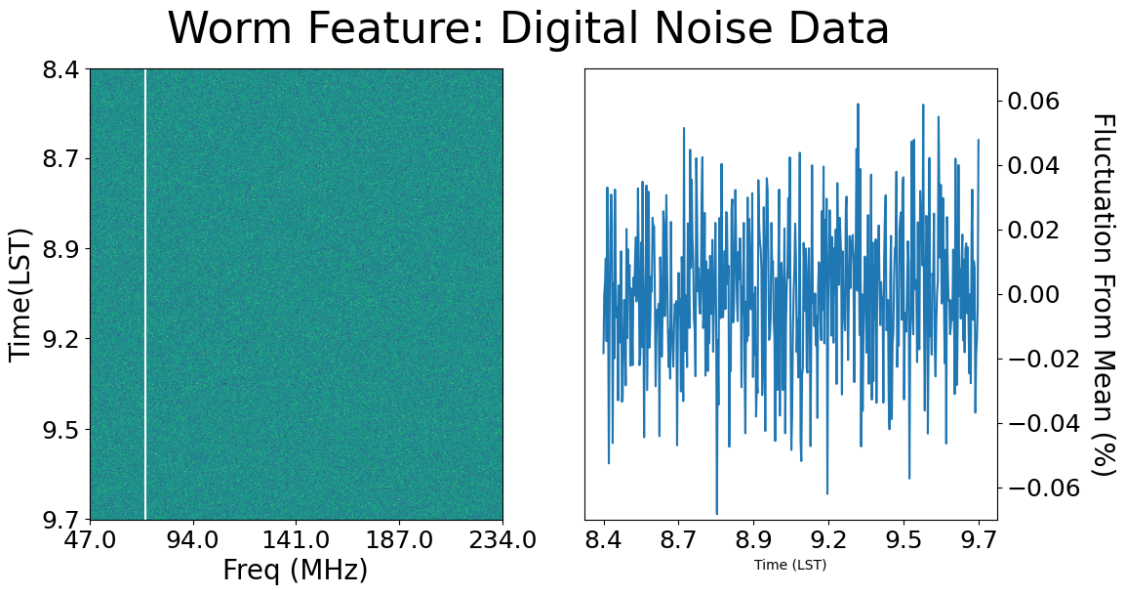


Figure 3.8: Both panels are constructed from digital noise data. The left panel is a mean-subtracted waterfall, and the right panel is a slice at the frequency indicated by a white line in the waterfall. Notably, this data shows no sign of spurious temporal or spectral structure.

and ADC primary suspects. However, we have previously seen that exceptionally bright spikes in the data, usually due to RFI, can cause problems in the digital system due to computational overflows. While this could not explain the features we see in the FEM load data, it is possible that some portion of the features seen in sky data could be introduced in the correlator. Our current understanding is that the worm systematic likely has multiple causes, hence the many different 'characters', and the digital system is still considered a possible culprit for some fraction of the worms. A more detailed investigation of the PAM and its potential contributions will be discussed in Chapter 4.

3.2.5 Temporal Structure

Temporal structure is defined as any case where data is changing more rapidly in time than can be explained by changes on the sky. As with the worms, we see that temporal structure is a broad category and can take many forms, so we must consider the likely outcome that it also has

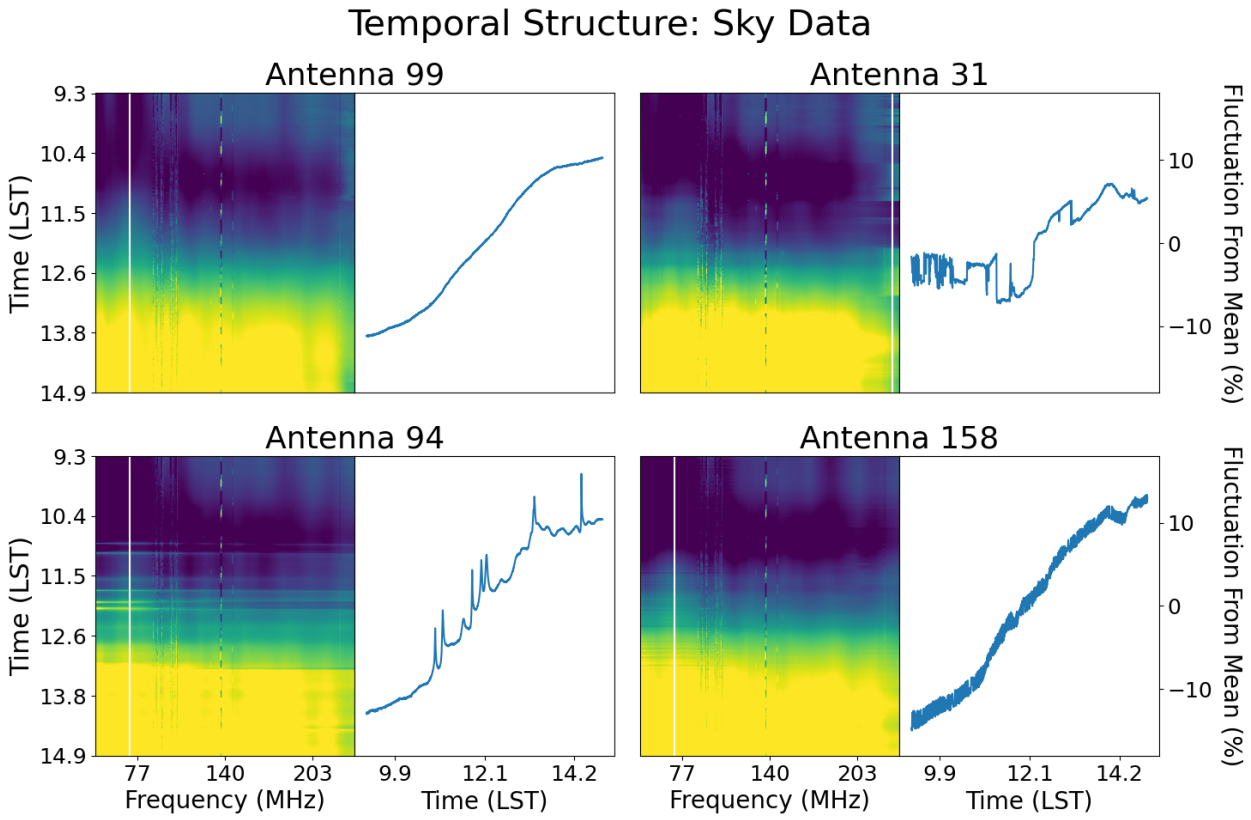


Figure 3.9: Data from four different antennas all observing sky data. The left hand panels are mean-subtracted waterfalls and the right hand panels are slices of those waterfalls at a frequency indicated by the white line that show the evolution of that frequency bin in time.

multiple root causes. Figure 3.9 shows auto-correlation mean-subtracted waterfalls and slices of sky data for three sample antennas that exhibit temporal structure and one clean antenna that does not, pictured in the top left panel. All four panels show data from the same polarization taken over the same time range, and therefore any deviation from the overall shape exhibited by antenna 99 is representative of spurious temporal structure. Antennas 94 and 31 both show cases of big spikes or dips in power, although the former spikes and recovers quickly while the latter resembles something more like state switching. In contrast, antenna 158 has more subtle structure that manifests somewhat like an excess in noise. Notably, we see very little evidence of either worms or temporal structure being correlated across polarizations or any other electronic subsystem, and

therefore we want to identify failure modes that could affect each signal chain independently. An initial hypothesis proposed to explain this structure was movement of the antenna feeds, possibly due to high winds or smaller vibrations, which would cause the antenna beam response to change and could happen very rapidly in time. We can easily test this hypothesis by looking for temporal structure in the FEM load, which is shown in Figure 3.10. We clearly see that there is temporal structure in the load data and that it exhibits a very similar character to the structure we observed in sky data. While it is still possible that beam changes due to feed movement are a contributing factor, those changes would not cause jumps in load data and therefore cannot explain the temporal structure we see in the load data. The fraction of antennas that exhibit temporal structure can change substantially between nights, but we found that the average fraction of susceptible antennas was similar between nights observing sky and nights observing load. This leads us to the conclusion that the majority, if not all, of the temporal structure must be coming from farther down the signal chain than the feed. As with the worm structure, we find that there is no detectable temporal structure in the digital noise data.

Due to our expectation of a spectrally smooth instrument response, temporal structure that has an associated spectral shape change (such as that seen in antenna 90 of Figure 3.10) particularly dangerous to our analysis. However, it is worthwhile to discuss why all forms of temporal structure are cause for concern. First and foremost, we must consider that whatever physical mechanism is producing this systematic could be having other, possibly more subtle, effects on the data that we are yet to observe. Depending on the night we might observe temporal structure on anywhere from ten to fifty percent of antennas - any component affecting this much of our system is of serious concern. Second, we see evidence that the temporal structure is not 'gain-like', meaning it does not behave as a simple per-antenna amplitude change that can be corrected through calibration. Rather, it behaves 'signal-like', showing up in the cross-correlations in a way that is very difficult to disentangle from the true signal. Lastly, our power spectrum calculations rely on the principle that the visibilities evolve sufficiently slowly that the difference between time-adjacent visibilities is an accurate representation of instrumental noise. If our instrument is introducing signal-like features that change on short enough timescales then this assumption would no longer hold true. This systematic is both severe and complex, and therefore will be discussed at length in the next chapter.

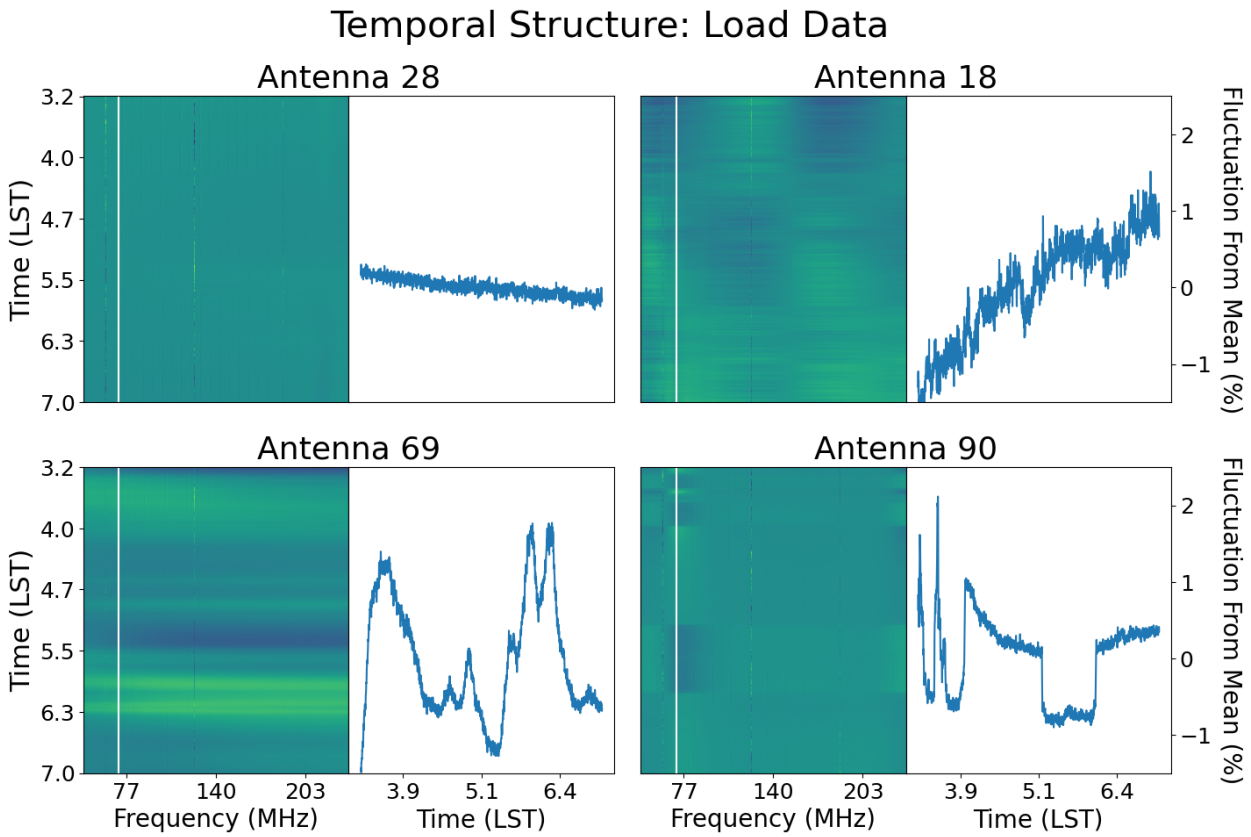


Figure 3.10: Data from four different antennas all observing FEM load data. The left hand panels are mean-subtracted waterfalls and the right hand panels are slices of those waterfalls at a frequency indicated by the white line that show the evolution of that frequency bin in time.

Chapter 4

EXCESS CORRELATION

In Section 2.1.1 we introduced the correlation metric as a tool for identifying particular antennas that are not properly correlating with the rest of the array. In this chapter, we will explore how the matrix visualization of this correlation metric allowed us to identify a much more nuanced systematic, which we will call 'excess correlation'. To begin, Section 4.1 will provide more context for the statistics we expect from the correlation metric and therefore what we expect a healthy correlation matrix visualization to look like. Next, Section 4.2 will dive deeper into pathologies we see in the correlation matrix and how we approached debugging this systematic. Finally, Section 4.3 will outline our attempts to fix the excess correlation systematic and describe the current state of affairs as it pertains to the data set used for analysis in this thesis.

4.1 Understanding the Correlation Metric

Figure 2.3 showed that there is a significant amount of structure in the correlation matrices, specifically related to node connections. Baselines within a node appear to have larger values of C_{ij} than baselines between nodes. We have previously noticed instances of severe node-based structure when there are timing mismatches between nodes due to a failure of the clock distribution system. Figure 4.1 is an example from an observation when the timing system was known to be broken, and we see clearly that timing mismatches depress the correlation metric between nodes. This causes much clearer node structure than the more common structure seen in Figure 2.3. Therefore, one wonders: are the larger C_{ij} values on the intra-node baselines due to some milder form of this clock distribution issue—perhaps a small error in timing—or is this structure otherwise explicable or even expected?

Put another way, with our definition of C_{ij} written as

$$C_{ij} \equiv \left\langle \frac{V_{ij}^{\text{even}} V_{ij}^{\text{odd}*}}{\left|V_{ij}^{\text{even}}\right| \left|V_{ij}^{\text{odd}}\right|} \right\rangle_{t,\nu}, \tag{4.1}$$

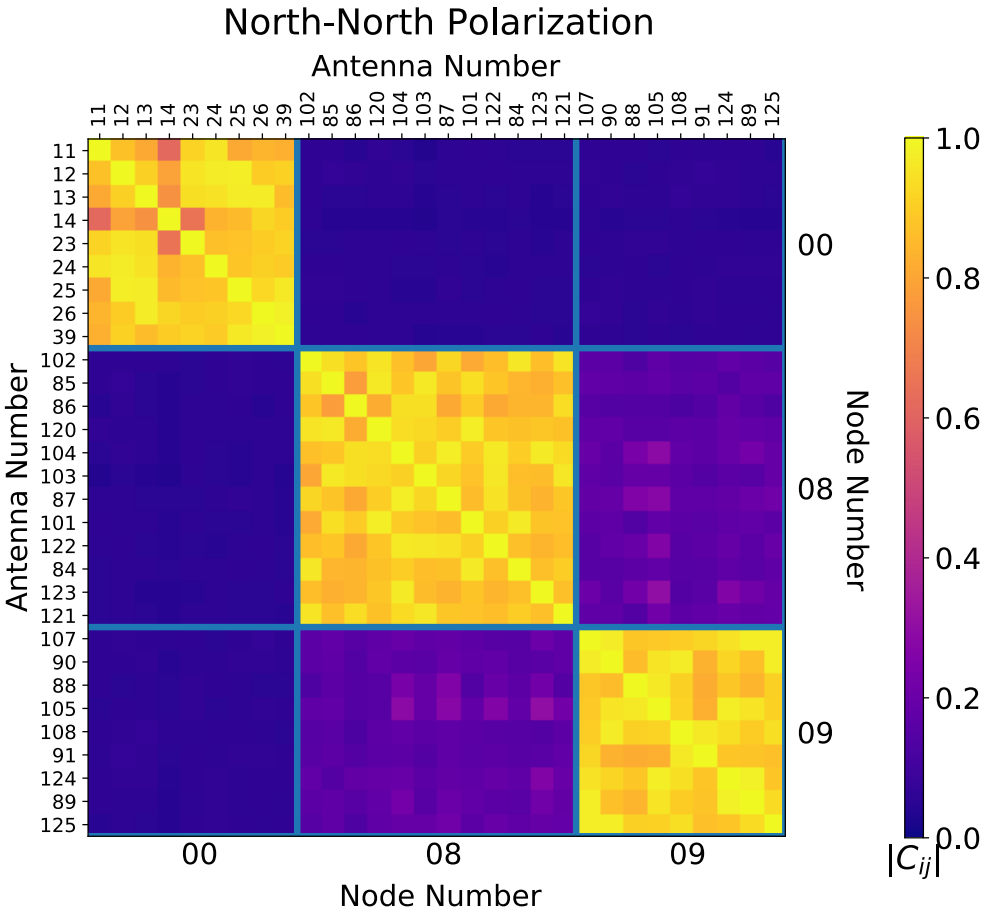


Figure 4.1: A correlation matrix for a single polarization of HERA data from October 21, 2019 (JD 2458778), taken at a time when the timing system was malfunctioning and antennas between different nodes were not correlating, showing a clear block-diagonal along node lines. This is a sample case where the auto-correlations are nominally acceptable, and investigation of the cross-correlations is necessary to see this type of failure mode.

what is the expectation value of C_{ij}? We can make the assumption for time-adjacent visibility sets V_{ij}^{even} and V_{ij}^{odd} that $\langle V_{ij}^{\text{even}} \rangle = \langle V_{ij}^{\text{odd}} \rangle \equiv V_{ij}^{\text{true}}$ and that the two only differ by their noise, n_{ij}, with mean 0 and variance σ_{ij}^2. Ignoring time and frequency dependence, then we can use Equation (4.1) to first order (ignoring correlations between the numerator and denominator) to find that

$$\langle C_{ij} \rangle = \left\langle \frac{(V_{ij}^{\text{true}} + n_{ij}^{\text{even}})(V_{ij}^{\text{true}} + n_{ij}^{\text{odd}})^*}{\left|V_{ij}^{\text{true}} + n_{ij}^{\text{even}}\right| \left|V_{ij}^{\text{true}} + n_{ij}^{\text{odd}}\right|} \right\rangle \approx \frac{\left|V_{ij}^{\text{true}}\right|^2}{\left|V_{ij}^{\text{true}}\right|^2 + \sigma_{ij}^2}. \tag{4.2}$$

This approximate expectation value shows us the importance of the signal-to-noise ratio (SNR). At high SNR, $\langle C_{ij} \rangle$ goes to 1, assuming the two even and odd signal terms are actually the same—i.e. that the array is correlating. At low SNR, $\langle C_{ij} \rangle$ goes to 0.

It follows then that the apparent node-based structure in C_{ij} might actually be the impact of the the relationship between SNR and baseline length. Inspecting the array configuration (see Figure 2.1) we see that baselines within the same node tend to be shorter than baselines involving two nodes. Shorter baselines are dominated by diffuse galactic synchrotron emission, which means that they tend to have a higher signal than longer baselines. Since all baselines have similar noise levels and since higher SNR leads to larger values of C_{ij}, this could account for the effect.

In order to test whether our node structure is explicable as a baseline length effect rather than some other systematic, we can implement a simple simulation with thermal noise. We calculate V_{ij}^{true} from our data as $(V_{ij}^{\text{even}} + V_{ij}^{\text{odd}})/2$, and take this as a reasonable stand-in for the sky signal, in lieu of a more sophisticated simulation, since it should have approximately correct relative power and should largely average out the instrumental noise. To each visibility V_{ij}^{true} we then add independent Gaussian-distributed thermal noise, with variance given by

$$\sigma_{ij}^2 = \frac{|V_{ii} V_{jj}|}{\Delta t \Delta \nu}, \tag{4.3}$$

where Δt is the integration time and $\Delta \nu$ is the channel width, and we have utilized the amplitude of the constituent auto-correlations to estimate the noise. This noise is uncorrelated between baselines, times, and frequencies. We then calculate C_{ij}. We compare the C_{ij} with simulated noise to the observed C_{ij} in Figure 4.2. We can see clearly that the node-based structure we observed in the original correlation matrices is largely reproduced when using a Gaussian noise estimate. This conclusion helps confirm that apparent node-based structure in C_{ij} is driven primarily by sky

feature amplitude, which sets the SNR. However, we have not ruled out more subtle effects due to systematics.

4.2 Investigating Excess Correlation

While we see that power on the sky is a driving factor for the node-based structure, we want to verify whether there are other factors contributing at a lower level. One clear way to do this is to create a correlation matrix of FEM load data - in this case sky power is not a factor, and we would nominally expect baselines of all lengths to behave the same. In Figure 4.3 we see this FEM load correlation matrix. In this case we are finally including all baselines in array at the time this data was taken, and therefore see a much larger matrix. We have neglected to include antenna labels as there is insufficient space to display them at a readable size, but it is still the case that each row and column references a single antenna such that each pixel is an individual baseline, with auto-correlations laying along the diagonal. Again, node boundaries are denoted by thick blue lines, and SNAP boundaries by thin blue lines. The reader may refer back to Figure 2.2 for a more thorough explanation on reading these correlation matrices. The value of C_{ij} displayed in all of the following matrices is averaged across eight integrations and all 1536 frequency channels.

We find in Figure 4.3 that there is indeed residual node-based structure even when we do not observe the sky. Again we emphasize that our digital system does not know about node boundaries and we do not expect signal contamination within a node, and therefore any node-based structure in these matrices is a serious concern. The color bar on this matrix is constructed such that values of C_{ij} within the range consistent with noise are shown on a linear gray scale, whereas values above the noise level are shown on a logarithmic color scale. As the source of this data is noise, we would expect the correlation metric to also be noise-like. We estimate the expected noise level here by utilizing the interleaved correlator outputs described in Section 2.1.1 - rather than summing the even and odd outputs to get our normal data products we can instead subtract them. These differences are written out and stored along with the data, and therefore can be used to calculate a correlation metric for the noise to get an estimate of expected metric values. Therefore, we need to note two features that we identified in Figure 4.3. First, C_{ij} is substantially elevated for baselines within a node, and is even more elevated for baselines sharing a SNAP. Second, even for baselines between nodes we see that the average value of C_{ij} across the array is significantly higher than the noise

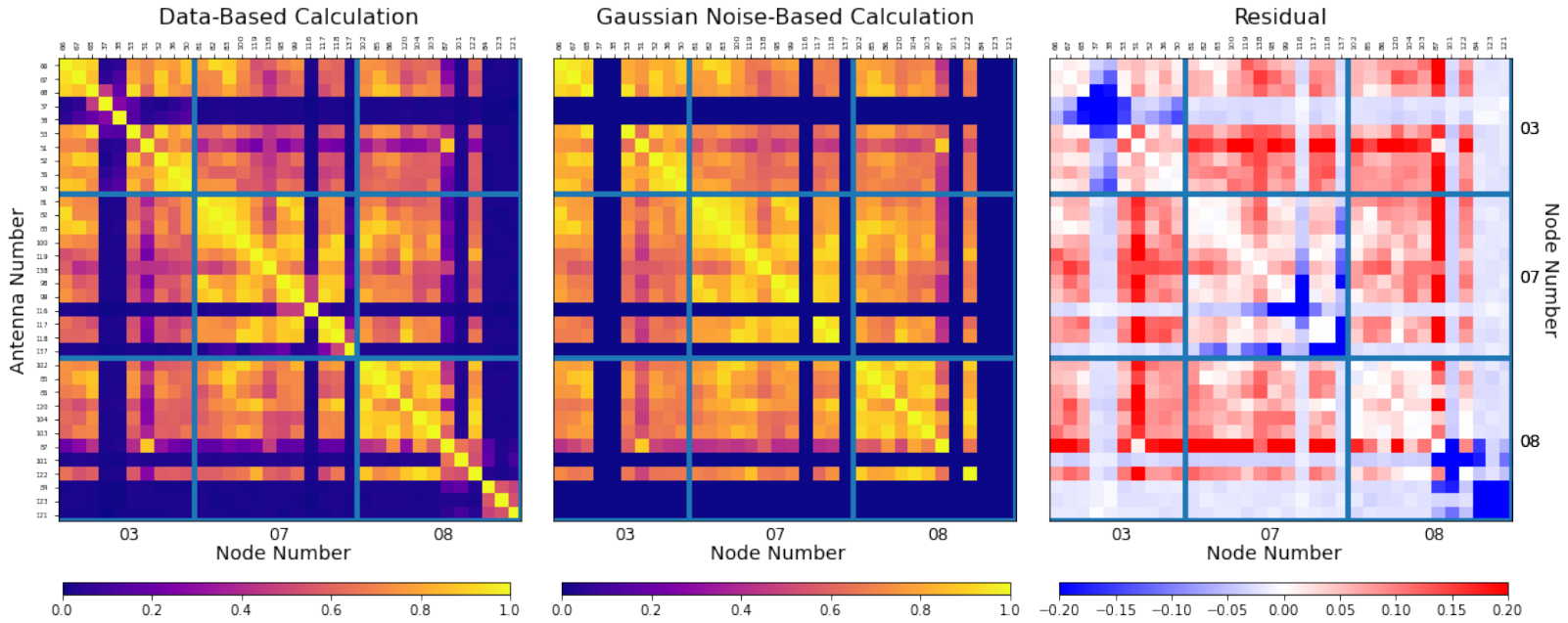


Figure 4.2: Comparison of the correlation metric computed using true noise from the data (left) and simulated Gaussian thermal noise calculated using the auto-correlations (middle), along with the residual (right). We see here clearly that the node-related structure observed in Figure 2.4 is fully reproduced using simulated Gaussian noise in lieu of the measured noise used in the original calculation.

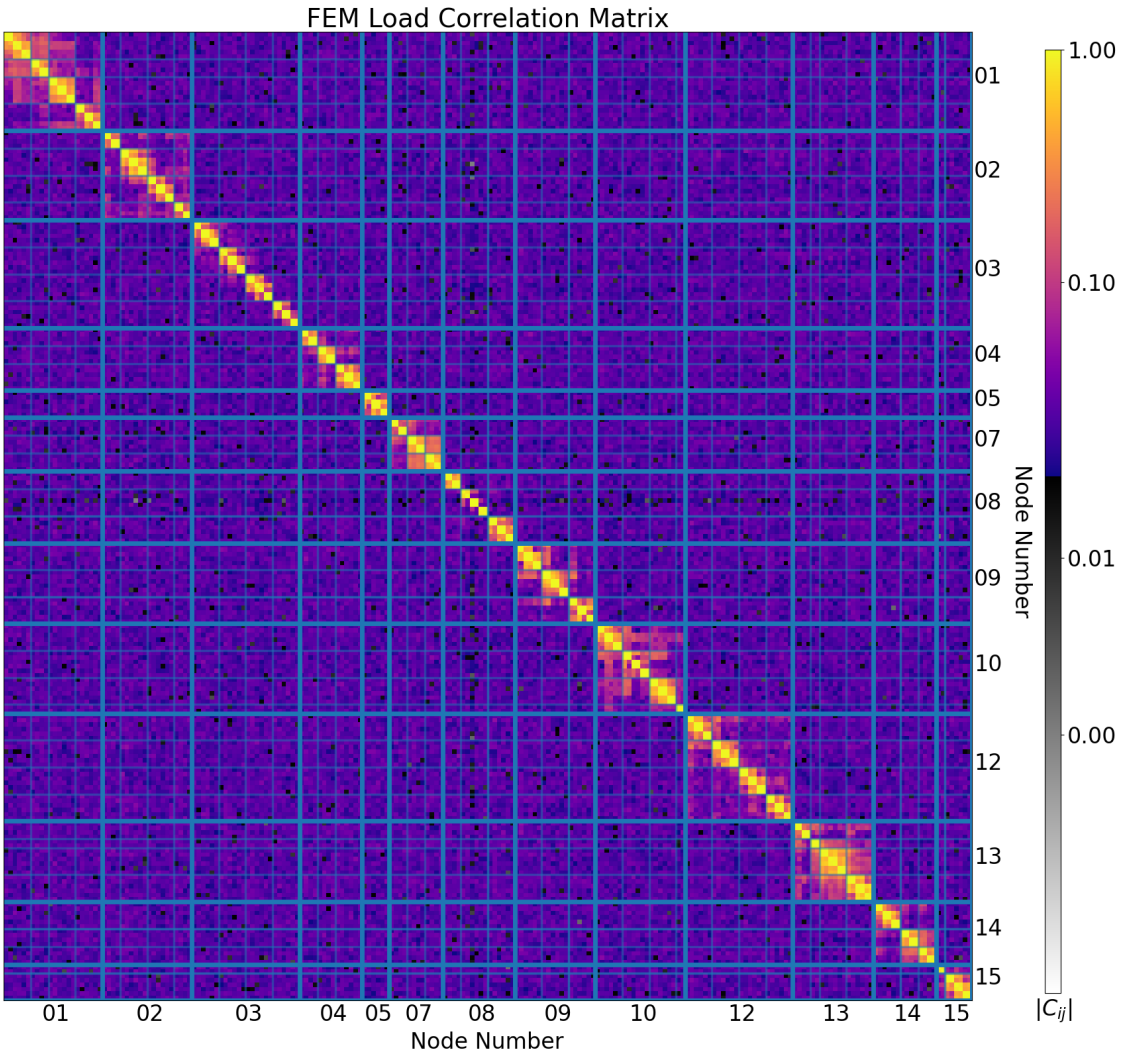


Figure 4.3: A correlation matrix using FEM load data where each pixel value C_{ij} is calculated as shown in Equation 4.1. The linear grayscale portion of the color bar represents values that are consistent with noise, while the logarithmic purple-yellow scale represents values above the noise level. More thorough description of how to interpret this matrix can be found in Chapter 2.

level. Both of these features were deeply concerning to observers and spurred collaboration-level efforts to find and mitigate their causes. The remainder of this chapter will primarily explore the cause of and solution to the first feature, and we will briefly discuss the second feature in Section 4.4.

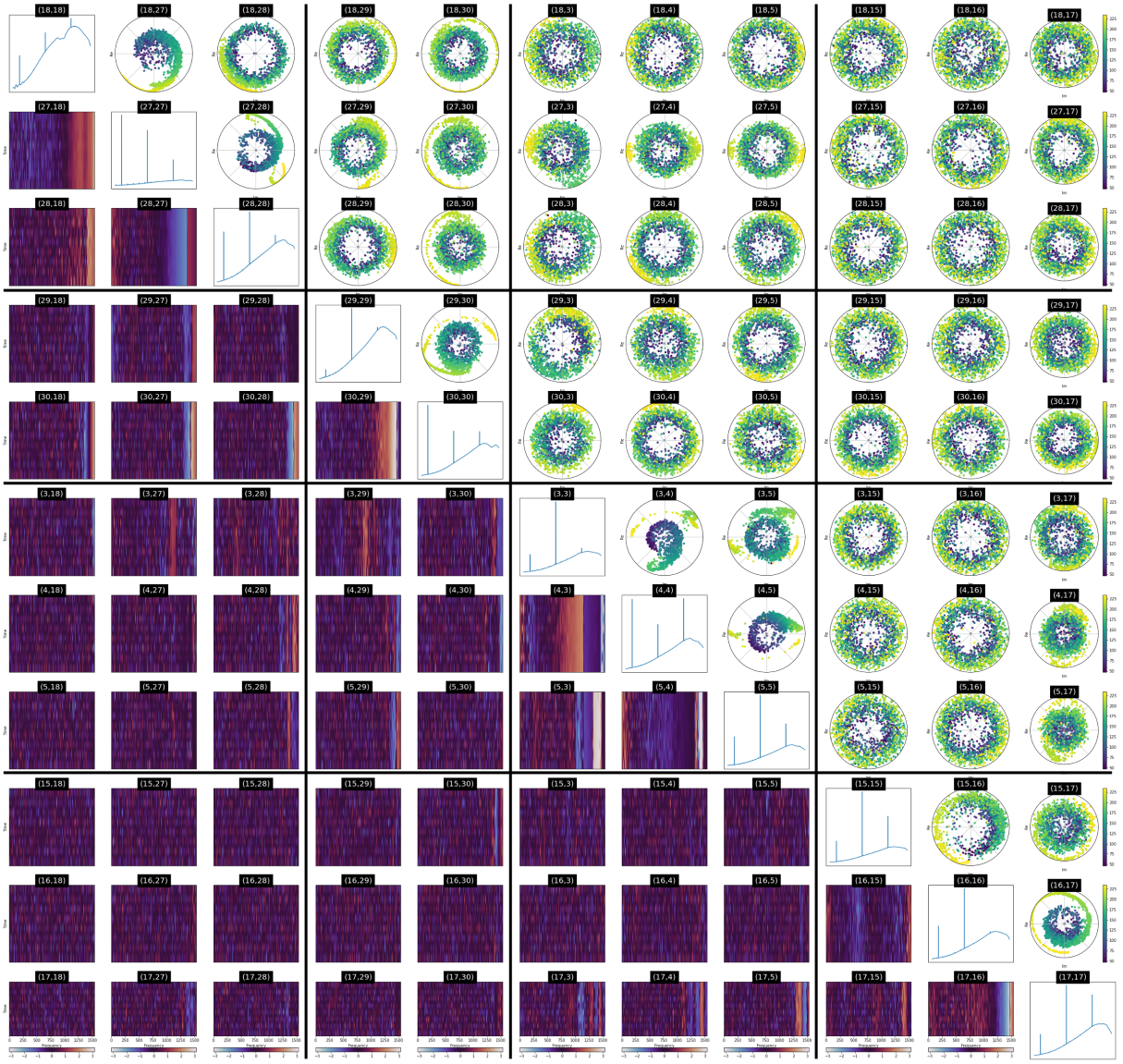


Figure 4.4: Visualization of FEM load visibilities for all baselines within a single node. On the diagonal are single time slices of auto-correlation spectra. Below the diagonal are phase waterfalls. Above the diagonal are charts similar to a Smith chart - amplitude is given by radius, phase given by angle, and frequency given by color. Solid black lines denote the boundary between different SNAPs.

Excess correlation for baselines within the same node implies some sort of leakage or cross-talk between signals as they pass through the node box. This is a form of signal injection which is very dangerous to our analysis and science goals, and therefore must be better understood. In an effort to visualize how this may be happening we turn to looking at the phase of our visibilites. For FEM load data we expect visibilities for all baselines to have phases that follow a Gaussian noise distribution. Any deviation from noise in the phases is an indication that the visibility contains some form of true signal. Figure 4.4 shows phase information for FEM load data from one node. Below the diagonal are phase waterfalls for every baseline within the node, shown using a wrapping colorbar such that phase wrapping appears continuous. The diagonal contains slices of the auto-correlation spectra amplitudes at a single time (auto-correlations are real-valued numbers, and therefore do not have phase information to display). Above the diagonal are visibility diagrams for each baseline that somewhat resemble a Smith chart. The log amplitude of each data point is shown by the distance from the origin, the phase is given by the angle around the circle, and the frequency is depicted by color. Black lines denote the boundary between different SNAPs. For a noise source, we expect these diagrams to have a noisy donut shape - the data should have a relatively stable amplitude, but the phase should be randomly distributed and have no frequency structure. Similarly, the phase waterfalls should have no structure beyond the noise. However, in both cases we see significant structure on some baselines, particularly on baselines within the same SNAP.

To further clarify this diagram, Figure 4.5 shows an enlarged diagram for the first SNAP from Figure 4.4. We see that baselines (18,27) and (27,28) show significantly correlated structure in both their phase waterfalls and 'Smith charts', especially at high frequency. Baseline (18,28) has a much cleaner phase diagram, especially at lower frequencies, where the phases are distributed in a noise-like way and the amplitude is a stable band, forming a donut shape. However, at high frequency we begin to see correlation between the frequency and phase. We expect a noise source like this to have noise-like phases, and therefore any structure beyond noise in these diagrams is considered evidence of a systematic.

As with other systematics, a natural next step is to create the same correlation matrix using digital noise data. The results of this are shown in Figure 4.6. We clearly see that with digital noise data the excess within nodes is not present. This means the primary component of the systematic

is being introduced somewhere between the FEM load input and the digital section of the SNAP, which makes the PAM a primary candidate as the source of this excess correlation. As a reminder for the reader, there is one PAM per antenna, and each PAM processes both polarizations of an antenna in parallel. At this point the best path forward is to perform targeted tests on the PAMs, which will be outlined in the next section. However, we do note that even with digital noise data the

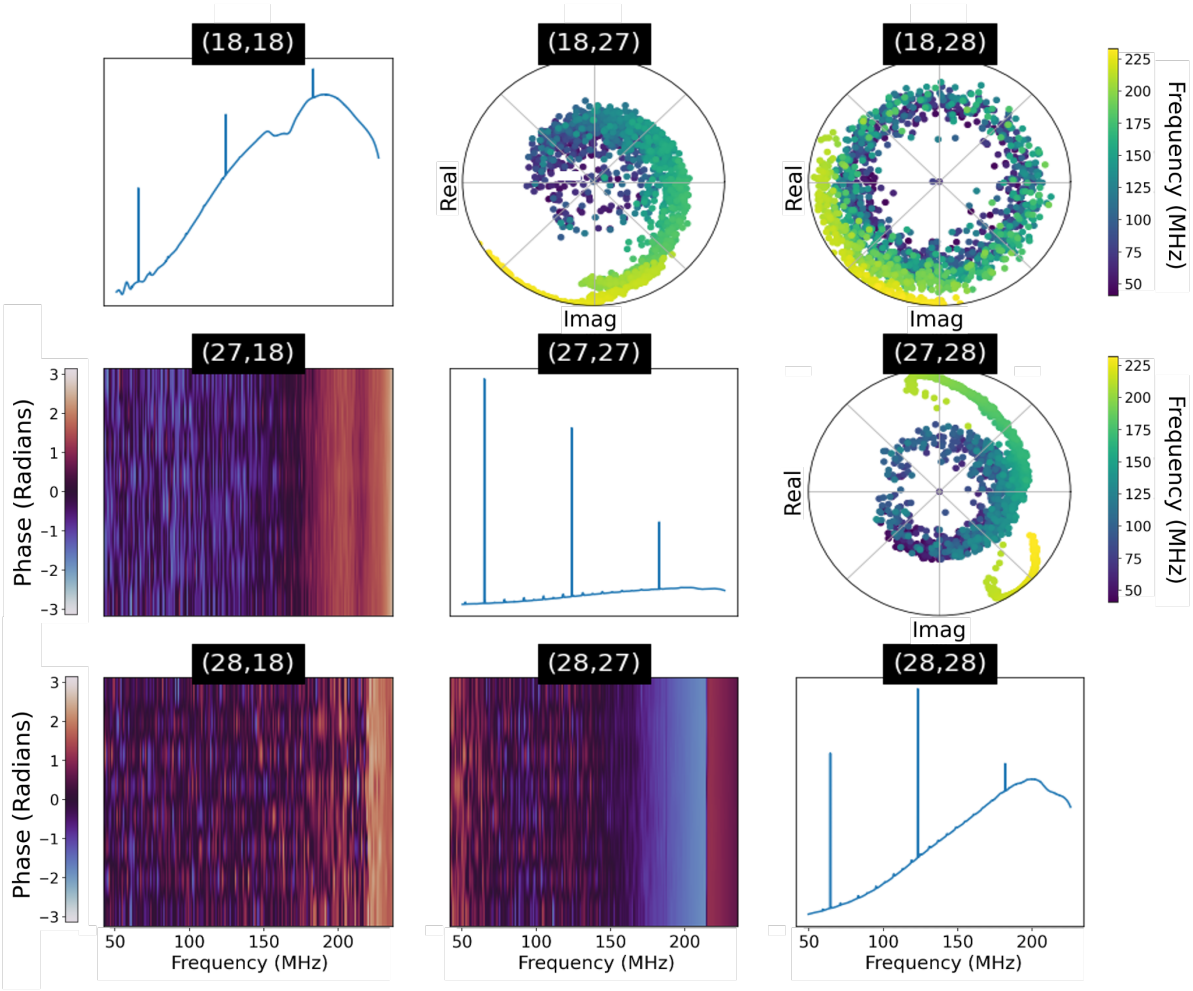


Figure 4.5: A single SNAP from Figure 4.4 shown at larger scale for better visualization. The diagonal contains auto-correlation spectrum amplitude, below the diagonal contains waterfalls of visibility phase, and above the diagonal contains visibility diagrams where the phase is given by the angle, amplitude by distance from the origin, and frequency by color.

overall correlation level across the array is elevated above the values predicted by noise, indicating that this array-wide excess may be a digital issue. We will discuss this further in Section 4.4.

4.3 Lab Experiments

In an effort to more effectively investigate the node-based excess correlation, we sent a subset of components back from the HERA site to construct a full replica of one node in the lab at UC Berkeley. This included the same noise sources that are used in the FEMs in the field feeding into the same signal chain all the way through where the signal would leave the node box in the field. This allowed experimenters to attempt to reproduce the systematic and perform unit tests by making changes to specific subcomponents and comparing against the baseline. While excess correlation was our most concerning systematic from an analysis perspective, we used this opportunity for lab testing to investigate other systematics as well. In fact, the reader will recall from Section 3.2.5 that the temporal structure systematic was also present in the FEM load data but not in the digital noise data, making the PAM a primary suspect in that case as well. Without a full correlator for our lab setup we were unable to produce correlation matrices from lab measurements, but we were able to examine the temporal and phase structure of the visibilities.

Figure 4.7 shows a mean-subtracted auto-correlation waterfall for 'FEM' load data measured in the lab. As evident in the figure, experimenters were quickly able to reproduce the temporal structure we observed in actual FEM load data. Not only do we see temporal structure, but we see the same spectral character in this waterfall that we saw in Figures 3.9 and 3.10. More specifically, we see that the character of Figure 4.7 is extremely similar to that of antenna 90 in Figure 3.10, giving hope that finding a solution in the lab could translate to a solution for the whole array. Additionally, we note that this temporal structure is effectively introducing artificial signal to the data - depending on the exact nature of this signal, it has the potential to correlate between inputs and cause an artificially elevated correlation metric. Therefore, we anticipate that mitigating this temporal structure may also be a remedy to our excess correlation problem.

This thesis will not detail the great many lab tests that experimenters ran, but rather will give the reader the great pleasure of jumping straight to the answer. Experimenters found that wiggling cables going into the PAMs could increase temporal structure, but found that the cable connections themselves were sound, and therefore became suspicious that the housing for the PAM board itself

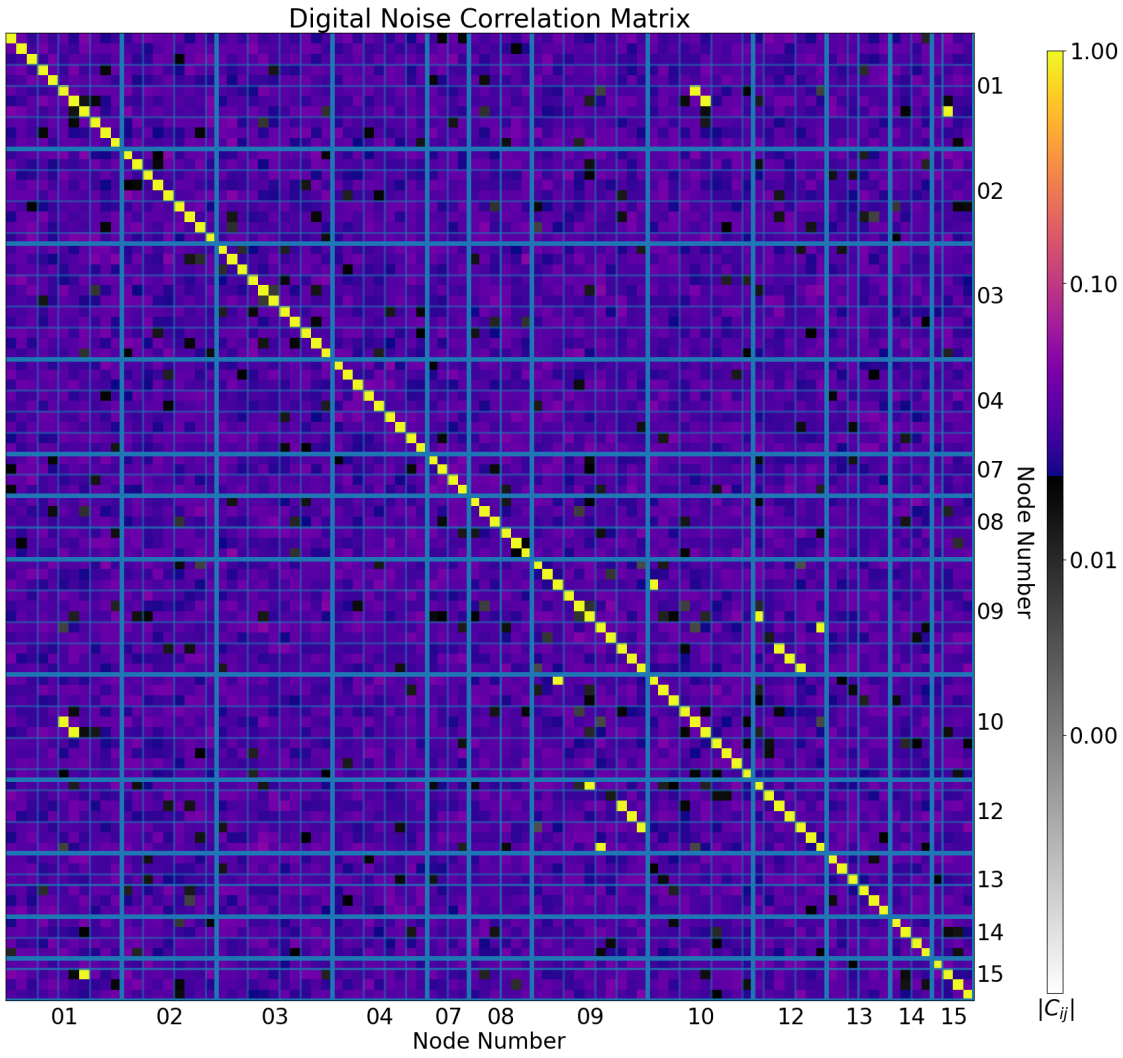


Figure 4.6: A correlation matrix using digital noise data where each pixel value C_{ij} is calculated as shown in Equation 4.1. The linear grayscale portion of the color bar represents values that are consistent with noise, while the logarithmic purple-yellow scale represents values above the noise level. More thorough description of how to interpret this matrix can be found in Chapter 2.

was loose, allowing the board to intermittently lose conductive contact with the housing. Figure 4.8 shows the configuration of the PAM board inside the housing in the left panel and outside the housing in the right panel. If the PAM board loses conductive contact with the housing along the edges this will create a cavity, which functions as a slot antenna that will both radiate and absorb radio-frequency signals. This has the potential to explain the spurious structure seen in the waterfalls, and could have ramifications for the excess correlation issue as well. In order to address this issue, experimenters proposed adding thin strips of conductive foam to the edges of the PAM board in order to create a tighter fit within the housing that will not allow for loss of contact. This conductive foam can be seen along the edges of the board in Figure 4.8b.

After applying the conductive foam experimenters continued to take lab data for several weeks and found that there was no temporal structure remaining in the data. Figure 4.9 shows a mean-subtracted waterfall using 13 hours of data collected in the lab with re-fitted PAMs - comparing against Figure 4.7 we see substantial improvement. In fact, even after looking at several weeks

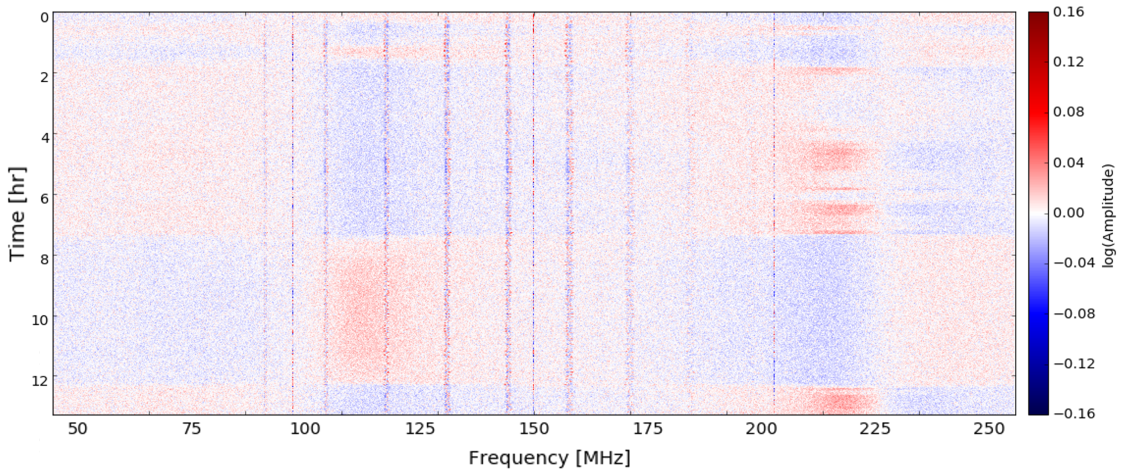
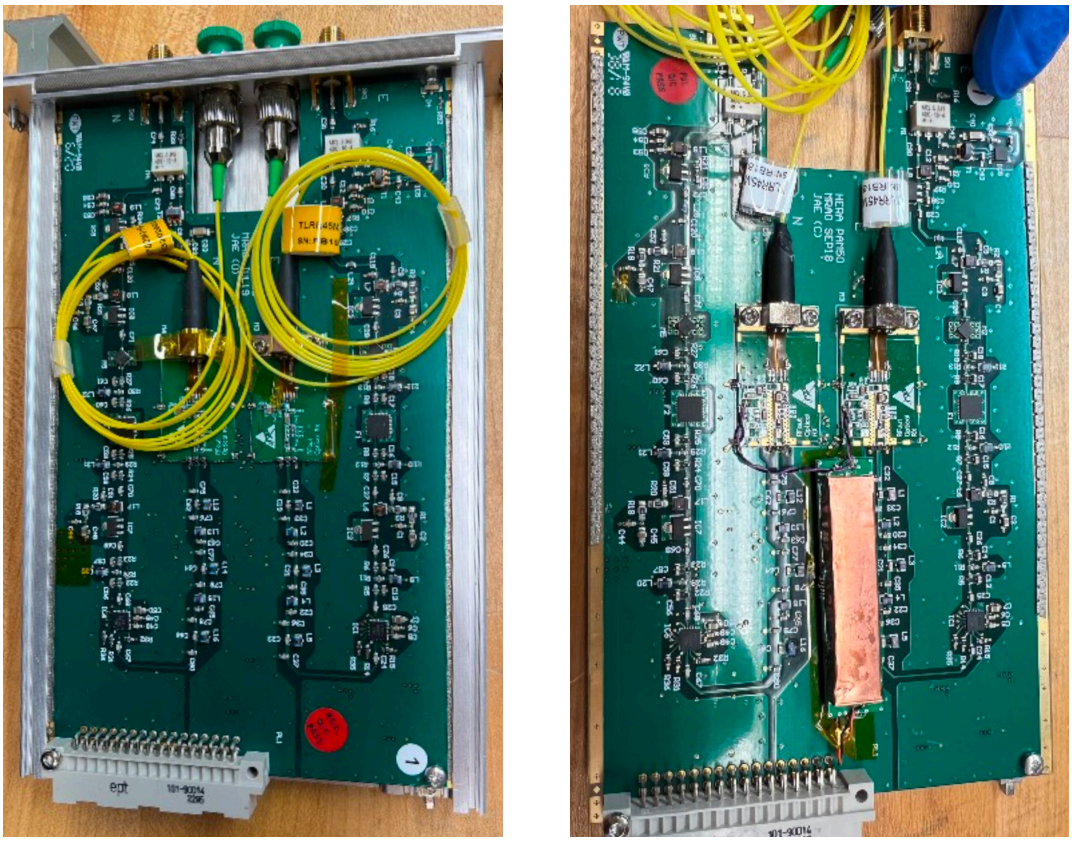


Figure 4.7: A mean-subtracted waterfall of auto-correlation noise data collected in the lab. This waterfall covers 13 hours of time, and shows significant temporal and spectral structure that is not expected from noise data.



(a) (b)

Figure 4.8: Pictures of the Post-Amplification Module (PAM) inside of the housing (a) and outside the housing (b). Panel (b) shows the conductive foam applied along the edge of the board to create a more secure fit with the housing.

worth of data, there was no sign of temporal structure after the PAM update. With such resounding results, the commissioning team decided to apply this refit to all PAMs in the array.

4.3.1 PAM Refit Results

In September 2022 several members of the commissioning team including myself traveled to the HERA site in South Africa to perform a refit of all PAM boards in the array. This involved uninstalling all PAMs from the node boxes, transporting them to the on-site lab, and the rather

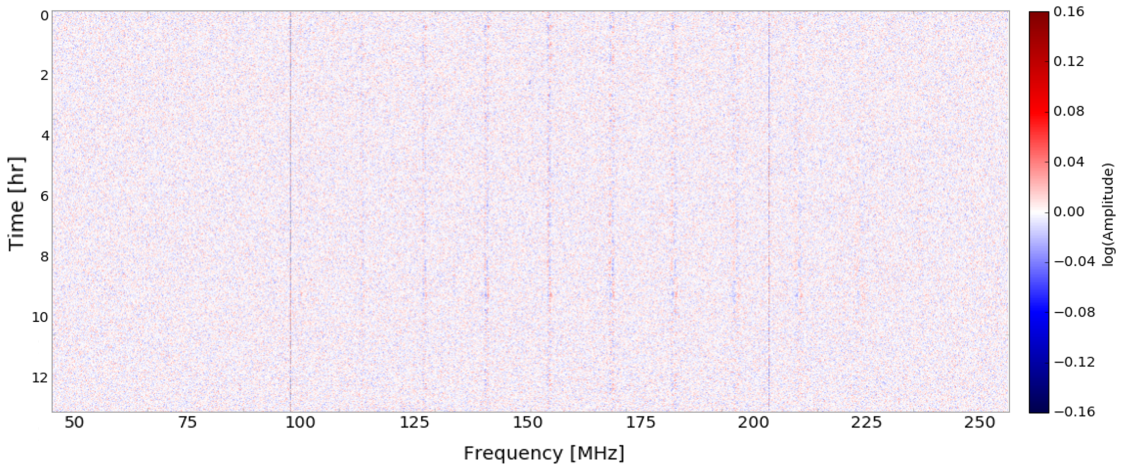


Figure 4.9: A mean-subtracted waterfall of auto-correlation noise data collected in the lab after conductive foam was applied to the PAM board. This waterfall covers 13 hours of time, and shows no visible temporal structure.

Figure 4.10: Picture of myself in the HERA onsite lab sliding a refitted PAM with conductive foam back into its housing. As part of the refit we removed as much dust and sand from both the boards and the housing as possible, but nonetheless grit on the rails made it quite strenuous to both remove and replace the PAM boards. This is evidenced by both my facial expression and my Band-Aid.

tedious task of precisely cutting a tremendous quantity of foam strips, and removing the boards from their housing to apply the conductive foam. Figure 4.10 shows a picture of myself in the HERA on-site lab replacing a refitted PAM board into its housing. Due to the accumulation of dust and sand, an unavoidable element of a desert telescope, this process was often rather strenuous.

After a full week at the HERA site refitting over 250 PAM boards, we could collect new FEM load data to inspect for both temporal structure and excess correlation features. First looking at temporal structure, Figure 4.11 shows a comparison between FEM load data collected before and after the PAM refit for five antennas that are representative of the array overall. We see that while there is still some temporal structure after the refit, the vast majority of the structure has been mitigated. Further, we see that the remaining temporal structure is spectrally smooth, which is less concerning from an analysis perspective.

Next, we create a new correlation matrix using FEM noise data collected after the PAM re-fit, shown in Figure 4.12. We immediately see that the node-based excess correlation structure is gone. There are still some cases of excess correlation for antennas within the same SNAP, but this is a minor systematic that effects significantly fewer baselines than the node-based structure. We also note that the overall correlation level across the array has been substantially depressed compared to Figure 4.3 - we will discuss this change further in the next section. We still see that the average correlation is slightly above the range that is expected based on noise (given by the gray scale color bar), but we see many more grey pixels than we did before the refit. Additionally, we don't see any concerning outliers on baselines between different SNAPs. To further verify our improvement we refer to Figure 4.13, which shows the phase structure for just one sample node using load data after the refit. Comparing against the version of this figure before the refit (shown in Figure 4.4), we see that the phase waterfalls and the 'Smith charts' show substantially less correlated structure than they did previously. In fact, both show very little structure beyond noise, which is exactly what we would expect. We inspected these phase diagrams for all nodes and polarizations in the array, and found the same level of improvement across all elements. For another view at this fix, Appendix A shows several histograms of the correlation metric before and after the fix, which also show clear resolution of the systematic. This is conclusive evidence that the temporal structure and node-based excess correlation were intricately linked systematics, and that the correlation issues have been resolved.

Unfortunately, while temporal structure has been largely mitigated in FEM load data we still see substantial temporal structure in sky data. This confirms our initial suspicions that the temporal structure and worm systematics are multi-causal. At this point our most likely culprits are changes in feed position or cable connectivity at the feed due to high winds or computational issues in the SNAP or correlator that only arise when processing higher power data, and therefore are not visible in the load data. We have sought a correlation between high winds and increased temporal structure, along with other weather patterns, and no such relationship has been found. Therefore, we remain suspicious of the digital system and see that more continuing commissioning efforts will

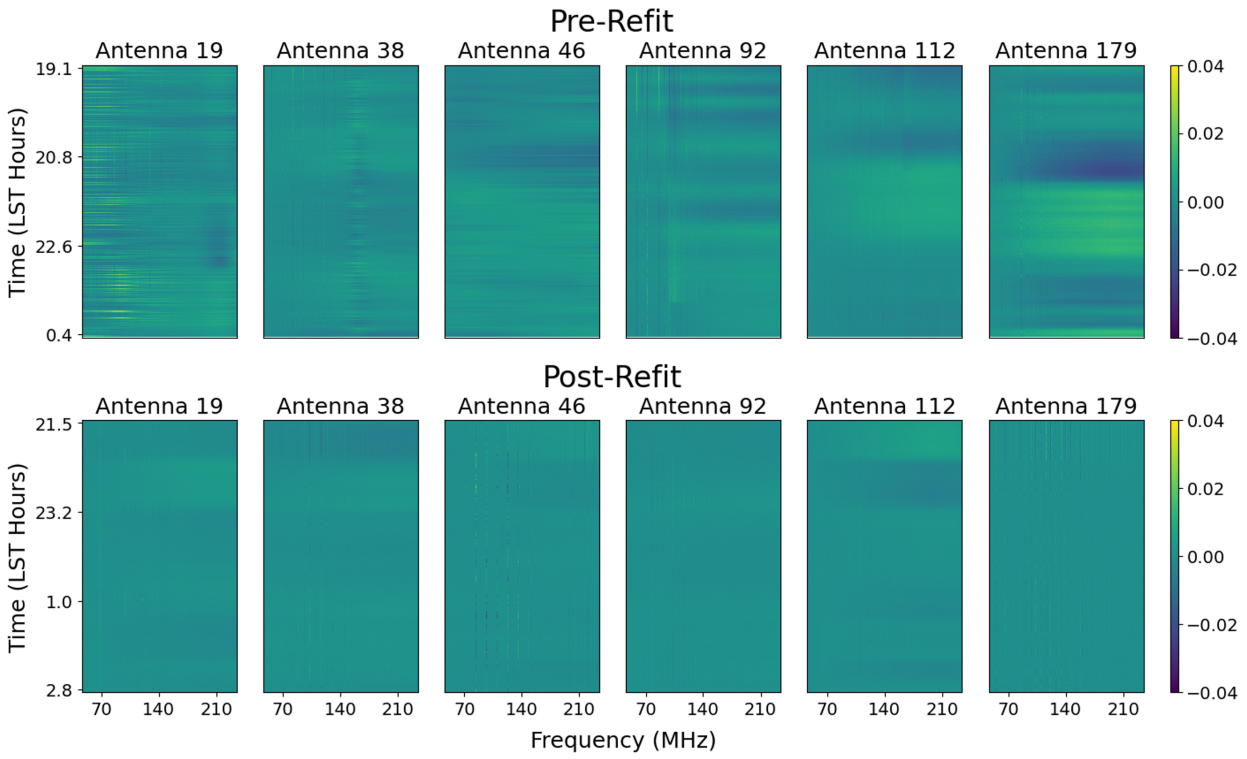


Figure 4.11: Waterfall plots of FEM load auto-correlations taken in the field for five representative antennas both before and after the PAMs were refitted with conductive foam. We clearly see a substantial reduction in temporal structure between the two data sets. Some low-level temporal structure remains in the data after the refit, but it is not as bright and has significantly less spectral shape.

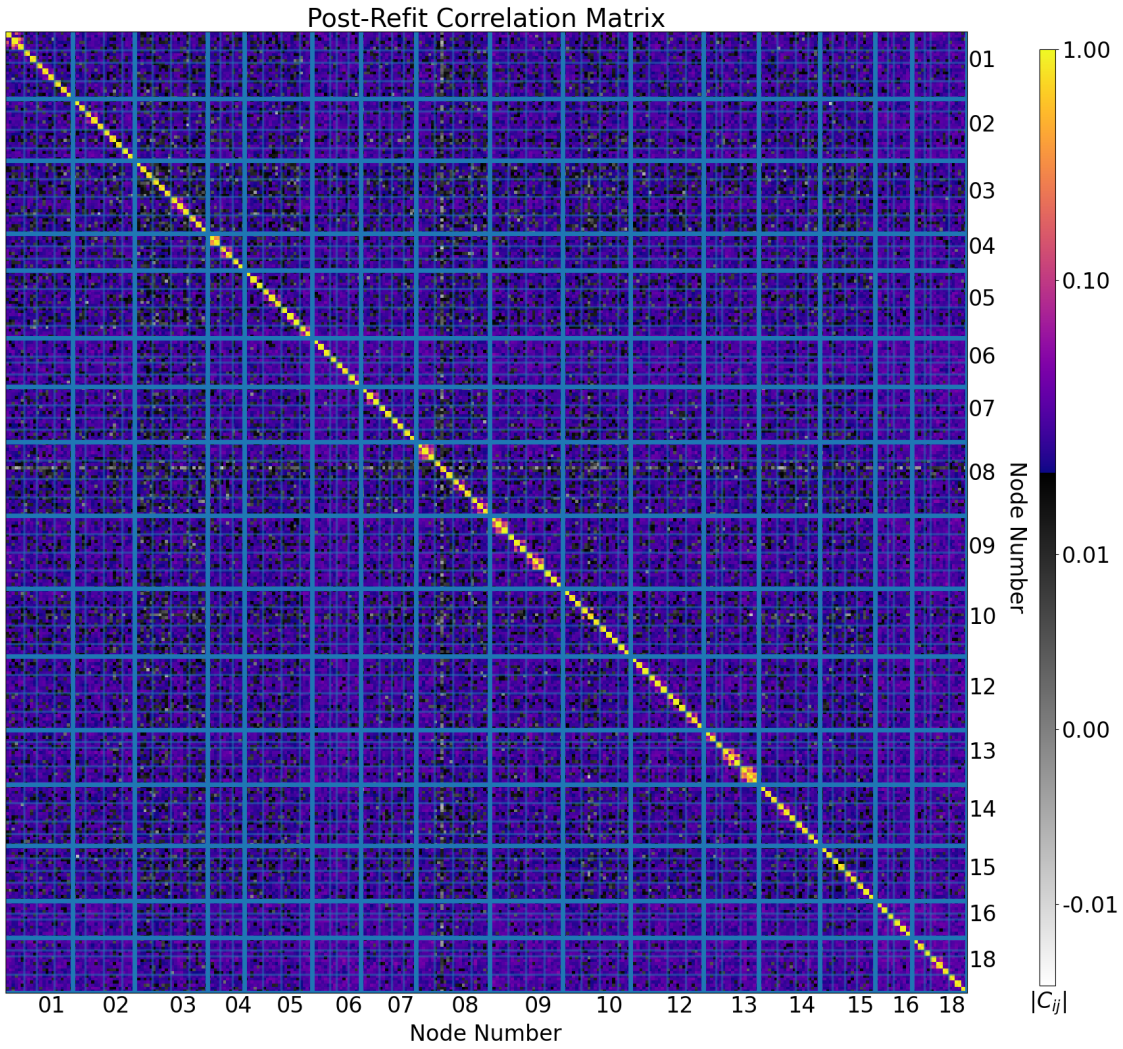


Figure 4.12: A correlation matrix using FEM load data collected after the PAM refit, where each pixel value C_{ij} is calculated as shown in Equation 4.1. The linear grayscale portion of the color bar represents values that are consistent with noise, while the logarithmic purple-yellow scale represents values above the noise level. More thorough description of how to interpret this matrix can be found in Chapter 2.

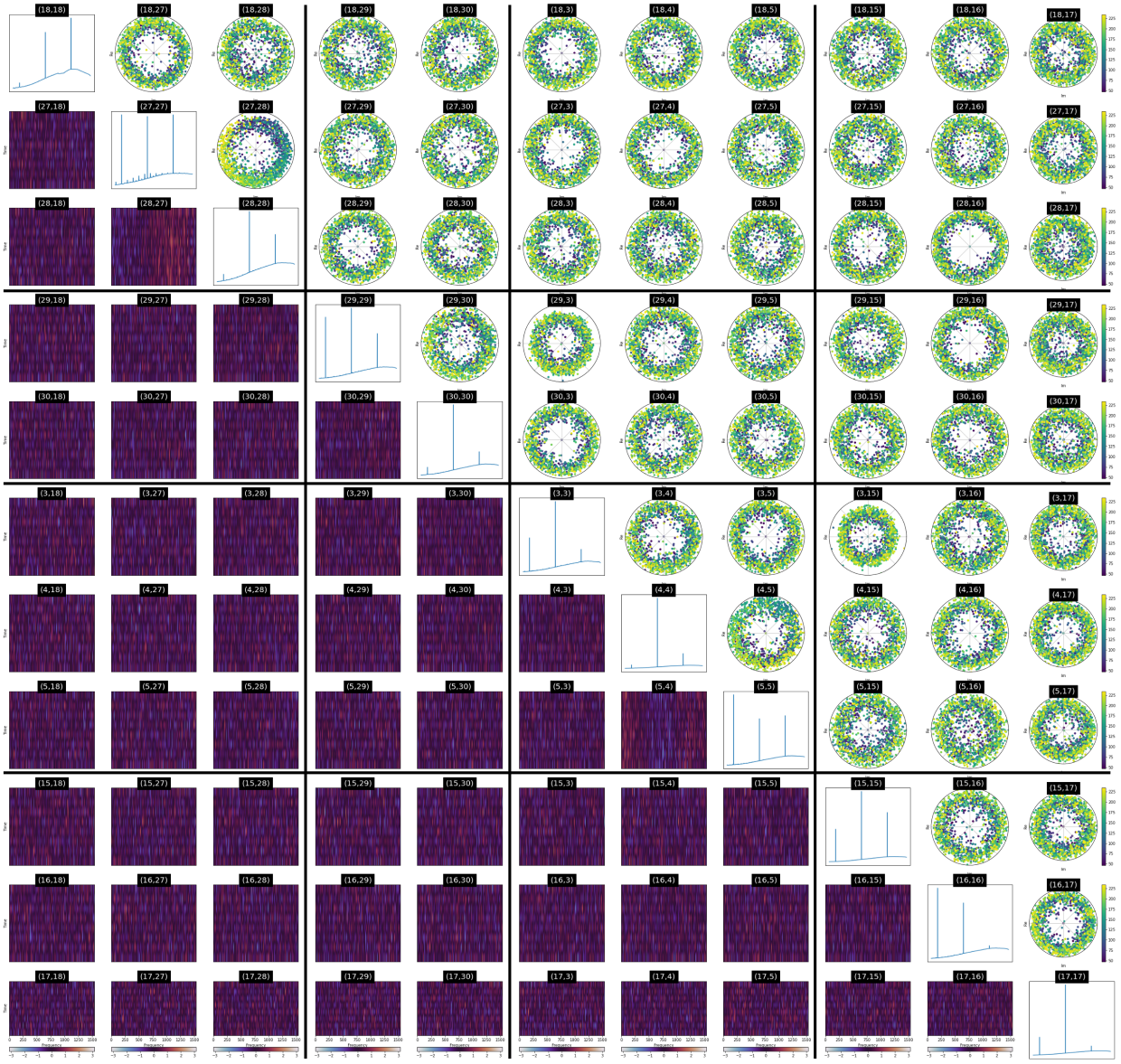


Figure 4.13: Visualization of FEM load visibilities for all baselines within a single node, using data taken after the PAM refit. On the diagonal are single time slices of auto-correlation spectra. Below the diagonal are phase waterfalls. Above the diagonal are charts similar to a Smith chart - amplitude is given by radius, phase given by angle, and frequency given by color. Solid black lines denote the boundary between different SNAPs.

be necessary to resolve the matter of temporal structure. For the purposes of the analysis in this thesis, we will use data selection and antenna flagging to avoid the remaining temporal structure to the extent possible.

4.4 Inter-Node Correlation Floor

As mentioned previously, in addition to the elevated correlation within nodes we were also concerned that the overall correlation level across the array, and particularly between different nodes, was much higher than the level predicted by the noise. This is evidenced quite clearly in Figure 4.3, where the values within the range predicted by noise are plotted on a gray scale, yet we see the vast majority of pixels are above this level on the color scale. Remembering that this matrix is computed using FEM load data, this excess above the noise indicates that there is some correlated signal being injected into the signal chain somewhere after the FEM. This is a deeply concerning systematic that our correlation metric allowed us to uncover, as any signal injection interferes with our ability to make a precision measurement of the sky. As experimenters were working to reproduce systematics in the lab, these matrices also spurred the correlator team to investigate any avenues for the digital system to be producing this systematic. The correlator team did successfully identify a digital bug that was causing signal leakage across inputs and implemented a fix in September 2022, the same week that we implemented the PAM refit. Therefore, the changes from this correlator upgrade are also reflected in Figure 4.12, where we see a substantial decrease in the overall correlation level with many more of the pixels being grey scale. Unfortunately, we do not have any data that contains only one of these upgrades, and therefore cannot conclusively separate the result of each upgrade. However, our understanding of the fundamental nature of each systematic and its fix makes observers confident that the PAM upgrade would only affect the node-based structure, while the digital upgrade would only affect the overall correlation floor.

However, despite substantial improvement from before the upgrades, we still see that the correlation floor in Figure 4.12 is more purple than gray, meaning that there is still an excess. This spurred our correlation team to further investigate potential causes. We discovered that the correlator was doing a much finer interleave than commissioners had previously thought, meaning that the even and odd visibilities were not sufficiently separated in time. This meant that the polyphase filter bank was incorrectly combining data across time stamps, leading to spurious signal correla-

tion. We can confirm that this is the source of elevated correlation by creating a correlation matrix where we use the full 9.6 second integrated visibilities to construct V_{ij}^{even} and V_{ij}^{odd} in Equation 4.1 rather than the finely interleaved visibilities. Such a matrix is shown in Figure 4.14, and we clearly see that the correlation floor has significantly dropped and is almost entirely within the range of values predicted by the noise. For this matrix we use eight integrations to construct the even set and eight to construct the odd set such that the even and odd sets have the same data volume as previous matrices.

The correlator was originally designed to perform a fine interleave and save the difference between the even and odd sets specifically so that the differences could be used as a precise measurement of noise. These differences would be calculated on timescales substantially shorter than any relevant sky rotation, and therefore would be considered true measurements of instrumental noise. Given this realization that the interleave time is too fine, we choose in our analysis to use the difference between adjacent 9.6 second visibilities rather than the finely interleaved differences as our noise estimate. Our full integrations are still short enough that changes on the sky are incredibly small, and we consider that effect to be less significant than the excess correlation seen in the finely interleaved set.

With the PAM and correlator system upgrades and this analysis choice, we consider our data to be free from excess correlation and are now able to proceed with a power spectrum measurement. In fact, the resolution of these systematics has been a major factor in determining the analysis data sets for all major HERA analysis efforts. In the next chapters we will outline our data selection criteria, describe our analysis pipeline, and present our final results.

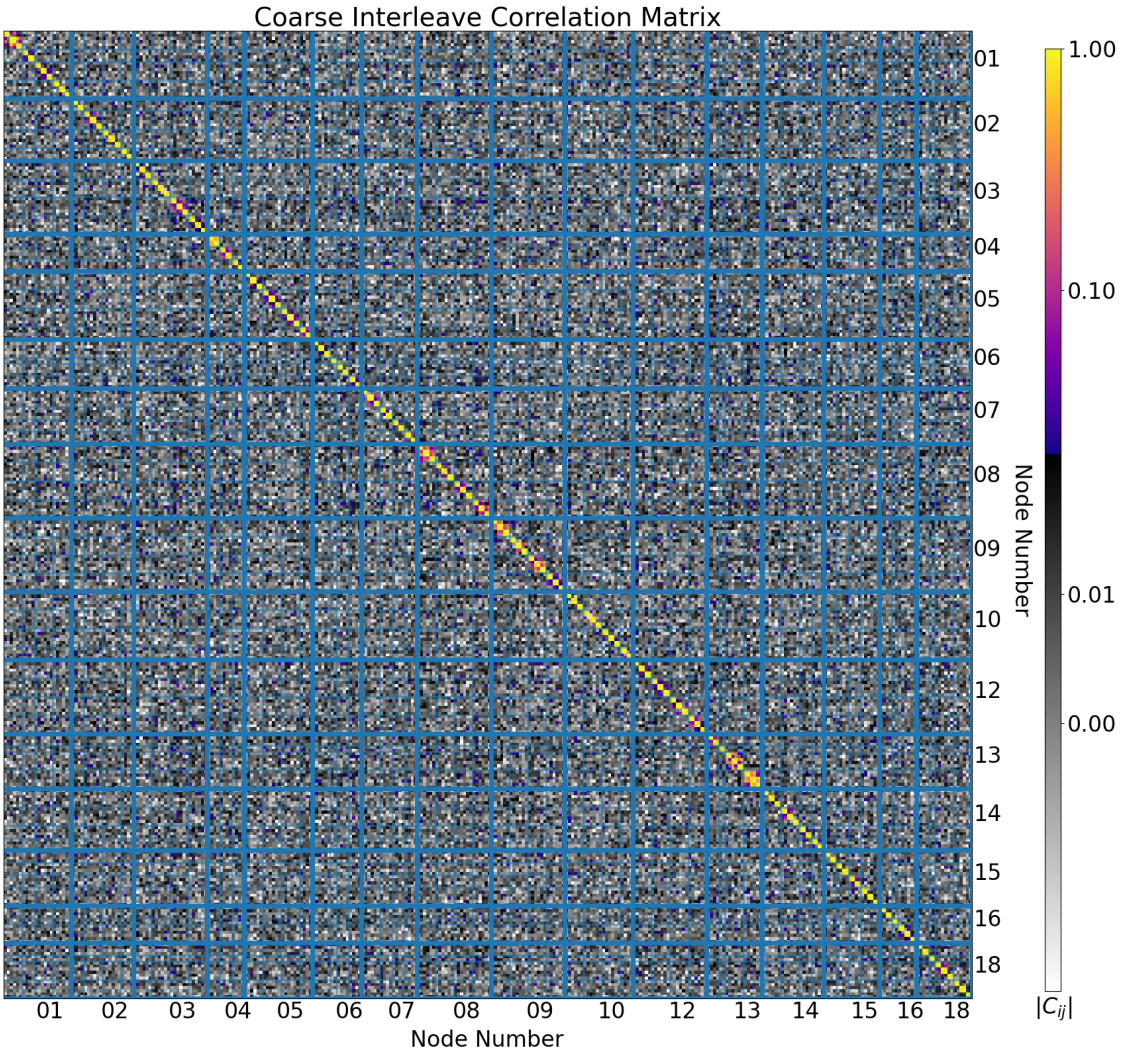


Figure 4.14: A correlation matrix using FEM load data collected after the PAM refit, where each pixel value C_{ij} is calculated as shown in Equation 4.1. However, in this case V_{ij}^{even} and V_{ij}^{odd} are constructed using time-adjacent full 9.6 second integrations, rather than the very finely interleaved even and odd visibilities used previously. The linear grayscale portion of the color bar represents values that are consistent with noise, while the logarithmic purple-yellow scale represents values above the noise level. More thorough description of how to interpret this matrix can be found in Chapter 2.

Chapter 5

HERA DATA SELECTION

To produce the best possible power spectrum result we need both high volume and high quality data. High volume largely refers to long integration time - this is primarily achieved by integrating over as many nights of data as possible. High quality refers to data that has minimal systematics present, only includes properly functioning antennas, and is sufficiently clean from RFI. We can improve the quality of any data set by 'flagging' problematic data products. Here we use flagging as a generalized term to include any procedure that removes specific data products from some or all analysis steps. Flagging can happen across times, frequencies, antennas, and entire nights, and can also happen at the individual visibility level.

This thesis represents a first look at an imaging-based power spectrum result with HERA data, with the primary goal of understanding how HERA performs in the power spectrum space given its known features and systematics. With this in mind, we aim to establish a particularly clean data set to form the basis of this initial analysis, with the eventual goal of expanding this data set in pursuit of an upper limit on the power spectrum. Particularly, if we find that HERA's systematics are manageable and our power spectrum measurement is noise-limited, we will require a significantly larger data set to achieve an optimal measurement. However, we choose to begin with a smaller data set to maintain a workable data volume for testing, visual inspection, and multiple comparative analysis runs. In the following sections we will present the reasoning for our choice of nights (5.1), frequency range (5.2), observation window (5.3), and antenna selection (5.4).

5.1 Observation Nights

First we must select which nights of observation to include in our analysis data set. HERA's observation season nominally runs from October to March, which is when the darkest parts of the radio sky are overhead while the sun is down. To aid with source subtraction during sky-based calibration and imaging, we only want to observe the darkest part of the sky when the fewest sources

are overhead. We do not have a good model of the galaxy, and therefore cannot perform successful 21 cm analysis when galactic emission is present in the primary beam (the central portion of the instrumental beam where the majority of our sensitivity lies). Our next constraint is antenna selection - as HERA is still in the commissioning process, new antennas are regularly coming online. As long baselines are particularly essential to our calibration and imaging procedures, our data set must prioritize times when antennas at the exterior of the array, and particularly outrigger antennas, are collecting data. Four outrigger antennas were added to the array and brought online part way through the 2022-2023 observing season. After these antennas were added we had three weeks of successful observing before a series of lightning storms that severely impacted the data, and eventually a correlator problem that ended the observing season. Therefore, we must use data within this three week window, which spans from November 6 to November 25 2022.[1] We will henceforth refer to observation nights using Julian Date (JD) rather than calendar dates, which corresponds to a date range of JD 2459890 to JD 2459909. Of the 20 nights of observation within this window, 17 of them were usable and will be included in the final analysis. The other three nights were considered unusable due to either excessive RFI or a lack of usable antennas.

5.2 Frequency Range

Both for data volume and analysis reasons we want to choose a subset of the HERA band to focus on. There are three primary considerations when choosing a frequency band. First, we require our frequency band to be one where the instrument has a smooth spectral response. HERA was designed with this goal in mind, and therefore this condition is met anywhere between 65 and 220MHz. Second, we need to consider the frequency limitations of the source catalog that we use for sky-based calibration. In our case, the source fluxes reported are based on measurements at 182MHz. Source fluxes at other frequencies are calculated based on a power law flux distribution using the spectral index of the source. Therefore, our flux measurements are most accurate at frequencies closest to 182MHz. Work by undergraduate student Kiana Henny found that source flux models are most reliable within $\sim$ 40MHz of the center frequency, which means we need to

[1] A persistent and debilitating correlator bug caused the 2023-2024 observing season to be unusable, and we choose not to use seasons prior to 2022 due to concerns about the excess correlation systematic (which was resolved just prior to the 2022 observing season).

observe between 140MHz and 220MHz to have the most success.

Third, we must consider at which frequencies RFI is most present. For this step, we will use the software package Sky Subtracted Incoherent Noise Spectra (SSINS, Wilensky et al. (2019)) to perform RFI flagging. SSINS calculates time-differenced baseline-averaged visibilities and makes use of the time-variability of RFI to differentiate it from sky signal. SSINS also incorporates a-priori knowledge about radio frequency allocation using a shape-matching filter. While SSINS is a high-precision algorithm that is very effective at finding and flagging even low-level RFI, we still expect better results from data sets with lower flag occupancy. This is true for two reasons. First, more flagged data means a lower usable data volume and an inconsistent data volume across

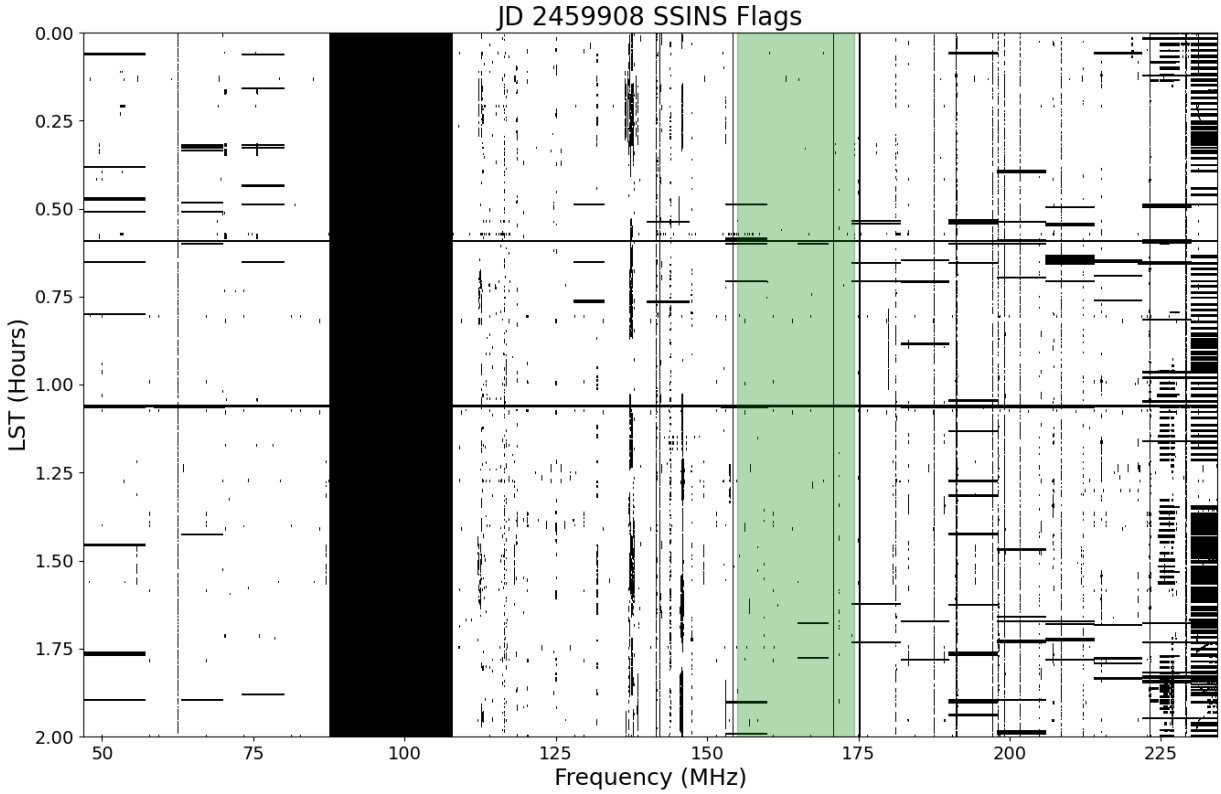


Figure 5.1: A waterfall of SSINS flags for two hours of data from JD 2459908 where black indicates flagged visibilities. The shaded green area represents the frequency band chosen for our analysis data set.

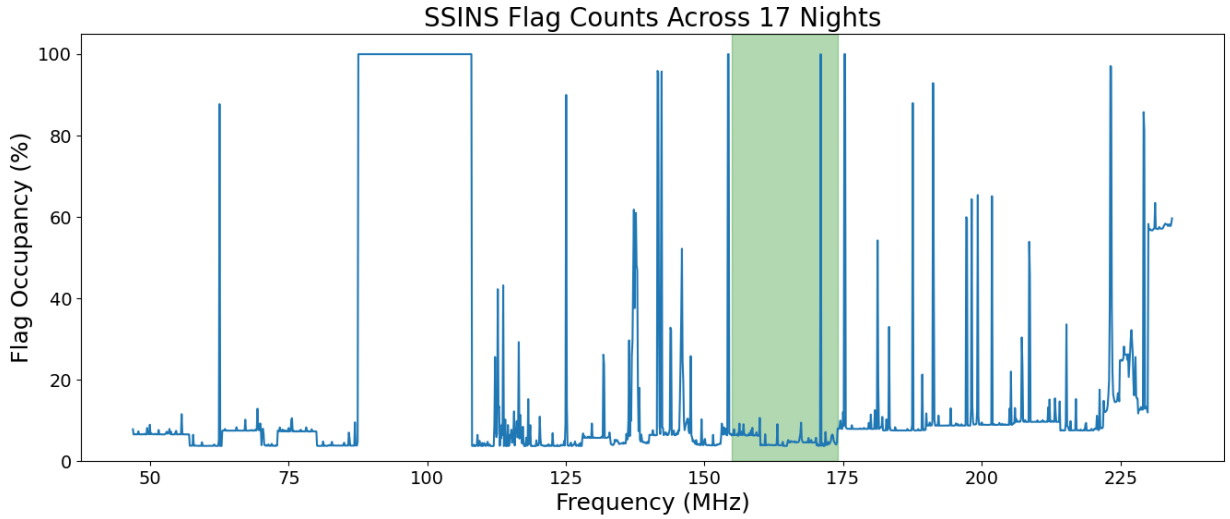


Figure 5.2: SSINS flag occupancy averaged over all nights in the analysis data set versus frequency. The shaded green area represents the frequency band chosen for our final analysis data set.

frequency bins, which can lead to a worse power spectrum result. Second, there is substantial evidence that RFI can have an extent in time and frequency and may exist at non-detectable levels in visibilities adjacent to those flagged by SSINS (Wilensky et al., 2020, 2023). Therefore, we have more confidence in unflagged data when there are fewer flags in the adjacent data. It is worthwhile to note that RFI flagging is applied per time and frequency bin, and such that all antennas and polarizations will have the same flags.

Figure 5.1 shows a waterfall of the SSINS flags for two hours of observation over the entire HERA frequency band. We see the entire FM radio band centered at 98MHz is flagged for all times. This flagging is enforced programmatically, as we know that the FM radio band is always contaminated to some degree.Other horizontal bands represent known frequency allocations that were flagged in their entirety for a given observation. Our final frequency selection is shown in green - in this case our frequency selection looks relatively clean from RFI, but we need to confirm this remains true across all times and nights we plan to use in the final analysis.

Figure 5.2 shows the flag occupancy as a function of frequency averaged over all nights in the data set. Once again the shaded green area represents our chosen frequency band. We see again

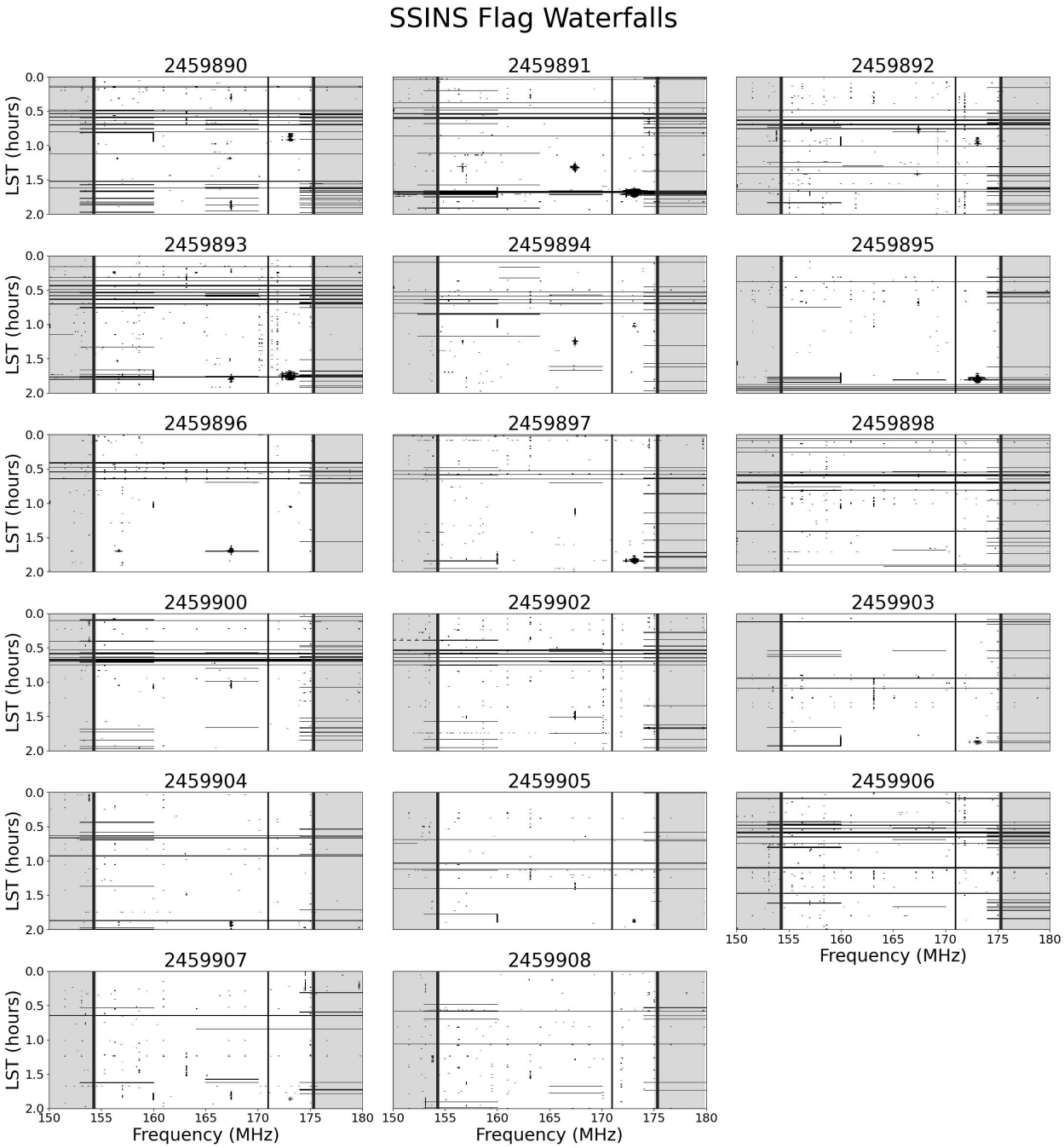


Figure 5.3: Waterfall plots of SSINS flags covering two hours of data for every night in the final analysis data set. This time range is not all included in the final analysis. The shaded grey regions are included in flagging but excluded from the final analysis set.

that the chosen band is particularly clean. We have intentionally chosen our band to only contain one known emitter at 170.9MHz, and otherwise have a particularly low average flag occupancy. While our final power spectrum result will combine information from all nights in our analysis set, it remains worthwhile to examine the flagging for each night individually. This helps inform our expectations of data and calibration performance, and helps us select the final set of usable nights. Figure 5.3 shows SSINS flag waterfalls over two hours of data for all nights in the data set. This two hour time range includes our final data range with a buffer before and after. Likewise, this plot is clipped to show only our chosen frequency band plus a small buffer on both sides, which is shaded in gray. Unsurprisingly, we find that the RFI environment is different across nights, and that some nights are significantly cleaner than others. We include the shaded regions in the SSINS flagging pipeline but not in the final analysis data set so as to include all shapes relevant to our chosen band in the flagging procedure, but exclude the two consistent emitters at 154.3MHz and 175.3MHz from the final analysis. In the next section, we will discuss time selection and how that can further aid with RFI avoidance.

5.3 Observation Window

Rather than integrating over the 10 hours of data we collect every night, we instead pick a 'window' to analyze and integrate over. This is partly because our power spectrum package `ϵppsilon` (introduced at length in Section 8.3) combines observations by aligning their measurements in sky space, and thus combining observations that measured entirely different portions of the sky is not sensible. We will see in Chapter 8 that our pipeline's usual method for combining data across time is problematic for HERA data, making this less relevant, but we mention it because it was a factor at the time we were performing data selection. Additionally, we expect some regions of the sky to have preferable foreground environments. Given the limited compute resources and scope of this thesis, our goal is to find a 30 to 40 minute window on the sky that we will analyze for all 17 nights of our data set. Referring to Figure 5.4, which shows a map of the radio sky, we see the bright galactic emission arching across the sky with a window of darker sky between the galactic center and anti-center. The range in Declination that HERA observes is given by the two blue dashed lines. We represent observation time using Local Sidereal Time (LST), which is locked on the sky such that the same LST always references the same sky. We therefore see a potential observation

window ranging from about -2 hours to 5 hours. This range does not go all the way to the edge of the galaxy because galactic emission is bright enough to affect our data even when the galaxy is on the horizon rather than overhead, and therefore we prioritize observing at times when the galaxy is as far from zenith as is feasible.

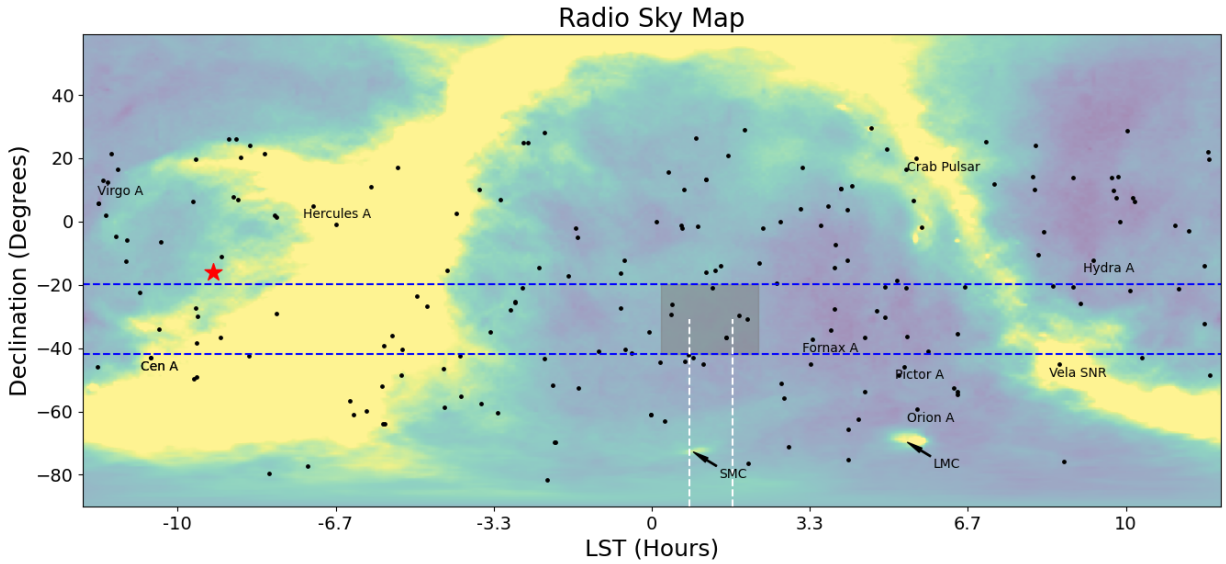


Figure 5.4: A map of the radio sky using the Haslam 2014 sky map (Remazeilles et al., 2015b), with the brightest radio sources shown as black dots, the Sun shown with a red star, and where HERA's observation band based on the extent of HERA's primary beam is between the two blue dashed lines. The shaded region represents the chosen observation window for the final analysis data set. The two dashed white lines indicate zenith of the first and last observations in this set, while the shaded region is extended to represent the primary beam.

Next, we note the source Fornax-A located at 3.3 hours. Fornax-A is an exceptionally bright radio galaxy that has a complex extended shape. We do not have very accurate models of Fornax-A, and therefore we experience difficulty calibrating properly when Fornax-A is in the primary beam or sidelobes. HERA's primary beam extends ten degrees from zenith, corresponding to 40 minutes of sky rotation, and therefore avoiding Fornax-A further limits our observation window to end at 2.6 hours, or at 2 hours if we also consider the beam sidelobes. Lastly, observations earlier in the night tend to see elevated RFI levels as people are still awake and using RFI-emitting electronics,

and LSTs -2 to 0 were found to have excessive RFI. All of these considerations lead to a usable LST window between 0 and 2 hours.

To narrow this window further, we look more quantitatively at RFI levels. Figure 5.5 shows the average flag occupancy across all 17 nights versus LST. We see a 40 minute window of clean data between LSTs 0.9 and 1.6, marked by the shaded green window. This window, also shown by the shaded region in Figure 5.4, constitutes our final observation window.

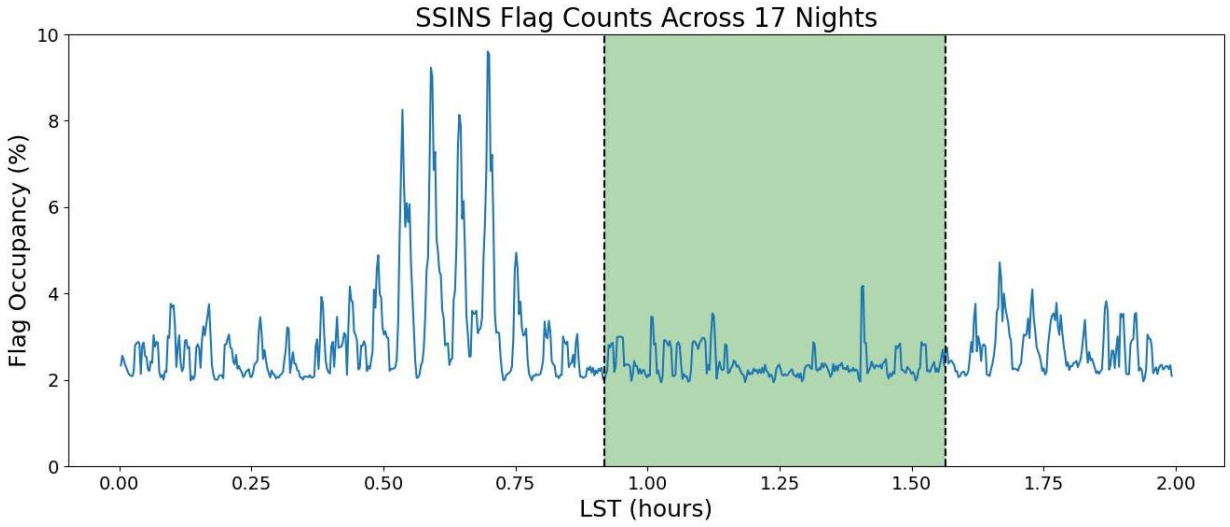


Figure 5.5: SSINS flag occupancy averaged across 17 nights of observation versus LST. The shaded green region represents the window chosen for the final analysis.

5.4 Antenna Selection

In order for calibration to work properly it is essential that all input antennas behave reasonably. Further, any antennas experiencing notable systematics will leak those systematics through into the power spectrum, and therefore must be excluded from analysis. We inspect and flag antennas using both the statistical metrics and visualization tools outlined in Chapter 2. There are seven major categories of concern that we validate during antenna flagging:

1. **Cross-Polarization:** Polarization cables are physically crossed in the field leading to incorrect measurements.

2. **Power Level:** Antennas that have too high power can cause problems in the digital system, particularly during the Fourier Transform stages. Antennas with abnormally low power are particularly susceptible to systematics and have a lower signal-to-noise, making them risky to include in analysis.

3. **Bad Bandpass Shape:** Any antenna that has a non-standard frequency response leading to a bandpass that is not sufficiently spectrally smooth. This spurious spectral structure is dangerous to our power spectrum measurements.

4. **No/Low Correlation** Any antenna where C_{ij}, as defined in Section 2.1.1, is below the acceptable threshold. This can happen at the node or antenna level due to issues with the clock system or failing to switch from observing load to sky, respectively.

5. **Worm Structure:** Any antenna exhibiting any type of frequency worm, as described in Section 3.2.4.

6. **No Sky:** Any antenna that is not observing the sky. This can be due to not receiving power from the FEM or a communcations issue where the FEM fails to switch from observing load to observing sky.

7. **Temporal Structure:** Any antenna whose visibilities do not evolve smoothly in time, as described in Section 3.2.5.

Table 5.1 shows what fraction of antennas were flagged based on each of these categories. The number shown represents the average across all nights in the analysis data set and where an antenna is considered flagged if either of its polarizations is flagged. Throughout the three weeks spanned by our analysis set there was one node whose clock was regularly failing, leading to low correlation flags for those antennas, and there were two nodes experiencing FEM power failures causing a higher than usual fraction of antennas to not observe the sky. Those two categories aside, we see that our primary contributors are frequency worms and temporal structure. This is not particularly surprising, but nonetheless unfortunate considering how challenging these systematics are to mitigate. Table 5.2 shows the number of unflagged antennas for each night in the analysis

Flagging Criteria	**Antennas Affected**
Cross-Polarization	0.8%
Power Level	2.7%
Bad Bandpass Shape	10.8%
No/Low Correlation	12.5%
Worm Structure	17.1%
No Sky	22.7%
Temporal Structure	32.6%

Table 5.1: Table showing the fraction of antennas excluded based on each metric used in antenna flagging. If an antenna was flagged based on multiple metrics it is counted in both categories.

JD	**North**	**East**	**Combined**	**JD**	**North**	**East**	**Combined**
2459890	121	119	96	2459891	109	112	80
2459892	103	110	81	2459893	109	105	82
2459894	88	109	60	2459895	96	103	61
2459896	85	89	66	2459897	100	99	58
2459898	104	111	77	2459900	112	115	91
2459902	94	96	60	2459903	94	96	60
2459904	99	99	63	2459905	83	78	43
2459906	112	121	90	2459907	102	106	69
2459908	100	91	66				

Table 5.2: Table showing the number of unflagged antennas for each analysis night out of a possible 201 antennas that were online for all 17 nights. The North and East columns show the number of flags for each polarization when they are flagged independently. This represents the set of flags used in the final analysis. The combined column shows the number of unflagged antennas in the case where polarizations are flagged together (flagging either polarization causes the other one to be flagged as well).

set out of a possible 201 connected antennas. The independent flags for the North and East polarizations represent the set of flags used in the final analysis.

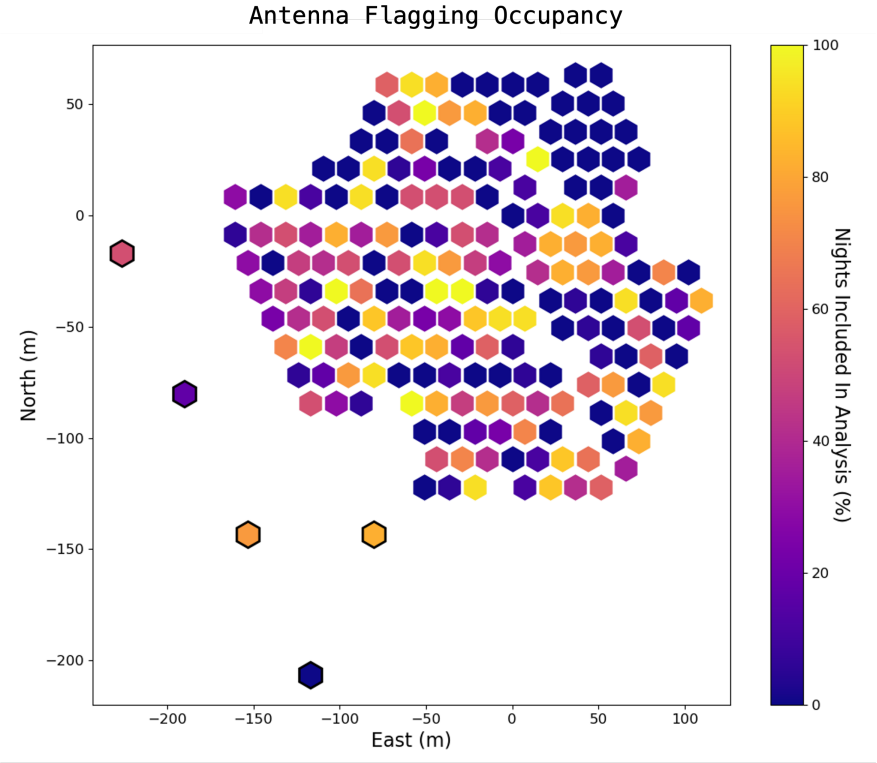


Figure 5.6: Antenna map showing how many nights of the analysis set each antenna is included in. Purple represents antennas that were always flagged, and yellow represents antennas included in all nights. The outrigger antennas, shown outlined in black, are plotted with their distance from the array core not to scale for the sake of readability.

Figure 5.6 shows a map of all antennas recording data at the time of our analysis set. Here each antenna is colored to show the fraction of analysis nights that the antenna remained unflagged. We see that antenna flagging is highly inconsistent between nights, and while each night has at least sixty unflagged antennas, there are only seven antennas that are considered usable for all nights of the analysis. In order to increase the number of antennas available for each night of analysis, we perform a new flagging procedure that treats each polarization independently. This approach is sub-optimal in terms of systematics mitigation, as we suspect that any systematic affecting one

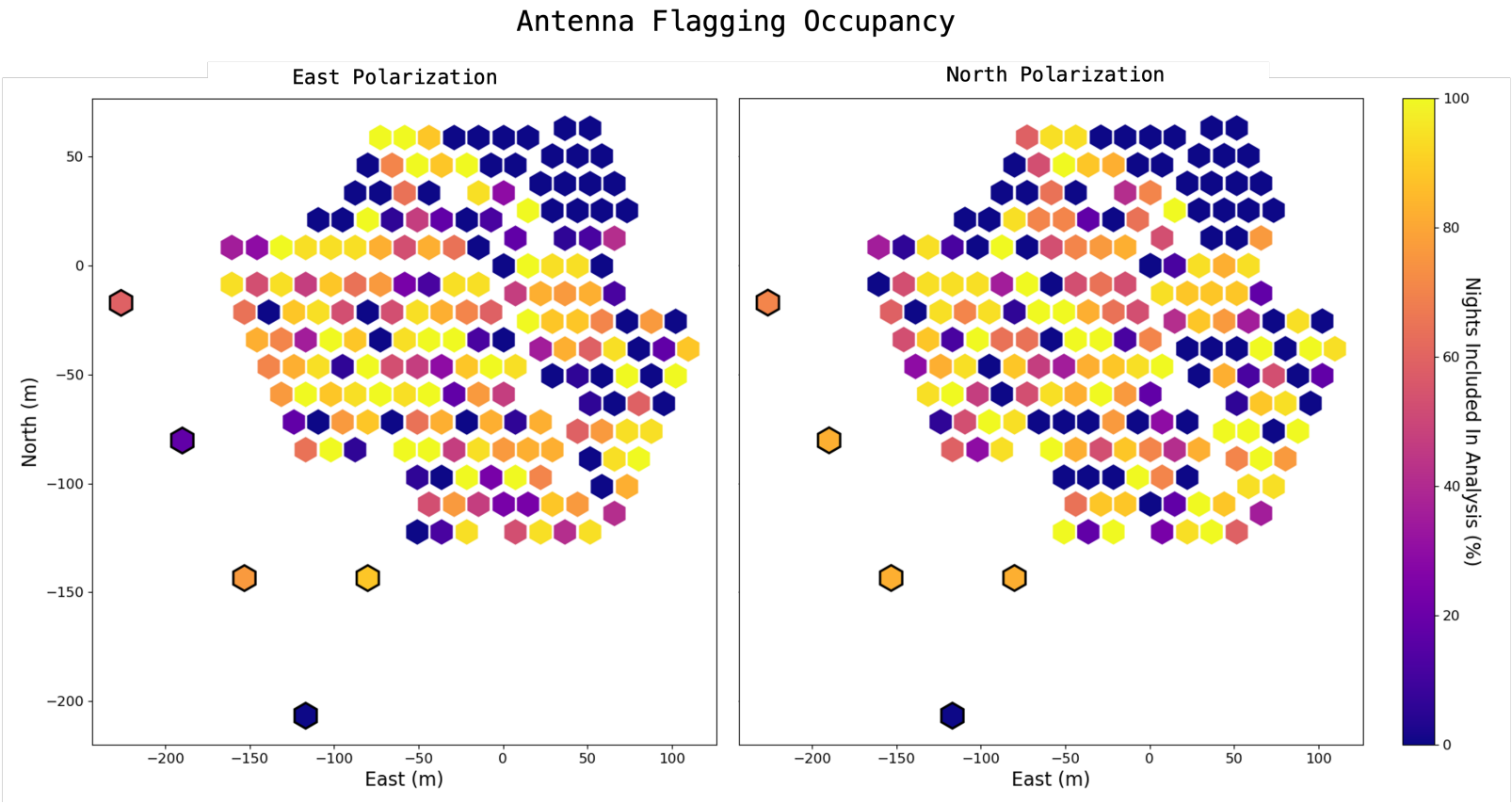


Figure 5.7: Antenna map showing how many nights of the analysis set each antenna-polarization is included in. In this case, flagging was done separately between the two polarizations. Purple represents antennas that were always flagged, and yellow represents antennas included in all nights. The outrigger antennas, shown outlined in black, are plotted with their distance from the array core not to scale for the sake of readability.

polarization is likely to contribute on some level to the other polarization as well. However, we require a large number of baselines to perform a succesfful calibration and have sufficient uv coverage to perfrom a power spectrum measurement. Figure 5.7 shows another flag occupancy antenna map in the case where flagging was performed per-polarization. This figure shows a substantially higher antenna utilization rate across all nights, and confirms that a per-polarization flagging regime significantly increases the number of available baselines.

To better quantify this improvement we refer to Figure 5.8, which shows a histogram of the length of all baselines available to analysis for every night in the data set. We plot a distribution for the case where polarizations are flagged together, and also a distribution for each polarization in the case they are flagged separately. We see a substantial increase in available baselines when we flag each polarization separately, and conclude that the risk of systematic leakage between polarizations is worth the increase in data volume and available baselines.

5.5 Final Data Set

With the above data selections in place we can now present our final analysis data set. Each night of data averages 224.6 unflagged time observations, each spanning 9.8 seconds, between LSTs 0.92 and 1.56 for a total of 36.7 minutes of average observation time. The nights also average 163.3 unflagged frequency bins spanning between 154.6MHz and 175.0MHz. Finally, the nights have an average of 100 unflagged antennas in the North polarization and 97 unflagged antennas in the East polarization, corresponding to an average of 4987 and 4727 baselines in each polarization, respectively. Figure 5.9 shows the exact visibility counts for each night. The first row shows the number of unflagged visibilities in time and frequency for any given antenna-polarization. The second row shows the number of unflagged baselines in each polarization, and the bottom row shows the total number of unflagged visibilities for each night and polarization.

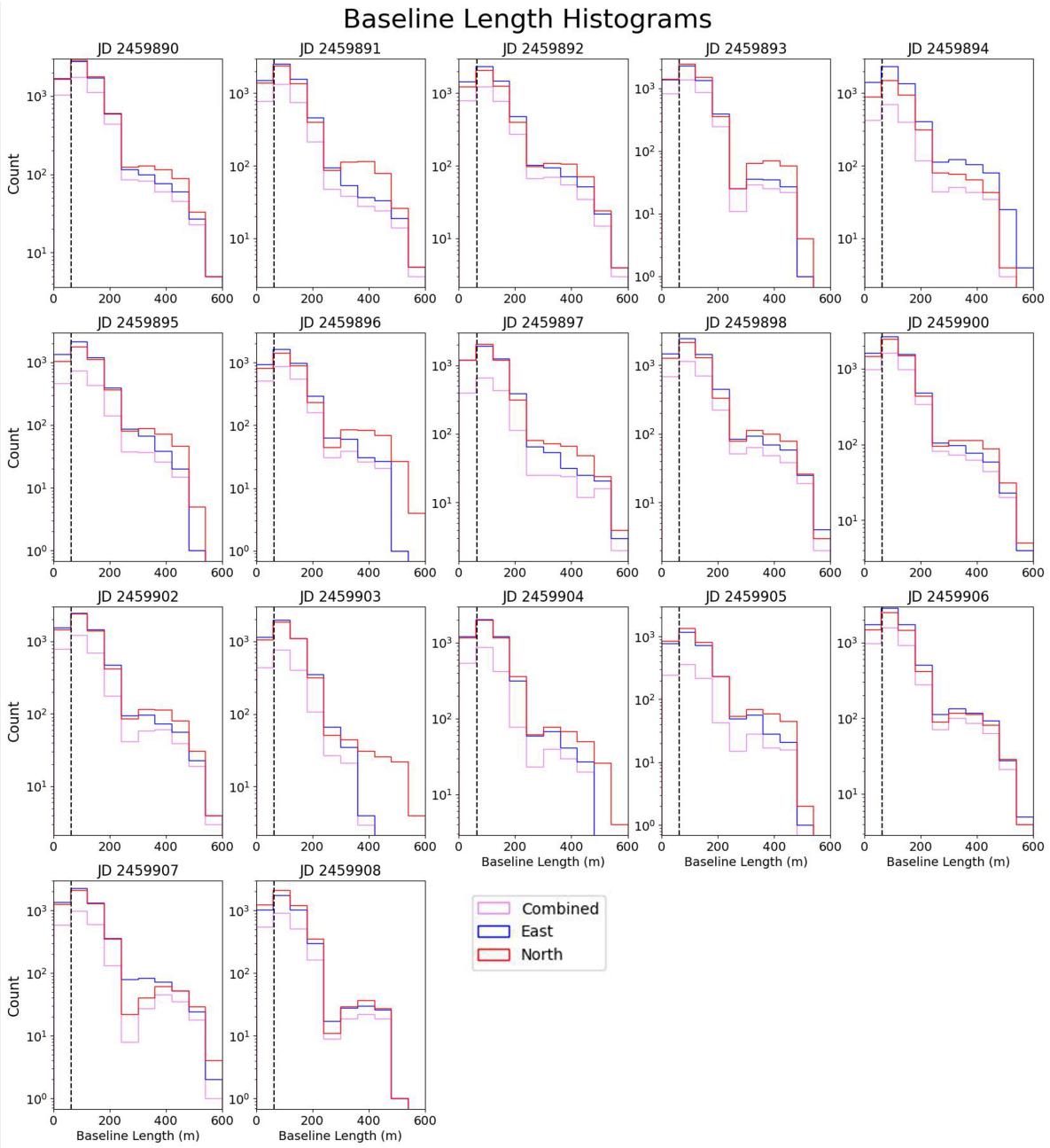


Figure 5.8: Baseline length histograms for all nights in the analysis data set. The East (blue) and North (red) histograms separately show the distribution of baselines that remain in the data after antenna flagging. Pink shows unflagged baselines in the case where per-polarization flags are combined. The black dashed line represents the baseline length cut applied to calibration, where shorter baselines are excluded from calibration.

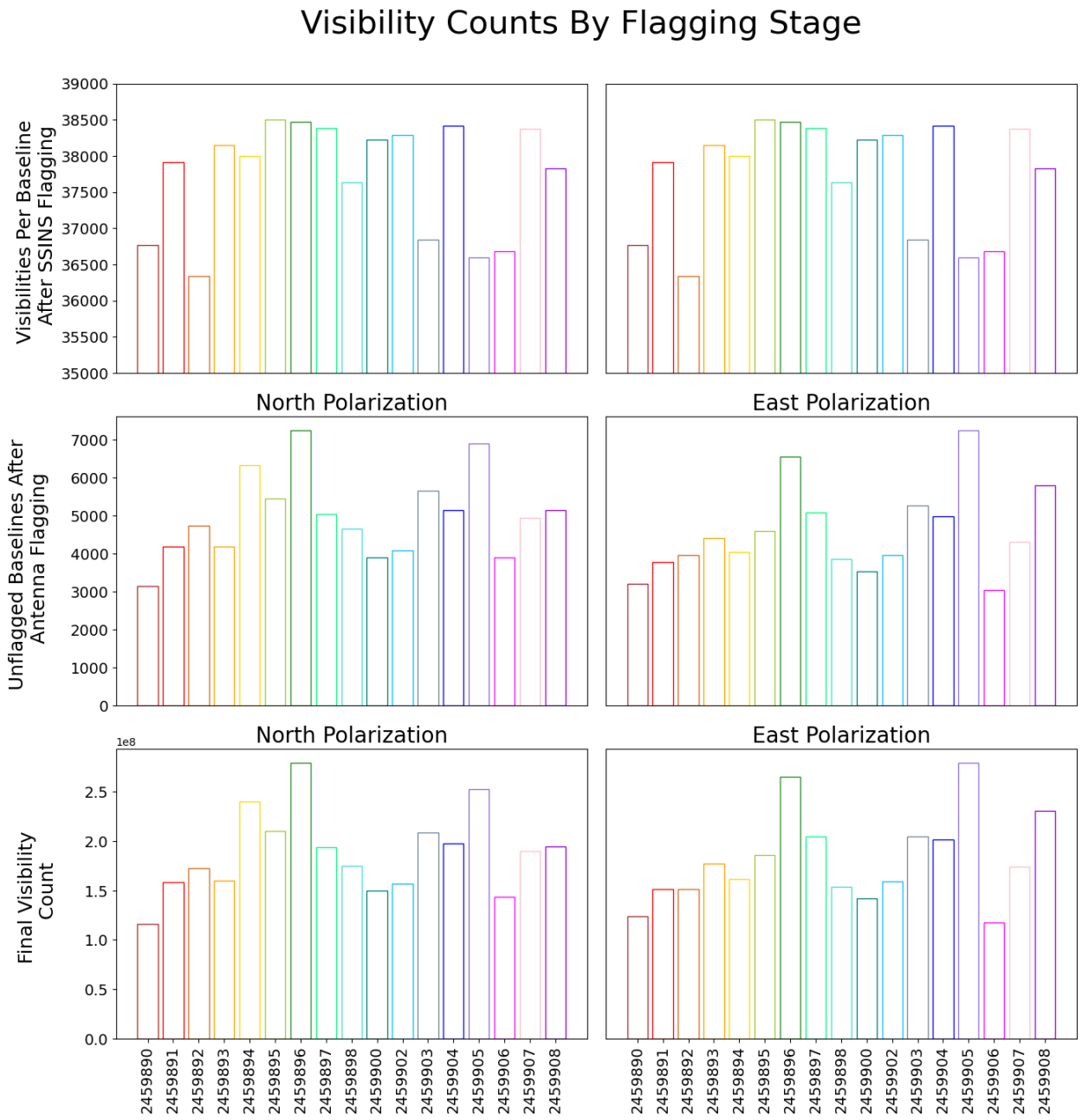


Figure 5.9: Bar chart of visibility counts for each night in the data set. The top row shows the number of visibilities per baseline after SSINS flagging - this is the same for both polarizations, and counts all times and frequencies in the nightly data. The second row shows how many baselines are unflagged for each night and polarization. The third row combines these two, showing the total number of unflagged visibilities for each night.

Chapter 6

CALIBRATION

In this chapter we will begin by providing a more explicit definition of our visibility measurement. Then, we will outline the basics of instrumental calibration, covering both redundant and sky-based calibration techniques. Finally, we will present the specifics of the calibration scheme used in this analysis and show results validating our final set of calibration solutions.

6.1 The Visibility Measurement

First, we recall from Section 1.3 that a visibility measurement samples the spatial coherence function of the electric field on the ground. We also recall that we have defined the uv plane, where u and v are spatial coordinates in the instrument plane, measured in units of wavelength, representing the displacement between two antennas. To begin, we can utilize the Van Cittert-Zernike theorem to define the coherence function on the ground that our visibilities will sample as it relates to the intensity on the sky:

$$\Gamma_{ij}(u,v) = \iint I(l,m) e^{-2\pi i(ul+vm)} dl\, dm, \tag{6.1}$$

where (l,m) are angular cosine coordinates on the sky plane. We see from this equation that the spatial coherence function in the uv plane is the Fourier transform of the intensity on the sky. If we were able to make continuous measurements of the entire uv plane this equivalence would be exactly true for our visibility measurements, but in practice our instrumental limitations mean that we have a non-continuous and incomplete sampling function. Therefore, we must calculate the instrumental beam, $B(l,m)$, which measures the sensitivity of our instrument as a function of position on the sky. Note that each antenna element is actually integrating the electric field over its collecting area, which means that baselines cannot truly be represented as sampling a point, but rather have a complex spatial response in the uv plane. In practice, this is estimated by doing EM simulations of individual antenna elements and combining that with knowledge of antenna positions

to create a representation of the entire instrument response in the uv plane, and making use of the Fourier relationship between the sky and instrument planes to get the instrument response on the sky, $B(l, m)$. Finally, considering that all instruments have chromatic features that effect the signal, we can now write down our visibility measurements, which are sampling the distribution in Equation 6.1, as

$$V_{ij}(u, v, f) = \iint I(l, m, f)B(l, m, f)A(f)e^{-2\pi i(ul+vm)}dl\ dm, \tag{6.2}$$

where u and v are the coordinates of baseline ij, and $A(f)$ represents the instrument's inherent frequency response. Here we have also indicated that both the sky intensity and the instrumental beam are functions of frequency.

6.2 Calibration Formalism

Once we have successfully taken visibility measurements we must calibrate our data. Most fundamentally, calibration is the process by which we try to reconstruct the true sky signal by accounting for the effects of the instrumental response on the observed data. In standard calibration regimes one can consider every signal path as having a multiplicative gain on the signal. This is called direction-independent calibration, in which each antenna has a single corrective gain per time, frequency, and polarization. It is also possible to perform direction-dependent calibration, which accounts for errors in the antenna response model and can compute distinct gains for different directions on the sky. Direction-dependent calibration is outside the scope of this thesis, and we will focus only on describing a per-frequency direction-independent calibration.

In the case where each antenna and polarization have an independent signal chain (as is true for HERA), we can consider any antenna-polarization i as measuring an electrical signal s given by

$$s_i^{\text{observed}} = g_i s_i^{\text{true}} + n_i \tag{6.3}$$

where s_i^{observed} is the observed signal, s_i^{true} is the true signal, g_i is a complex frequency-dependent instrumental gain, and n_i is a Gaussian-distributed thermal noise term. Note that when we refer to the gains as per-antenna this is including every electronic element along that antenna's signal

chain. A visibility is then the cross-correlation of two such signals, or $V_{ij} = \langle s_i s_j^* \rangle_t$. We can expand and simplify to write the visibility as

$$V_{ij}^{\text{observed}} = g_i g_j^* V_{ij}^{\text{true}} + n_{ij} \tag{6.4}$$

where assumptions that an antenna's noise is uncorrelated with other antennas and is uncorrelated with the sky signal means that the cross-terms $\langle s_i n_j \rangle$ fall to zero.[1] As with the visibilities, we calculate a complex gain per antenna, polarization, frequency, and observation. The gains are assumed to be slowly varying in time, and in practice are often calculated over a longer integration window than the visibilities.

The next sections will outline the two primary techniques used to solve for the per-antenna gains in Equation 6.4: redundant calibration and sky-based calibration. Recent efforts have found effective techniques for unifying these calibration frameworks to optimize each of their strengths (Byrne et al., 2021), but we will not discuss or utilize those techniques in this thesis.

6.2.1 Redundant Calibration

Referring to Equation 6.2 we see that any baseline with the same antenna separation, and therefore the same coordinates in the uv plane and same beam shape, will in principle measure the exact same visibilities. We refer to any set of baselines that have the same uv coordinates as redundant baselines. Therefore, we can in theory ascribe any discrepancies in the visibility measurements of redundant baselines to electrical instrumental effects and we can utilize this to put constraints on the calibration equation (Wieringa, 1992; Liu et al., 2010). Assuming that a large fraction of visibilities come from redundant baselines this approach is able to constrain the majority of calibration parameters using just the measured visibilities. However, there are four degenerate parameters per polarization and per frequency that cannot be constrained through redundant calibration, and therefore some knowledge of the sky is still required (Dillon et al., 2018). Our current sky models have substantial limitations, and therefore the limited sky knowledge required for a redundant calibration approach is extremely interesting. However, redundant calibration has its own failures. For

[1] The excess correlation systematic outlined in Chapter 4 caused this assumption to fail. This is a big part of why mitigating that systematic was so essential.

example, even slight errors in antenna placement will cause baselines to not be truly redundant, and therefore can cause the underlying assumptions of redundant calibration to fail. Additionally, even in the case of perfect redundancy, a sky model cannot adequately constrain all degeneracies (Byrne et al., 2019).

HERA was specifically designed with a highly redundant antenna layout to make it suitable for redundant calibration techniques. The primary analysis pipeline used within the HERA collaboration therefore uses redundant calibration as their primary calibration technique. More detailed explanation of this technique in theory and in practice as it pertains to HERA can be found in (Dillon et al., 2020a). For the purposes of this thesis we are pursuing an imaging-based power spectrum, which inherently requires extensive use of a sky model and therefore favors a sky-based calibration approach. It is important to note that each of these techniques have their own advantages and their relative performance varies depending on the instrument and the foreground environment being observed (Li et al., 2018; Zhang et al., 2020). For the remainder of this work we will exclusively use sky-based calibration techniques, which will be explained in the next section.

6.2.2 Sky-Based Calibration

With sky-based calibration we use a-priori knowledge of the sky rather than redundancy to solve for the per-antenna gains. In general, sky-based calibration is performed by using a catalog of known sky sources to create model visibilities, substituting those in place of V_{ij}^{true} in Equation 6.4, and using a χ-squared minimization to solve for the gains. One can calculate model visibilities using equation 6.2, where a source catalog is used to estimate I, a model instrument beam is used for B, and the known instrumental bandpass response is used for A. We do not yet have a detailed model of diffuse sky structures, and therefore most sky-based calibration schemes rely either primarily or entirely on point source catalogs to create the model visibilities. Given the Fourier relationship between the uv plane and the sky, longer baselines are more sensitive to small spatial structures on the sky while short baselines are more sensitive to diffuse emission. Therefore, where redundant calibration relies on a regularly gridded antenna layout to be successful, sky-based calibration depends on a sufficient set of long baselines. This is one of the motivators behind the outriggers in HERA's antenna layout, as their substantial distance from the core array introduces

a set of much longer baselines that are helpful to calibration.

The details of sky-based calibration can be implemented in a number of ways. In this work we utilize the software package `Fast Holographic Deconvolution` (FHD) to perform both calibration and imaging (Sullivan et al., 2012). In the next section we will review several calibration choices specific to our analysis and validate our final calibration solutions. More details about the specifics of the `FHD` calibration scheme can be found in Barry et al. (2019a).

6.3 Calibration with HERA

In this section we will present the details of our calibration scheme and specific choices we make within the `FHD` framework. We begin by describing the sky model (6.3.1) and beam model (6.3.2) that we use. Then, we describe the specifics of our calibration pipeline and show the raw calibration results (6.3.3) and introduce a polynomial fitting scheme (6.3.4).

6.3.1 Sky Model

We use the GaLactic and Extragalactic All-sky MWA survey (GLEAM, Wayth et al. (2015)) as our primary source catalog for calibration. The GLEAM catalog includes flux density measurements for 307,455 radio sources with 20 different flux measurements between 72-231MHz (Wayth et al., 2015). It includes sources over a majority of the sky, but does not include measurements in the galactic plane or the Magellanic Clouds. GLEAM also omits some important bright sources, and so we supplement the GLEAM catalog with model information for nine additional bright sources (Byrne, 2021). In this analysis we do not include any extended source information and treat all sources as point sources.

6.3.2 Beam Model

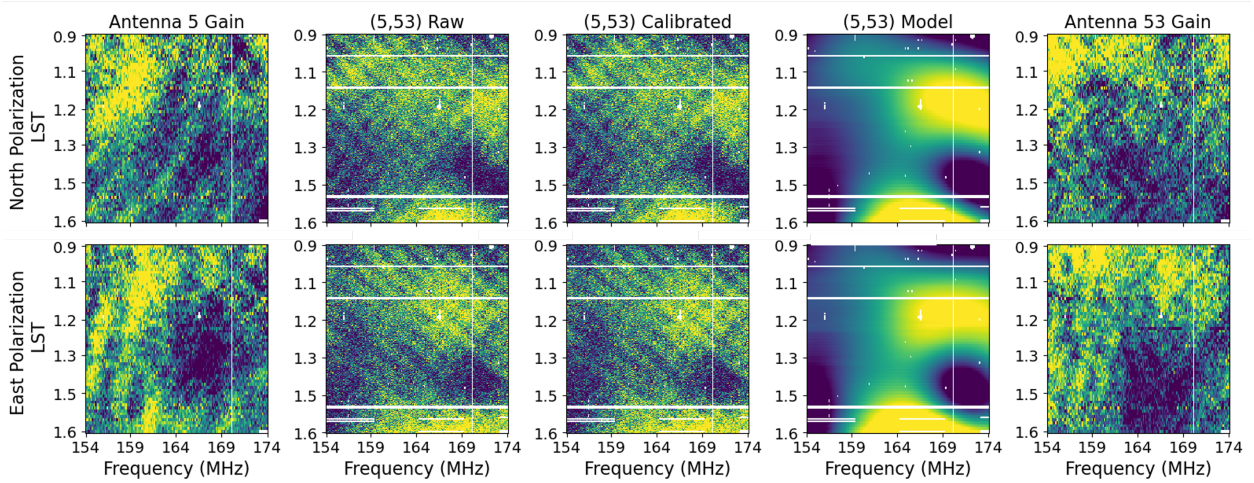
In order to create the model visibilities required for sky-based calibration we need a high precision instrumental beam model. There are several effective methods for creating a beam model, including using a known satellite signal to estimate the instrument response (Line et al., 2018; Neben et al., 2016), using holography techniques with a very bright point source (Berger et al., 2016), utilizing Unmanned Aerial Vehicles (UAVs) (Jacobs et al., 2017; de Lera Acedo et al., 2017), and using

drift–scan widefield observations (Pober et al., 2012; Nunhokee et al., 2020). In this work we use the HERA beam model presented in Fagnoni et al. (2021), which calculates the predicted instrument response using detailed electromagnetic simulations of the antenna element and Vivaldi feed. This simulated beam is validated using in-situ techniques in Nunhokee et al. (2020).

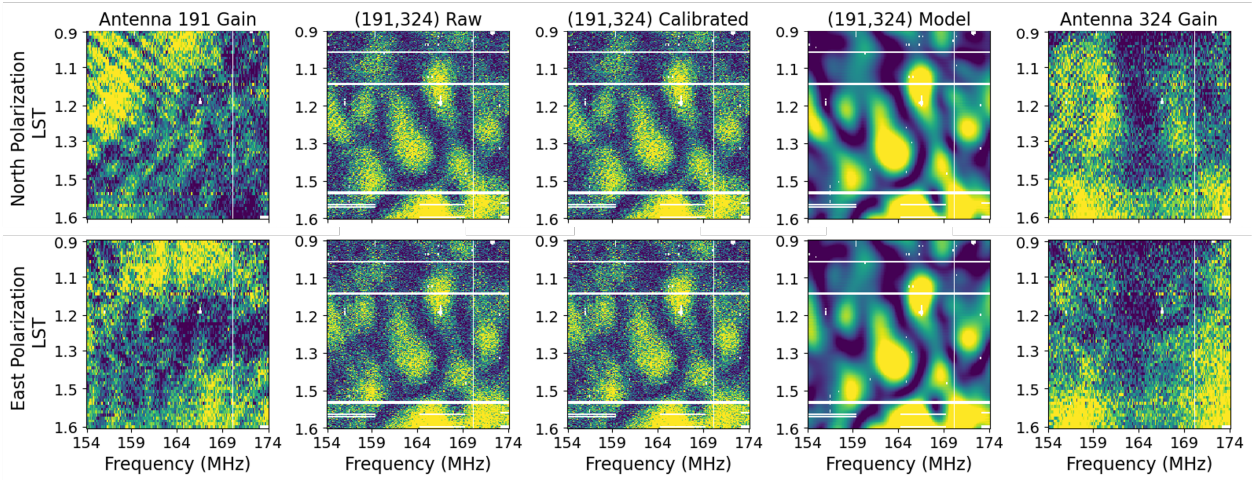
6.3.3 Raw Calibration Solutions

With our sky catalog and beam model established we can now use `FHD` to perform calibration. Our calibration solves for one complex gain per antenna, frequency, polarization, and observation set. For this analysis we calibrate data in sets of four observations, or 39 seconds, and only solve for one gain solution per set. We then apply that one gain solution to all four time stamps in the set. Solving for one gain across multiple times allows for better constraint on our calibration equation and is considered accurate on short timescales relative to the rate of rotation of the sky. We also assume that our instrument's response changes slowly with time. This assumption is generally accepted for most instruments, but remains a concern with HERA given the severity of temporal structure we observe in the data. Therefore, we chose a calibration time scale shorter than that required by sky rotation so as to mitigate the effects of temporal structure as much as possible. We also choose to only use baselines longer than 35 wavelengths for calibration, as un-modelled diffuse emission becomes important on shorter baselines Byrne et al. (2021).

Figure 6.1 shows amplitude waterfall plots of the visibilities and gain solutions for two baselines. Panel (a) shows a 100 meter baseline, which is important to validate because, as shown in Figure 5.8, 100 meters (or 56λ) is our most common baseline length and therefore is where we will have the most sensitivity in our power spectrum calculation. We also want to emphasize a long baseline that includes an outrigger antenna, shown in panel (b), as these baselines' finer resolution on the sky makes them particularly essential to our calibration scheme. As expected, the longer baselines have more structure in the visibilities, and their structure is more precisely matched by the model visibilities. However, in both cases we see that the model visibilities successfully reproduce the overall structure of the raw visibilities. On the shorter baseline, we do see a cross-hatched structure in the raw visibilities that is not in the model. We believe this is due to one or more bright sources near the horizon interacting unfavorably with the sidelobes of the beam in a way that is not



(a) 100 Meter Baseline



(b) 475 Meter Baseline

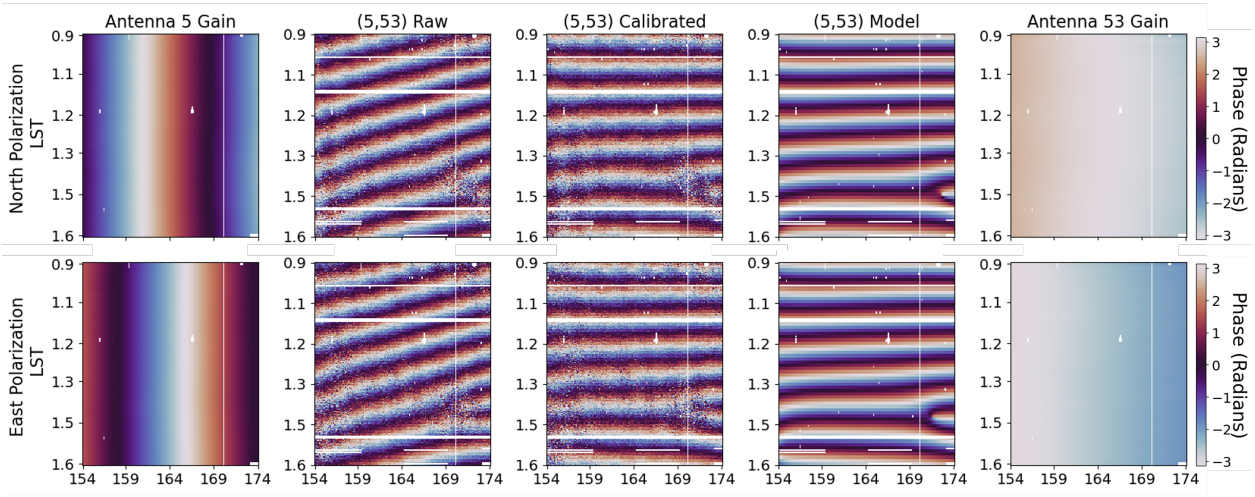
Figure 6.1: Amplitude waterfall plots of for two baselines 100 meters (a) and 475 meters (b) long, showing the North (top) and East (bottom) polarizations. The leftmost and rightmost columns show the gain solutions for each of the baseline's constituent antennas. The middle columns show the raw, calibrated, and model visibilities for that baseline, respectively. The visibility flags are over-plotted in white.

captured by our model. While we only see this structure in the cross-correlations of the raw data, we note that every antenna's gains are calculated with input from all baselines above our baseline length cut. Therefore, we see evidence of this structure leaking into the gains, as is particularly clear for antenna 191 (bottom left panel). As this structure is due to a real interaction between the sky and our instrument it should not be carried into the gain solutions.

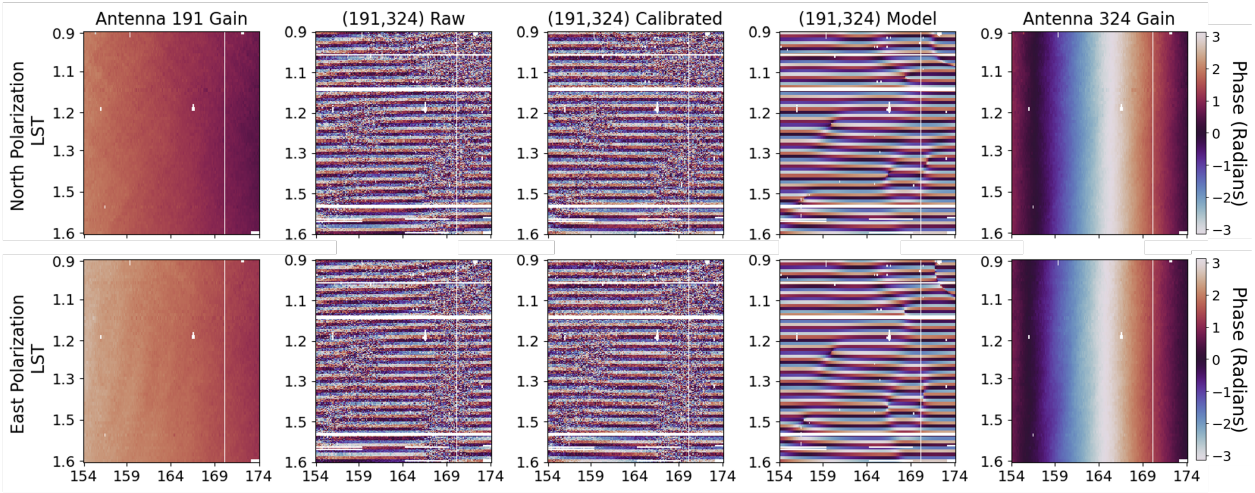
In figure 6.2 we show gains and visibilities for the same two baselines as the previous figure, but now we are plotting phase instead of amplitude. In both the long and short baselines we see that the phase structure is well matched between the calibrated and model visibilities. On the 100 meter baseline we see that there was a slope to the phase ramp in the raw data that has been corrected by calibration. We also note that the phase wraps much faster in time on longer baselines, which is expected. On the 475 meter baseline we see that the cross-correlation phases appear striped - this is due to the interaction between the rapid phase wrapping and the 39 second gain interval. Importantly, the phase of the gain solutions is stable in time. HERA antennas have consistently shown two distinct characters of gain phases; they either have a nearly flat slope with frequency or they have a steep ramp that causes a phase wrap in frequency. We see both cases in Figure 6.2, where antennas 191 and 53 show a flat phase ramp and antennas 5 and 324 show a shallow one.

6.3.4 Fit Calibration Solutions

Thus far we have computed gains independently per-frequency, but we must be careful that our gains don't introduce unwanted or spectral structure into our calibrated visibilities. Barry et al. (2016) found that sources on the sky that are captured by the raw data but not included in the model will introduce spectral structure through the gains that is substantial enough to obscure the expected EoR signal. Additionally, we saw from Figure 6.1 that our raw data exhibit a frequency ripple that wanders with time and is leaking from the cross-correlations into the per-antenna gains. We found through careful inspection that this spectral ripple consistently drifts in time the same way at the same LSTs across different nights. We also found that the ripple is strong at the beginning of the night, fades away, and then reappears at the end of the night oriented the opposite direction that it was at the beginning. This behavior is much more consistent with an interaction with un-modeled diffuse emission than with an analog or digital systematic, and therefore we treat



(a) 100 Meter Baseline



(b) 475 Meter Baseline

Figure 6.2: Phase waterfall plots of for two baselines 100 meters (a) and 475 meters (b) long, showing the North (top) and East (bottom) polarizations. The leftmost and rightmost columns show the gain solutions for each of the baseline's constituent antennas. The middle columns show the raw, calibrated, and model visibilities for that baseline, respectively. The visibility flags are over-plotted in white.

this as real sky structure.

To ensure that our gains are not contaminated with spurious spectral structure we choose to fit a polynomial to our per-frequency gain solutions and apply those fitted gains to our data rather than the raw gains. We apply a separate polynomial fit to the amplitude and phase terms of our complex gains. We tested fitting our gain amplitude solutions with polynomials up to order 10, but found that beyond the quadratic order the spectral structure of the fit was not stable in time. To best separate instrumental terms from the diffuse emission contamination we choose to fit our amplitude only up to a second order polynomial. A first order polynomial (amplitude and slope terms) is sufficient for fitting the phase solutions.

Figure 6.3 shows an example of this polynomial gain fitting for antennas 5 and 53, which comprise the same baseline shown in Figure 6.1a. The first row shows the auto-correlations for both polarizations of each antenna, taking a slice at the first time from Figure 6.1a. The second row shows the amplitude of both the raw gains and the applied polynomial fit, with the residual between the two beneath them. Finally, the bottom row shows the phase of the raw and fit gains, again with the residual underneath. We see that the phase ramp is exceptionally well fit and the residuals are primarily noise, which indicates that we have successfully found the appropriate phase ramp. The residual of the amplitudes shows some remaining structure beyond the noise, particularly in antenna 53. This structure is indicative of the cross-hatch structure that we wanted to eliminate from the gains, and therefore we consider this a successful amplitude fit.

While the calibration solutions and fitting in Figure 6.3 appear successful, these only reflect forty seconds of data for two antennas. To build confidence in our calibration scheme and to further validate our antenna selection from Section 5.4 it is important to examine the calibration solutions and fitting across all antennas. Figures 6.4 and 6.5 show the raw gain solutions with the fit gains overplotted for the North and East polarizations, respectively, of every unflagged antenna in the first night of our analysis set. Each panel of these plots represents forty seconds of data for an individual antenna, and the scale is fixed between all panels to show relative gain amplitude between antennas. In all cases we see that the polynomial fitting is accurately capturing the overall gain amplitude and general shape, while not capturing any sub-structure or rippling that we have identified as non-instrumental.

We note that while some gains are fairly smooth even before the fitting, some have more am-

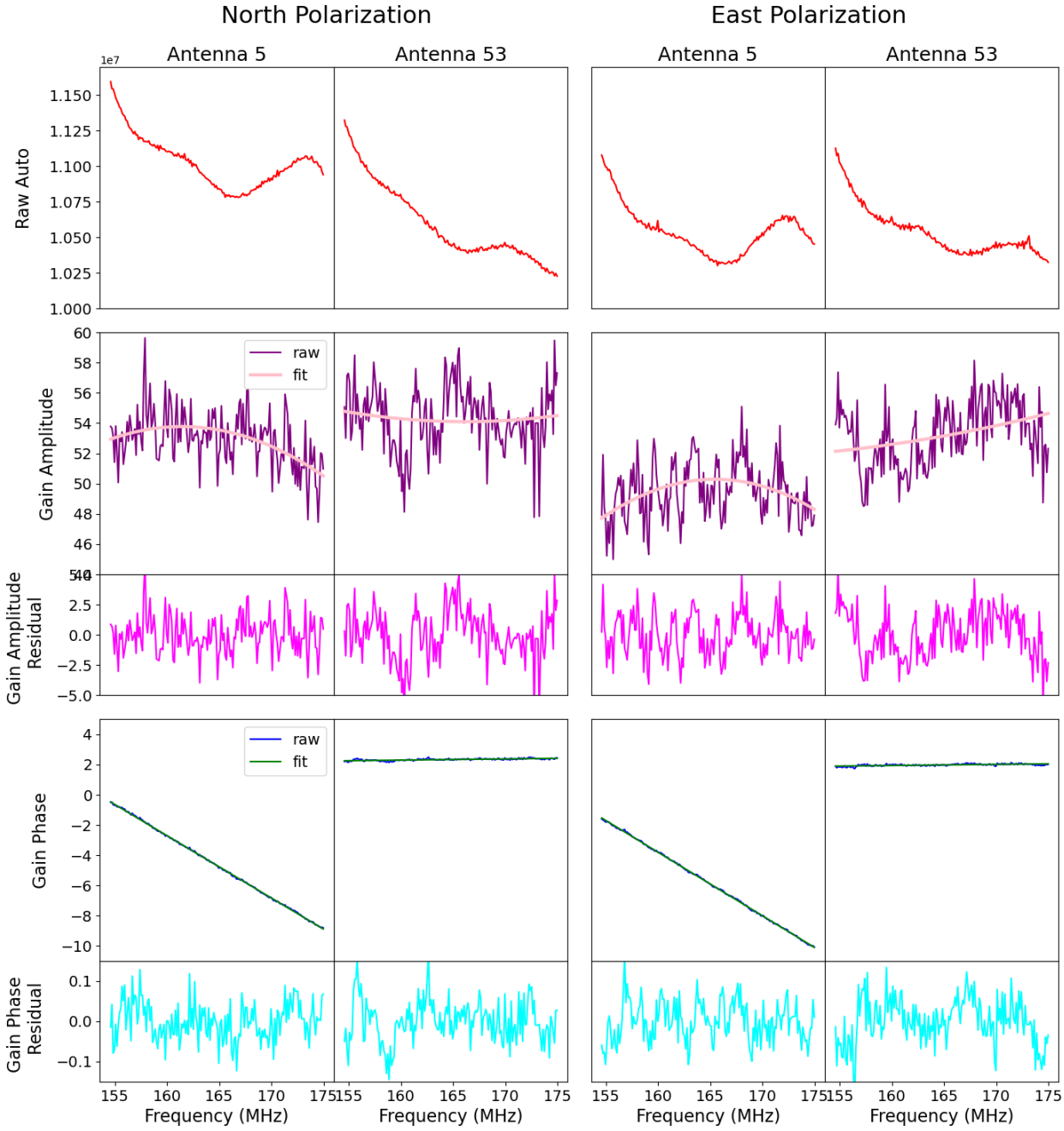


Figure 6.3: Single observation visibility slices for antennas 5 and 53 and for both the North (left) and East (right) polarizations. The top row contains the raw auto-correlation spectra, which is not included in calibration, but rather is shown to emphasize the relative scale and shape of the per-antenna gain solutions. (Caption continued.)

Figure 6.3: (Previous page.) The middle two rows show the raw gain amplitude (purple), the polynomial gain amplitude fit (pink), and the residual between the two beneath them (magenta). The bottom two rows show the raw gain (blue), the gain phase fit (green) and the residual between the two beneath them (cyan). These two antennas correspond to the upper cross-correlations shown in Figure 6.1.

plitude fluctuation and appear to have a rippling structure. We have identified that antennas with these more ripply gains are those whose cross-correlations are more likely to have a strong ripple or cross-hatched structure. We have seen that antennas who participate in fewer long baselines are more likely to show this feature, although it remains an open question what the true cause is. In extreme cases we may choose to remove an antenna from the data set at this stage due to highly structured gains, but this is rare. It is worthwhile to emphasize that while the gains can help us identify the antennas that most exhibit this feature we do *not* want it captured by the gains as we do not believe it to be a gain-like electronic effect of the instrument. This helps build confidence in our polynomial fitting.

Next, Figure 6.6 shows a histogram of the polynomial amplitude fitting parameters for each polynomial order and for each night in our final analysis data set. We see that order zero, which represents the overall gain amplitude, has a notable tail in its distribution. This is not concerning as HERA's antennas have significant variability in their power level that we rely on the gains to correct for. The narrower distributions in the first and second order terms is also expected, as a tail or outlier in these cases would indicate that some antennas have particularly abnormal bandpass shapes. While this does happen in HERA, we have already selected out any such antennas. Similarly, Figure 6.7 shows the fit parameters for the first and second order phase terms. Again, we see a relatively narrow distribution for both orders and both polarizations. We require our visibilities to have a zero-mean phase term, which is confirmed by the zero-mean distribution of the amplitude term. We see no notable outliers among any of the analysis nights, which builds confidence that all nights in our analysis set are performing nominally.

In the next chapter we will briefly overview the fundamentals of imaging, particularly as they

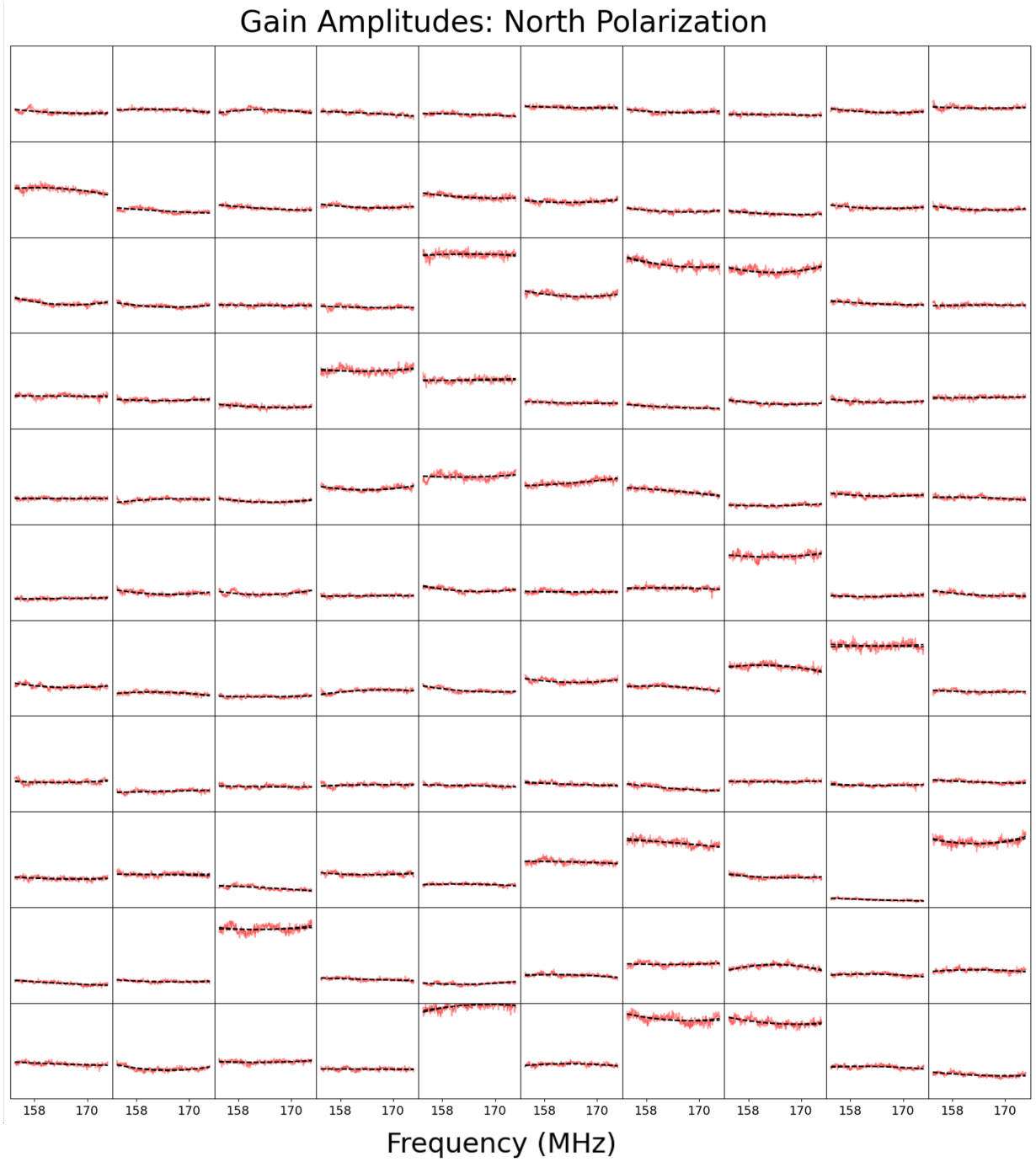


Figure 6.4: Per-frequency gain solutions for all antennas in the North polarization. The raw gain amplitudes are shown in red, and the amplitude of the polynomial fit gains is overplotted in black. Each panel is an individual antenna.

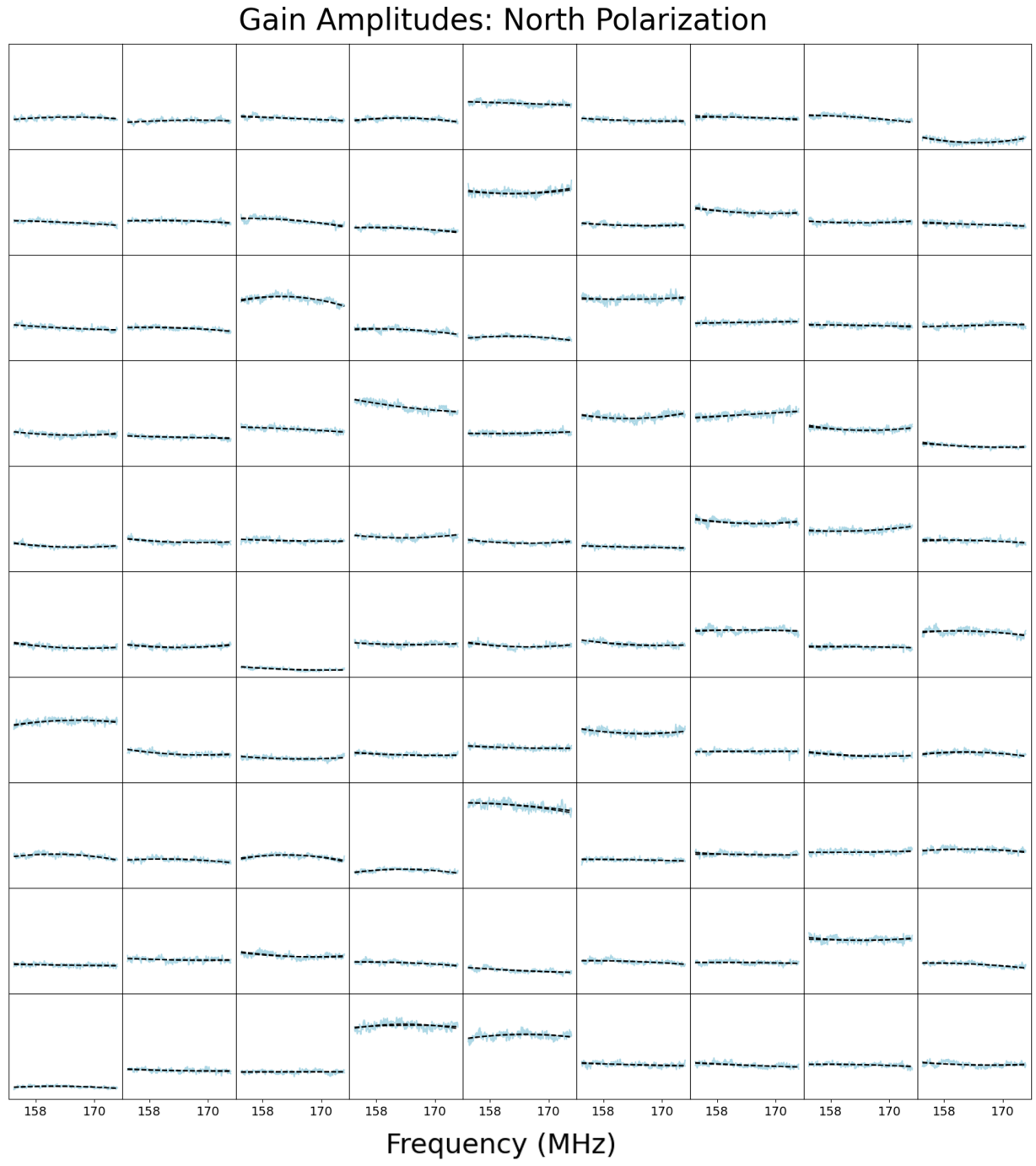


Figure 6.5: Per-frequency gain solutions for all antennas in the East polarization. The raw gain amplitudes are shown in blue, and the amplitude of the polynomial fit gains is overplotted in black. Each panel is an individual antenna.

pertain to the power spectrum pipeline, and show some example images from HERA. Then, in Chapter 8 we will outline the essentials of power spectrum calculation and present our results.

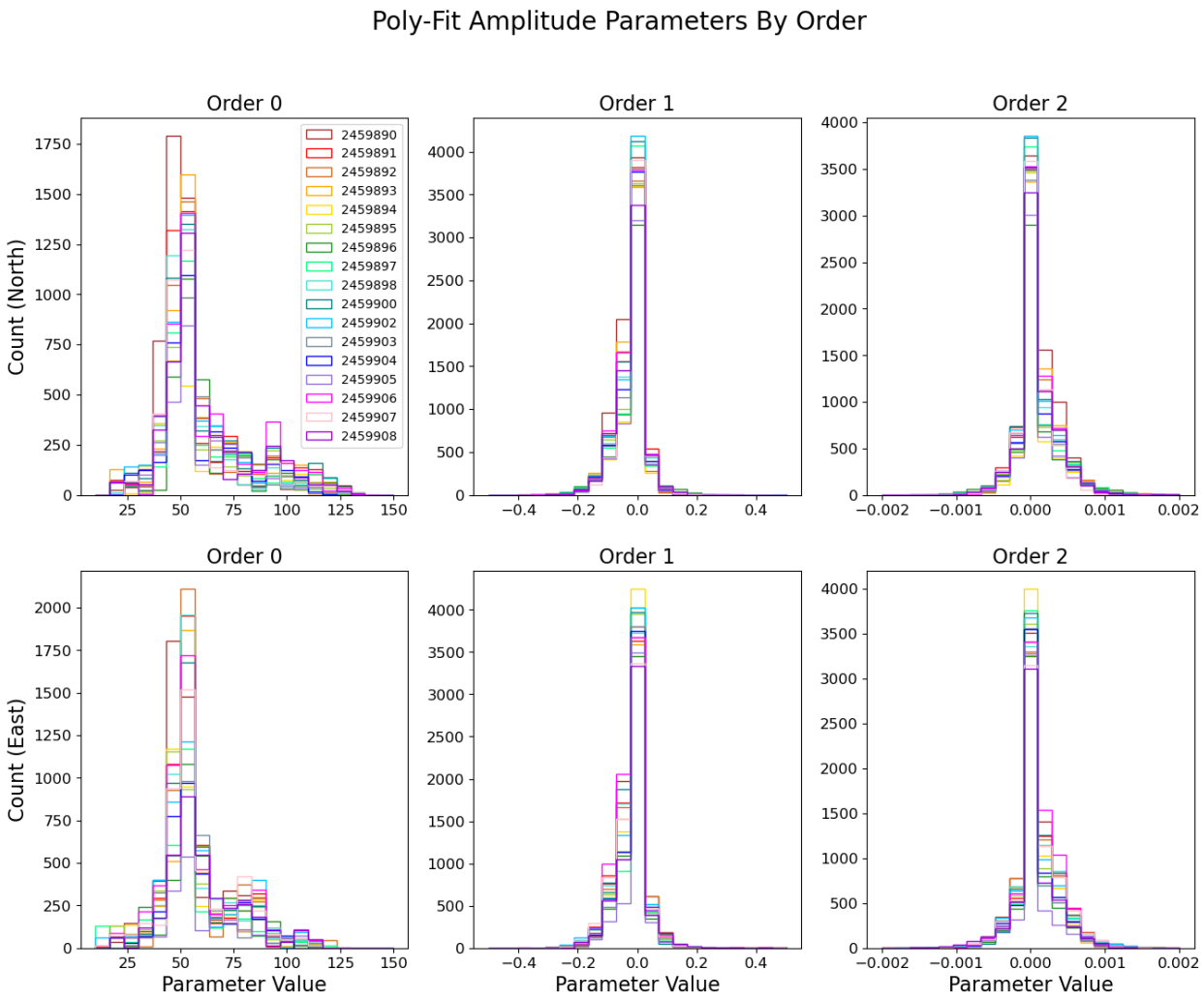


Figure 6.6: Histograms of the polynomial fitting parameters applied to the amplitude of the gain solutions during calibration for the North polarization (top) and East polarization (bottom). Order zero corresponds to an amplitude term, order one to a slope term, and order two to a square term. There is one histogram per night of data to confirm that all nights are behaving similarly and that there are no outliers.

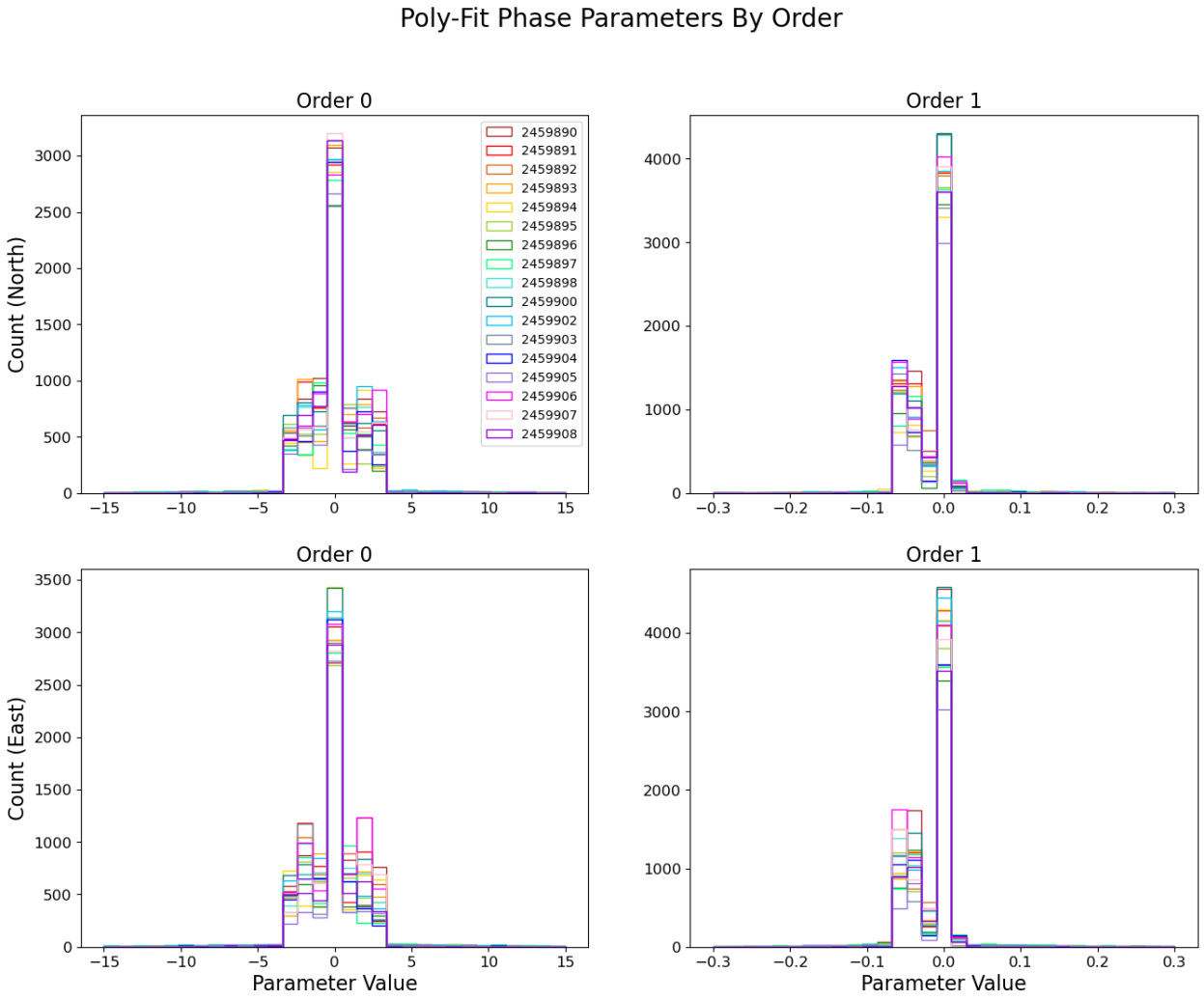


Figure 6.7: Histograms of the polynomial fitting parameters applied to the phase of the gain solutions during calibration for the North polarization (top) and East polarization (bottom). Order zero corresponds to an amplitude term and order one to a slope term. There is one histogram per night of data to confirm that all nights are behaving similarly and that there are no outliers.

Chapter 7

IMAGING

In the next chapter we will present power spectrum results created using an imaging-based analysis pipeline, and therefore it is worthwhile to provide some background on the fundamentals of our imaging procedure. A more detailed description of interferometric imaging in general can be found in Taylor et al. (1999) or Thompson et al. (2017), and for more details on the specifics of our pipeline we refer to Sullivan et al. (2012) and Barry et al. (2019a). In Section 7.1 we will introduce the fundamentals of interferometric imaging and in Section 7.2 we will present some results of imaging with HERA.

7.1 Imaging Fundamentals

We begin by recalling our visibility measurement equation from the previous chapter:

$$V(u,v,f) = \iint I(l,m,f)B(l,m,f)e^{-2\pi i(ul+vm)}dl\,dm, \tag{7.1}$$

where V is the full set of measured visibilities, u and v are the set of all baseline coordinates, I is the intensity on the sky, and B is the instrumental beam. We have dropped the instrumental bandpass term $A(f)$ in this case because we are now working with calibrated visibilities. With this representation of a visibility, we can consider 'imaging' to be the process of reconstructing an observed sky I^{obs} given a visibility measurement and known instrumental beam B. In practice an individual beam model can be provided for each antenna (making B a per-baseline term), but we will only consider the case where all antennas have the same beam model, as is the case with our analysis. Utilizing the Fourier relationship between the visibility and sky planes, we can rewrite Equation 7.1 as:

$$I(l,m,f)B(l,m,f) = \iint V(u,v)e^{2\pi i(ul+vm)}du\,dv. \tag{7.2}$$

In reality our visibilities are not a continuous function but rather a discrete set of measurements, and so we must begin by gridding our data to the uv plane. Gridding is the process by which we project our set of visibility measurements onto a uniform set of points in the uv plane. This involves both placing the baseline coordinates on a fixed grid and applying a kernel to smear each visibility about its baseline center on the uv plane. The gridding kernel is necessary to accurately represent the spatial antenna response (Morales and Matejek, 2009). Once gridded, we can take a Discrete Fourier Transform (DFT) of our visibilities to generate a set of measurements in the sky plane.

When we make images with our measured visibilities we typically call these 'dirty' images, which refers to the fact that they have not been deconvolved (sources have not been identified and removed). We can also create images in the same way using the model visibilities described in the previous chapter, which we call model images. It is also useful to look at the difference between the dirty and model images, which we call the residual.

7.2 Imaging with HERA

While working in image space is useful for power spectrum estimation (as we will demonstrate in the next chapter), it is also a useful tool for validating our measurements and instrumental models. Further, images can be a useful space for identifying systematics both in the instrument and the analysis pipeline. As with calibration, we implement all imaging using `FHD` and we utilize the instrumental beam model from Fagnoni et al. (2020).

Figure 7.1 shows an image of this beam model for both the North and East polarizations. We show only the central forty degrees of the beam as this is where we have the most confidence in our beam model and beyond which we have very little sensitivity. The full beam model used for all analysis steps extends out to the horizon, which is important in accounting for sources far from zenith. The central twenty degrees of the beam (between the nulls) accounts for the vast majority of our sensitivity, and is the only portion of the beam where our model is confident enough to expect an exceptionally good match between the data and model images.

As we've spent much of this thesis outlining various systematics in the HERA telescope (and there are many more we haven't included), it is worthwhile to begin by reviewing how `FHD` images of HERA data looked at the beginning of the commissioning process. Figure 7.2 shows the first set

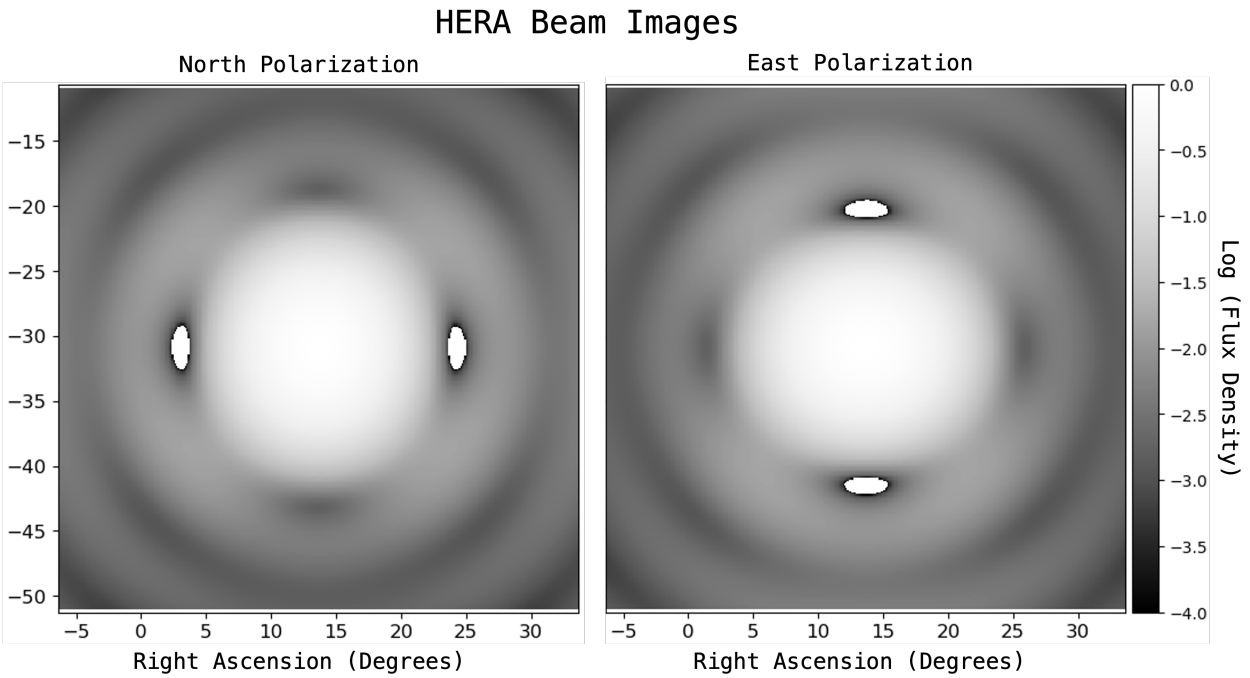


Figure 7.1: Image of HERA's instrument beam model in both the North and East polarizations. The beam model imaged here is described in Fagnoni et al. (2020).

of HERA images from `FHD`, using data from October, 2019. This was soon after HERA upgraded from phase I to phase II, which involved switching from a dipole to a Vivaldi feed and an entirely new digital system. This was also early in HERA's commissioning process and there were very few antennas online. We also note that this was before any of the systematics mitigation described in this thesis (particularly the PAM upgrade described in Chapter 4) and the character of phase II data was not yet well understood. The majority of images from the initial run failed so catastrophically that they were just a noisy hash structure without sources, and so we choose to show the *best* images from that observing season.

We first note that these images were created before the Fagnoni et al. (2020) beam model was established, and instead utilized a much more rudimentary beam model. We also had errors in how `FHD` was interpreting the beam model. Between these two factors, we see that the shape and width of the beam imprint is much different in the images than in the beam model shown in Figure 7.1, but this is expected. Of greater concern is that the interaction between sources and the beam model is creating significant large-scale structure that is not reproduced by the model, and therefore shows up very brightly in the residual. Even worse, we note that the data and the model are seeing a vastly different set of sources, and so many sources show up in the residual. In particular, we note the line of four consecutive sources in the data, circled in red in the residual. There is no such set of sources on the sky, and therefore we suspect some form of aliasing. This level of discrepancy between the model and the data indicates that either our data or our analysis (or both) are failing.

Next, Figure 7.3 shows a set of dirty, model, and residual images of both polarizations from a single 38 second observation taken during the first night of our analysis data set. We see the shape and nulls of the beam model are accurately represented in these images. Visual comparison between the dirty and model indicates that our images are represented well by the model. This is further confirmed by examining the residual images, where we see that almost all sources are exceptionally well matched between the two. There is only one source at approximately RA 22° where our data measured less power than the model predicted, and therefore we see an under-subtraction. We only show images for one time stamp here for the sake of brevity, but there are cases at other times when the reverse can happen and we measure more power than predicted, causing an over-subtraction. There are also many times when all sources are matched well and there are no visible sources in the residual image. We also note that the interaction between the beam nulls and sources is fairly

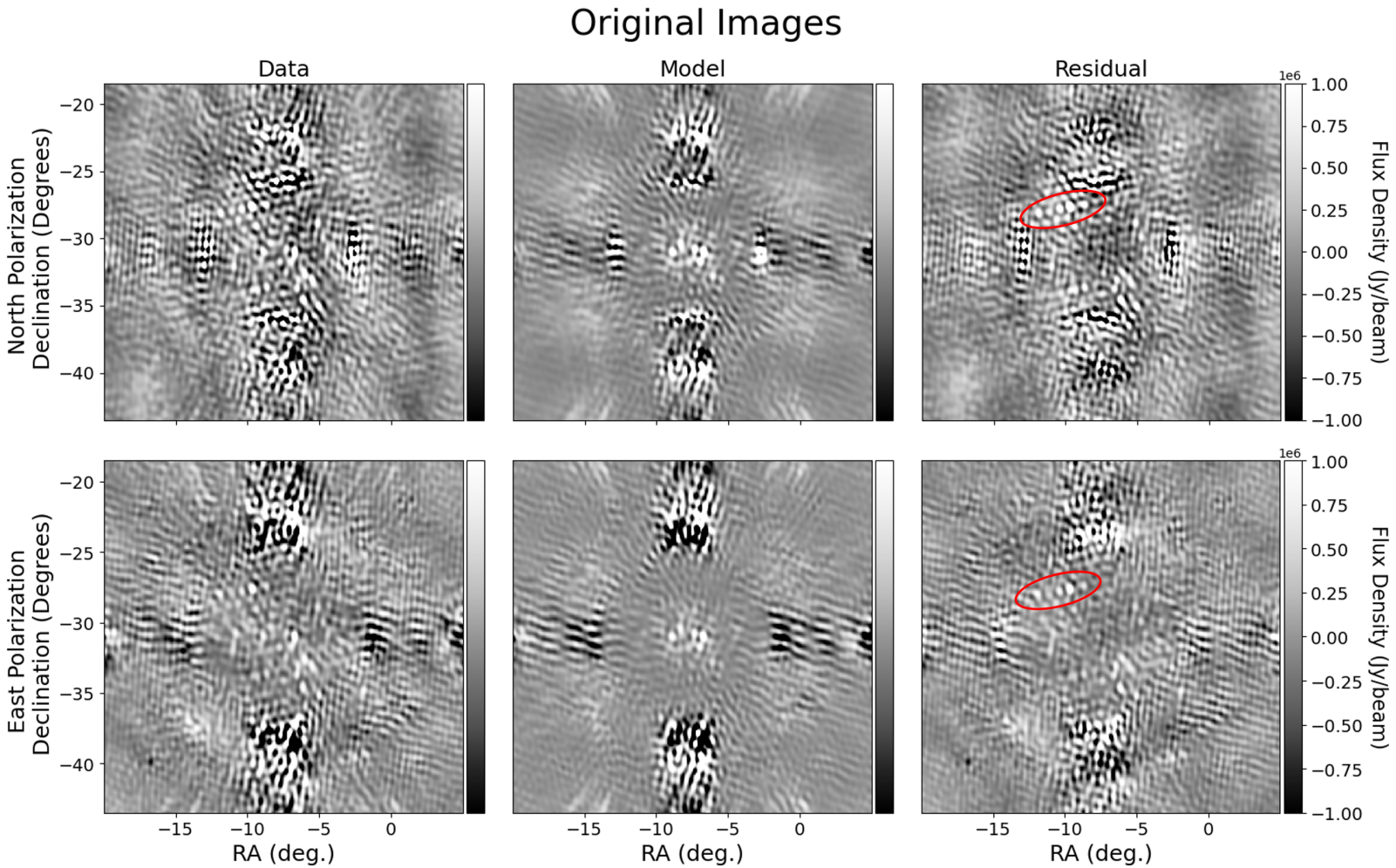


Figure 7.2: Sample images of HERA data from October 2019 for both the North (top) and East (bottom) polarizations from a single 9.6 second observation and averaged across 28.6MHz. The left column is dirty images, or images constructed from calibrated visibilities. The middle column is images of the model visibilities, and the right column is the difference between the dirty and model. We see that sources do not match between the model and the dirty, indicating a significant issue in our data and/or our imaging pipeline.

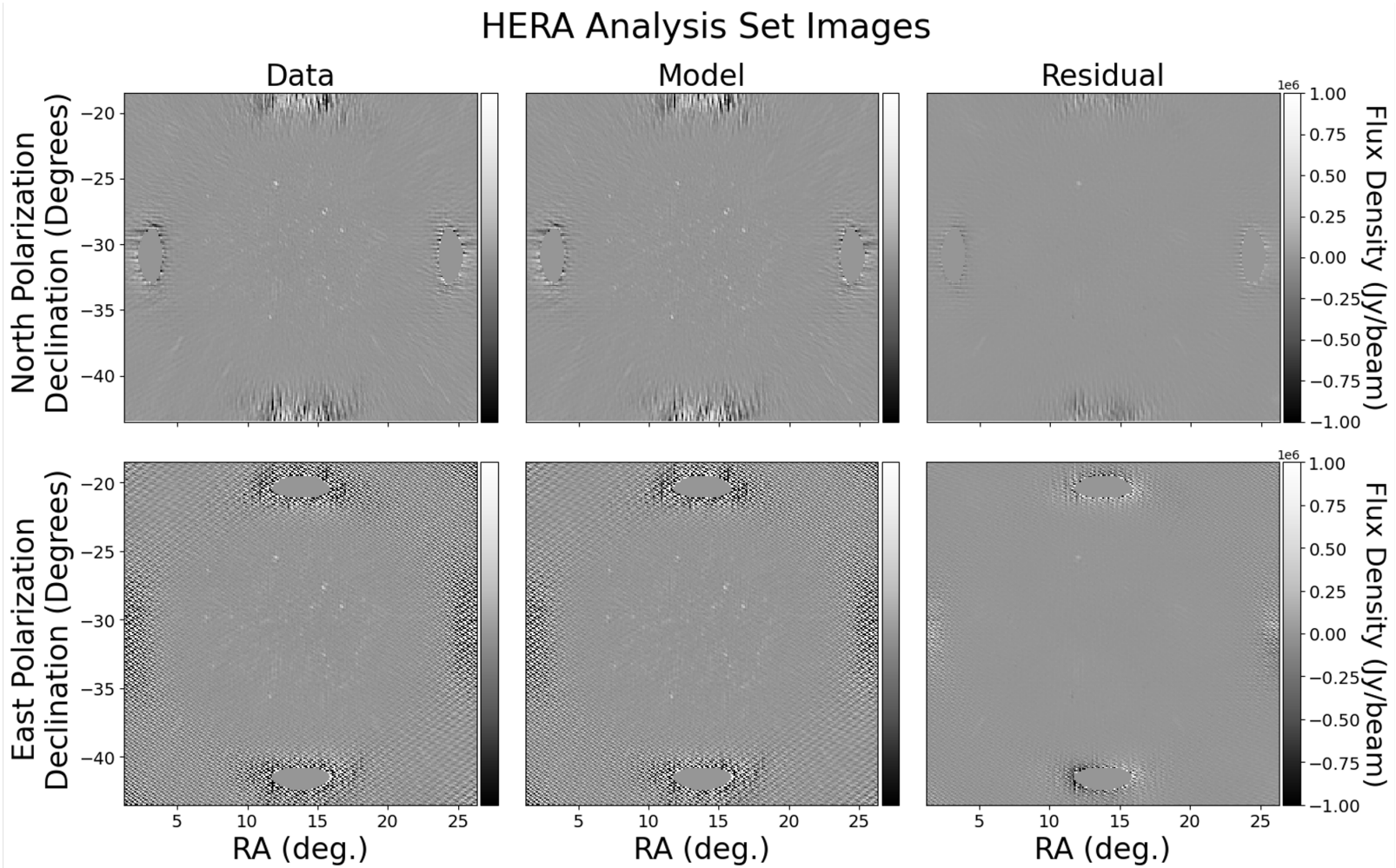
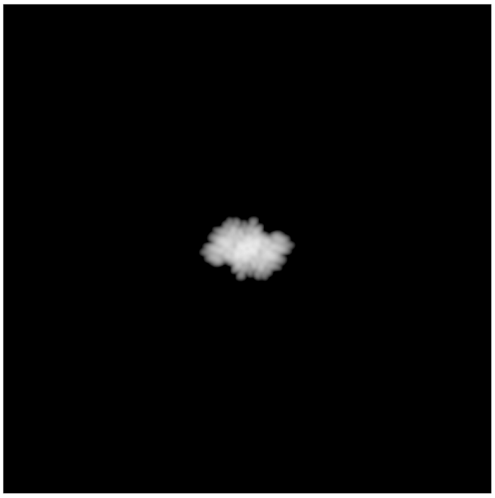


Figure 7.3: Sample images of current HERA data for both the North (top) and East (bottom) polarizations from a single 9.6 second observation and averaged across 20.4MHz. The left column is dirty images, or images constructed from calibrated visibilities. The middle column is images of the model visibilities, and the right column is the difference between the dirty and model. We see that sources match well between the model and dirty images, and we have a clean residual within the primary beam.

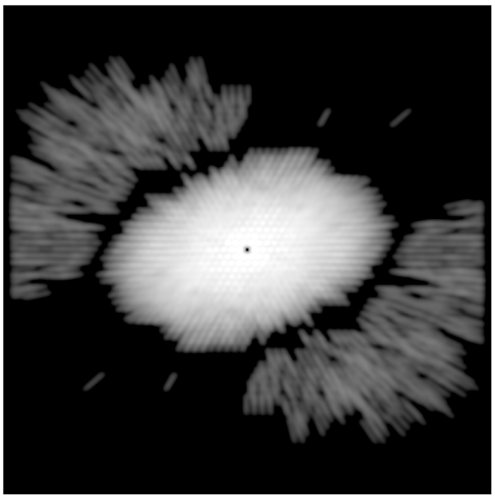
well modeled, as the power of those features is significantly suppressed in the residual.

There have been far too many changes to both the data and the analysis pipeline to describe every difference between these two image sets. However, we can say that the majority of changes to the quality of our images come from mitigation of systematics (particularly the excess correlation), a more accurate interpretation of the beam model throughout the code (which required extensive debugging to achieve), and above all a larger antenna set. This is partly attributable to continued construction of HERA and partly to developing a deeper understanding of how to identify malfunctioning antennas and removing them from the data set. In particular, we found that just a few misbehaving antennas could cause calibration to fail and produce very noisy images. The most notable antenna malfunction in image space is an antenna that is not correlating with its peers. In this case even just one uncorrelated antenna can prevent calibration from finding accurate phase solutions causing all antennas to have uncalibrated phases, which severely impedes the imaging process. The understanding of this systematic and development of flagging techniques outlined in Chapter 2 was one of the major advances that allowed us to start resolving sources correctly in our images.

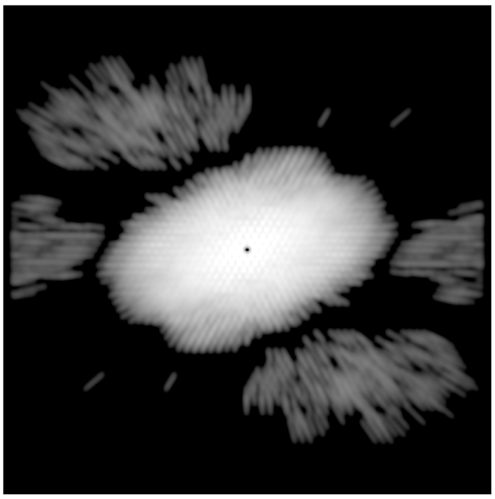
By a wide margin the biggest factor in improving image quality was increasing the number of usable antennas, both through the continued construction of HERA and through systematics mitigation. The number of baselines scales with the square of the number of antennas, meaning that a marginal increase in antennas can cause a significant increase in the number of baselines. The original imaging run included only 34 antennas, whereas the final analysis nights have between 80 and 120 antennas per polarization. The number of available baselines is critically important because each unique baseline acts as an additional sampling point in the uv plane, increasing the uv coverage. Figure 7.4 shows the uv weights for both the initial imaging run and the first night of the final analysis run. The weights reflect how much sensitivity we have in the uv plane based on our baseline set. In Figure 7.4a we see that the initial imaging run only had usable uv coverage for exceptionally short baselines, and therefore would have had exceptionally little sensitivity to point sources on the sky. Conversely, panels 7.4b and 7.4c show that the final analysis set has much more extensive coverage. In particular, we note the outer wings in these panels are entirely attributable to baselines including outrigger antennas. While these will not add much usable sensitivity for a power spectrum measurement at those scales, they do aid tremendously with calibration, which



(a) First run: both polarizations



(b) Analysis set weights: North polarization



(c) Analysis set weights: East polarization

Figure 7.4: Weights in the UV plane for a single observation, averaged over 20.4 MHz. Panel (a) shows weights for the first `FHD` imaging run and corresponds to the images in Figure 7.2. At this point antenna selection was the same across polarizations, and therefore there is only one set of weights. Panels (b) and (c) show the weights for the first night of the final analysis set for the North and East polarizations, respectively.

has ramifications for both imaging and power spectra. A baseline's location in the uv plane is dependent on both frequency and the phase center of the observation, and therefore uv coverage is improved with both frequency and time synthesis. The weights shown in all three panels are synthesized over 20.4 MHz, which causes the blurring effects. We will explore the effects of time synthesis on our weights in the next chapter.

While the images in Figure 7.3 are just one snapshot, we have examined images for all nights in our analysis set and found that they are all producing good images with clean residuals. Given this success, we feel confident that our data selection (and particularly our antenna selection) was effective at removing the vast majority of problematic data, and we are prepared to move forward with a power spectrum measurement.

Chapter 8

POWER SPECTRUM METHOD AND RESULTS

While a primary goal of 21 cm cosmology is to produce tomographic maps of neutral hydrogen during the Epoch of Reionization, this is not feasible with current technology and analysis techniques. Instead, we turn to the power spectrum measurement as a powerful statistical tool for detecting the 21 cm signal. Most fundamentally, the power spectrum reflects the brightness variance we measure at Fourier modes corresponding to the 21 cm signal. The ultimate goal of this thesis is to create a sufficiently systematics-free data set that we can perform a successful power spectrum measurement.

There are two distinct categories of power spectrum estimation within the field: delay-based and imaging based. These are sometimes also called measured and reconstructed power spectra, respectively. For a comprehensive review of the distinction between these approaches we refer the reader to Morales et al. (2018). For partly historical and partly practical reasons the large majority of the HERA collaboration analysis efforts are directed towards a delay-based power spectrum measurement. In fact, the collaboration has already produced power spectrum upper limits using this approach (Abdurashidova et al., 2022). While imaging-based power spectra are more common across the field, this thesis represents the first efforts towards an imaging-based power spectrum measurement with HERA phase II.[1] This has its own set of advantages and difficulties compared to the delay-based approach with HERA and the imaging-based approach with other instruments.

To properly understand the differences between this work and the mainline HERA analysis efforts we will begin by outlining the properties of the delay-based power spectrum in Section 8.1 and the imaging-based power spectrum in Section 8.2. Then, in Section 8.3 we will present the software package we use for power spectrum calculation, describe the diagnostic tools we utilize, and provide a final summary of our entire analysis pipeline. Finally, we will present our 2D power spectrum results in Section 8.4, along with a discussion of some important features, and finally we

[1] See Xu et al. (2022) for discussion of imaging power spectra with HERA phase I.

present our 1D results in Section 8.5.

8.1 Delay-Based Power Spectra

The delay-spectrum approach to power spectrum estimation was first introduced in Parsons et al. (2012) and has most notably been used to produce upper limits with PAPER (Kolopanis et al., 2019), the MWA (Kolopanis et al., 2023), and HERA (Abdurashidova et al., 2022). While this approach is well-developed and appreciated within the community, it is both newer and less common than the imaging approach. In this approach a visibility in in (u, v, f) space is Fourier transformed along the frequency axis and then squared to create a power spectrum. The result of this direct Fourier transform across frequency is commonly called k_τ, and can be thought of as measuring the time-delay between antennas (hence the name 'delay-spectrum'). Because the location of a baseline in the uv plane, and therefore the spatial mode it samples on the sky, is frequency dependent k_τ is actually slightly different than the line-of-sight axis, which we call $k_\parallel$. Notably this means that baseline contributions are the same across frequency, and therefore that redundant baseline sets can be coherently averaged before the Fourier transformation and square. Averaging that happens before squaring is more sensitive because it produces a lower noise measurement, and so this ability to utilize HERA's redundant antenna layout and perform baseline averaging is a critical advantage of the mainline HERA delay-based pipeline.

The primary advantage of a delay (or measured) power spectrum approach is that compared to an imaging approach it requires less precise calibration to be successful. This is largely because imaging-based power spectra are highly sensitive to frequency-independent calibration errors, while these have minimal effect on delay-based power spectra (Morales et al., 2018). Additionally, the delay approach is much less reliant on both the sky catalog and the instrumental beam model, and therefore is less sensitive to errors in either. Additionally, this approach is simpler to implement in software. A less demanding baseline code base has allowed those implementing HERA's delay approach more time to invest in developing systematics mitigation tools as part of their analysis pipeline. In particular, this approach lends itself nicely to delay rate and fringe rate filtering techniques, which have been used at length on HERA data (Garsden et al., 2024, Rath et al., 2024, in prep).

However, the delay approach also has some fundamental drawbacks. First, it cannot work within

the foreground wedge (described in Figure 8.2), which an imaging power spectrum theoretically can. Second, delay-based power spectrum measurements cannot be cross-correlated with measurements from other instruments. Third, it is not compatible with tomographic mapping, which is the long-term goal of this field. Therefore, the delay approach is an exciting technique for placing upper limits on the 21 cm power spectrum, or possibly making first detections, but does not have as much potential for long-term applications as an imaging approach.

8.2 Imaging-Based Power Spectra

The imaging (or reconstructed) power spectrum approach was initially developed in Morales and Hewitt (2004), and remains the most common technique used in the field. In this approach visibilities in (θ, f) space (or image space) are Fourier transformed along all three axes and then squared to create a power spectrum cube. This approach is called a 'reconstructed' approach because it involves reconstructing a sky estimate, or image, whereas the 'measured' approach never involves transforming to image space. We will discuss this process in more detail in Section 8.3, but will emphasize a few key features of the imaging approach here. First, we note that because visibilities are used to reconstruct an estimate of the sky before taking the frequency Fourier transform the baseline contributions are not the same at all frequencies. Further, the product of the frequency Fourier transform in this case is by construction equal to the line-of-sight axis $k_{\parallel}$.

The imaging approach requires more precise calibration and a more advanced software base than the delay approach, it is a more mature technique and we can utilize the `FHD-`ϵ`ppsilon` pipeline that has been rigorously developed and validated for more than a decade. This is a particular advantage with HERA data, as we have found throughout this work that our data is prone to unexpected systematics and instability. With the newer delay-based pipeline it is less straightforward to identify if an issue is sourced from the data or the analysis techniques. Additionally, working in image space makes our data and pipeline easier to validate by comparing to a known sky model, whereas the delay pipeline is more abstract and has no intuitive reference.

Unlike the delay approach, an imaging approach has potential applications within the wedge. Additionally, by measuring our data in image space our approach allows for future cross-correlations with other instruments and has the potential to be used in tomographic mapping. While both of these are future goals it is important to validate and advance the techniques that will eventually

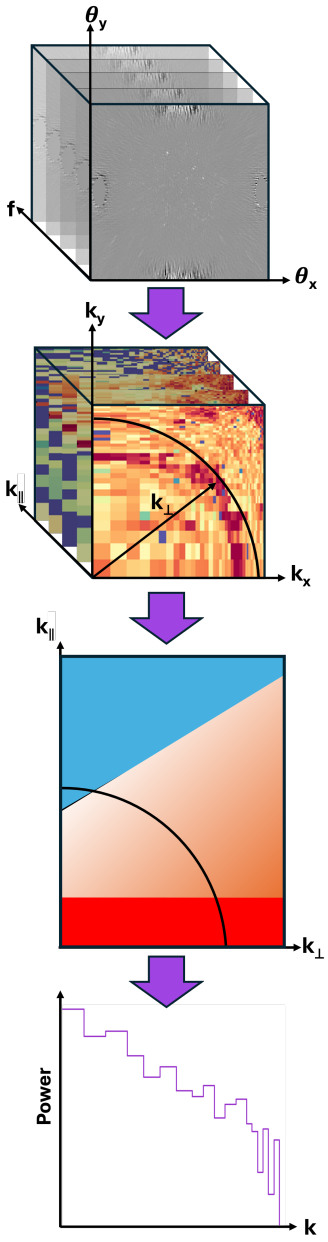
make those goals possible.

Imaging is the more common power spectrum approach across the field, but there are several unique aspects of HERA that make it an interesting instrument to image with. HERA's highly redundant antenna configuration is beneficial to a delay-based approach by enabling more redundant baseline averaging, but hinders an imaging approach by providing less coverage on the uv plane. However, HERA is a second-generation instrument and was designed based on many lessons learned from instruments like LOFAR and the MWA. For example, there are many spectral features in the MWA bandpass that are difficult to handle in analysis and were a key factor in HERA being designed with a particularly smooth bandpass shape. Additionally, when completed HERA will have 53,878m^2 of collecting area compared to 3,584m^2 for the MWA, 1,188m^2 for PAPER, and 35,762m^2 for LOFAR (DeBoer et al., 2017). More collecting area enhances signal to noise, which in turn makes systematics more identifiable. Understanding HERA's capacity for an imaging-based power spectrum and identifying any prohibitive systematics will help inform continued analysis with HERA and design considerations for future 21 cm experiments.

8.3 The Power Spectrum Pipeline

We implement all power spectrum estimation using the software package `ϵppsilon`, which was designed to work in tandem with `FHD` and is described at length in Barry et al. (2019a). Together, the `FHD-ϵppsilon` pipeline forms the reference pipeline for imaging and power spectrum estimation in the field, and has been used to produce several competitive power spectrum upper limits (Beardsley et al., 2016b; Barry et al., 2019b; Li et al., 2019).

Figure 8.1 shows a simplified schematic of the imaging-based power spectrum as implemented through the `FHD-ϵppsilon` pipeline. We begin with an image cube (8.1a), which is constructed from an image at each frequency (note the difference from images in the previous chapter, which were averaged over the frequency band). In this case we construct our images using Hierarchical Equal Area isoLatitude Pixelation (HEALPix) to create a set of equal area pixels on the sky. Previous analyses using the `FHD-ϵppsilon` pipeline with other instruments have done all time averaging in this space, but this is not feasible for HERA. Instead, we choose only to coherently average forty second LST sets across all 17 nights in this space, but not average across LSTs. See Appendix B for discussion of our averaging choices. After our chosen averaging in the image cube, we do a



(a)

Images at every frequency
form the 'image cube'.

(b)

We Fourier transform all three
axes to create a power spectrum cube.

(c)

Averaging the power spectrum
cube in cyllindrical shells
creates a 2D power spectrum.

(d)

Averaging the remaining **k** modes
(or averaging the power spectrum
cube in spherical shells) creates
a 1D power spectrum.

Figure 8.1: Diagram of the `ϵppsilon` power spectrum pipeline. We begin with an image cube (images at every frequency), Fourier transform along all axes to form a power spectrum cube, and then do either a cylindrical or spherical average to produce 1D and 2D power spectra. The 2D power spectra are primarily for diagnostic uses, while the 1D constitute a science result. See Figure 8.2 for further details on understanding the 2D power spectrum space. All analysis for the East and North polarizations is handled independently.

Fourier transform along all three axes and then square the data to get a 3D power spectrum cube $P(k_x, k_y, k_\parallel)$ (8.1b). Here, k_x and k_y are the Fourier transforms of the two spatial axes in image space θ_x and θ_y, and $k_\parallel$ is the Fourier transform along the frequency axis. We relate k_x and k_y to u and v by

$$u = \frac{k_x D_M(z)}{2\pi}, \quad v = \frac{k_y D_M(z)}{2\pi}, \tag{8.1}$$

where $D_M(z)$ is the transverse comoving distance (Hogg, 2000; Morales and Hewitt, 2004). Doing this Fourier transform and square on an integrated HEALPix cube will cause the noise to self-correlate, resulting in positive-definite noise. Instead, we split all observations into even and odd interleaved sets and integrate them separately. Then, instead of squaring the data we cross-multiply these interleaved sets. The sky change is negligible on these time scales and so the signal will correlate but the noise terms will remain independent, leading to zero-mean noise. While we choose not to coherently average across LSTs this interleave still pertains to our analysis. We process data through `FHD` in sets of four integrations, or forty seconds, and these sets of four are split into interleaved sets and propagated into ϵ`ppsilon`. We then coherently integrate those sets across all 17 nights before Fourier transforming and cross-multiplying. Then, we incoherently average all sets across LSTs. We use the volume of the instrumental beam model to normalize the power spectrum.

Next, we can average in cylindrical shells about the line-of-sight axis to get a 2D power spectrum $P(k_\perp, k_\parallel)$. This space is not treated as a science result, but rather is useful for identifying features in the power spectrum such as RFI, systematics leakage, or model errors. Figure 8.2 shows a more detailed schematic of this space and some of the most prominent features. The spectrally smooth foreground sources are mostly isolated to the lowest $k_\parallel$ mode, shown in red. However, the inherent frequency response of the instrument interacts with foreground sources causing their power to leak into higher $k_\parallel$ modes. This creates a wedge-like feature of excess foreground power, shown in orange. The wedge is a well-studied feature of interferometric power spectra; more details can be found in Parsons et al. (2012); Vedantham et al. (2012); Trott et al. (2012); Morales et al. (2012); Hazelton et al. (2013). We refer to the portion of Figure 8.2 above the wedge as the EoR window, as this is the region of $(k_\perp, k_\parallel)$ with the least contamination and therefore where we expect to be most sensitive to the EoR. The horizon and field of view lines indicate the maximum $k_\parallel$ modes that can

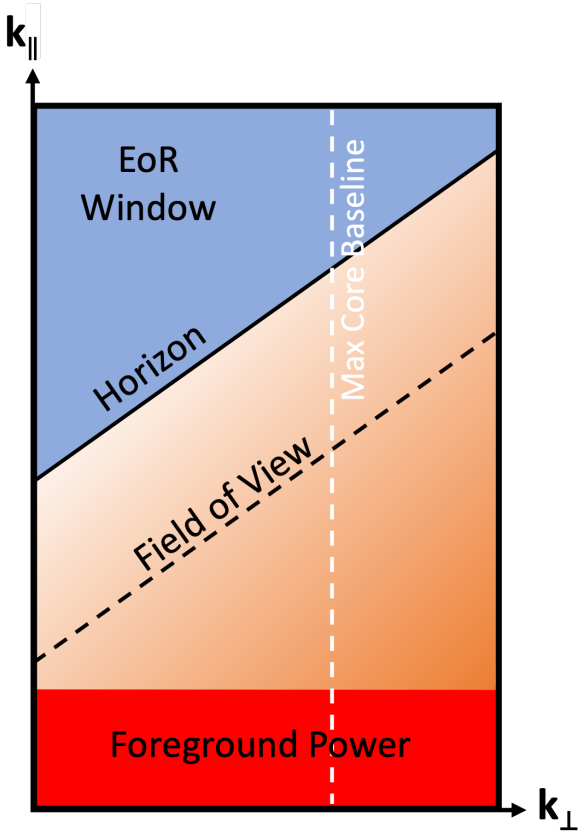


Figure 8.2: Schematic of the 2D power spectrum and some relevant features. The y-axis represents $k_{\parallel}$, which is the Fourier transform of the frequency axis and measures power spectrum structure along the line of sight. Power from the spectrally smooth foreground sources fall largely in the lowest $k_{\parallel}$ bin, shown in red. The x-axis represents $k_{\perp}$, which is the Fourier transform of the spatial coordinates in image space and measures the power spectrum at angular scales perpendicular to the line of sight. Large scales (or short baselines) are at at small $k_{\perp}$ and small angular scales (or long baselines) are at large $k_{\perp}$. The white dashed line shows the largest HERA baseline formed by two antennas within the array core. Above this baseline length only baselines including an outrigger exist, which there are very few of, and so we expect a significant drop in sensitivity. The spectral response of the instrument causes foreground power to bleed into higher $k_{\parallel}$ as a function of baseline length, creating the orange wedge. Above the wedge is an area that should be free from foreground power and therefore the region we are most sensitive to the 21 cm signal, giving it the name 'EoR window'.

be contaminated by this effect due to foreground sources on the horizon and at the edge of the primary beam, respectively.

We also use a white dashed line to indicate the maximum baseline length that can be achieved by two antennas in HERA's core. Larger $k_\perp$ modes can only be achieved by incorporating an outrigger antenna. This is an important feature to note for HERA, as there are substantially fewer baselines involving outriggers and therefore we expect a significant drop in sensitivity above this threshold. We can also visualize this threshold as the border of the bright core seen in the uv weights in Figure 7.4.

Finally, we can average the power spectrum cube in cylindrical shells to get a 1D power spectrum $P(k)$. This is shown pictorially in Figure 8.1 as following the 2D power spectrum because we make all analysis choices based on the 2D before computing the 1D. In practice, both the 1D and 2D power spectra are computed directly from the 3D power spectrum cube, and are not inherently sequential. This 1D power spectrum is the scientifically relevant result that can be directly compared against both theory and measurements from other instruments. The North and East polarizations are handled independently throughout the entire pipeline. We also propagate model visibilities through the pipeline in the same way as the data to create a model power spectrum.

8.4 2D Power Spectrum Results

Here we present our 2D power spectrum results. Figure 8.3 shows the data, model, and residual (data minus model) 2D power spectra for both polarizations. Red areas indicate high power, and blue speckled regions are noise-dominated. We note several important but expected features. First, the lowest $k_\parallel$ mode (which is dominated by foreground power) is dark red in both the data and model but not in the residual, indicating good subtraction. Second, the white dashed line indicates the baseline length corresponding to the maximum baseline not including an outrigger, beyond which we expect to see a drop in sensitivity. In particular we see noise-like pixels just right of the white dashed line where we have a gap in uv coverage.

There are two primary features in Figure 8.3 that are unexpected and indicative of a potential systematic. First, the purple dashed line denotes a power 'roof' that has leaked into the EoR window. Second, we see that the third $k_\perp$ bin from the left has very low power in all $k_\parallel$ bins compared to its peers. To somewhat better understand this systematic Figure 8.4 shows the 2D

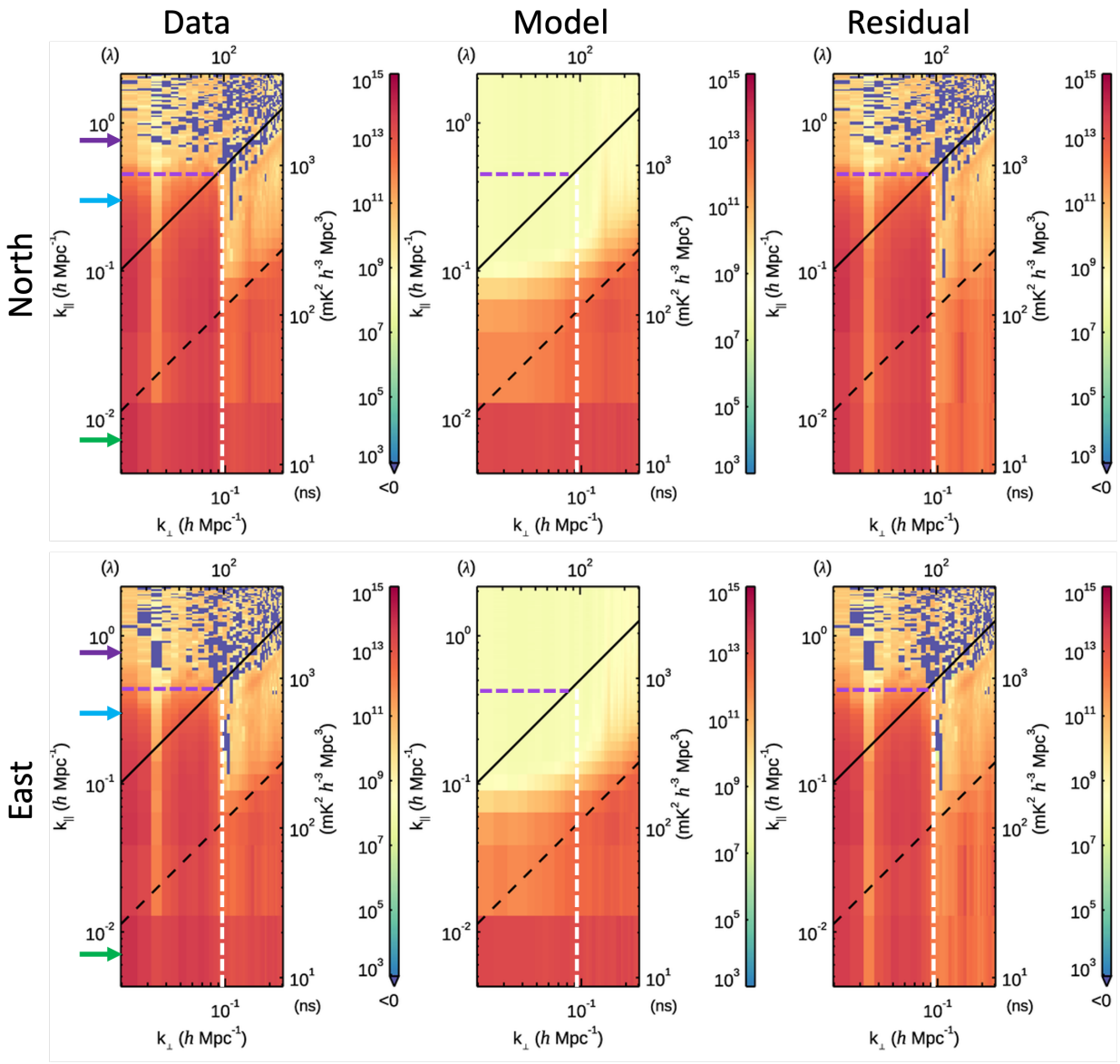


Figure 8.3: Data, model, and residual 2D power spectra for both polarizations, calculated as described in Section 8.3. The white dashed line indicates the baseline length of the maximum core baseline, beyond which we expect and observe a significant drop in sensitivity. The purple dashed line indicates the top of a block of power that is leaking into the window. The three arrows indicate the location of the slices shown in Figures 8.5, 8.6, and 8.7. See Figure B.7 for a non-annotated version.

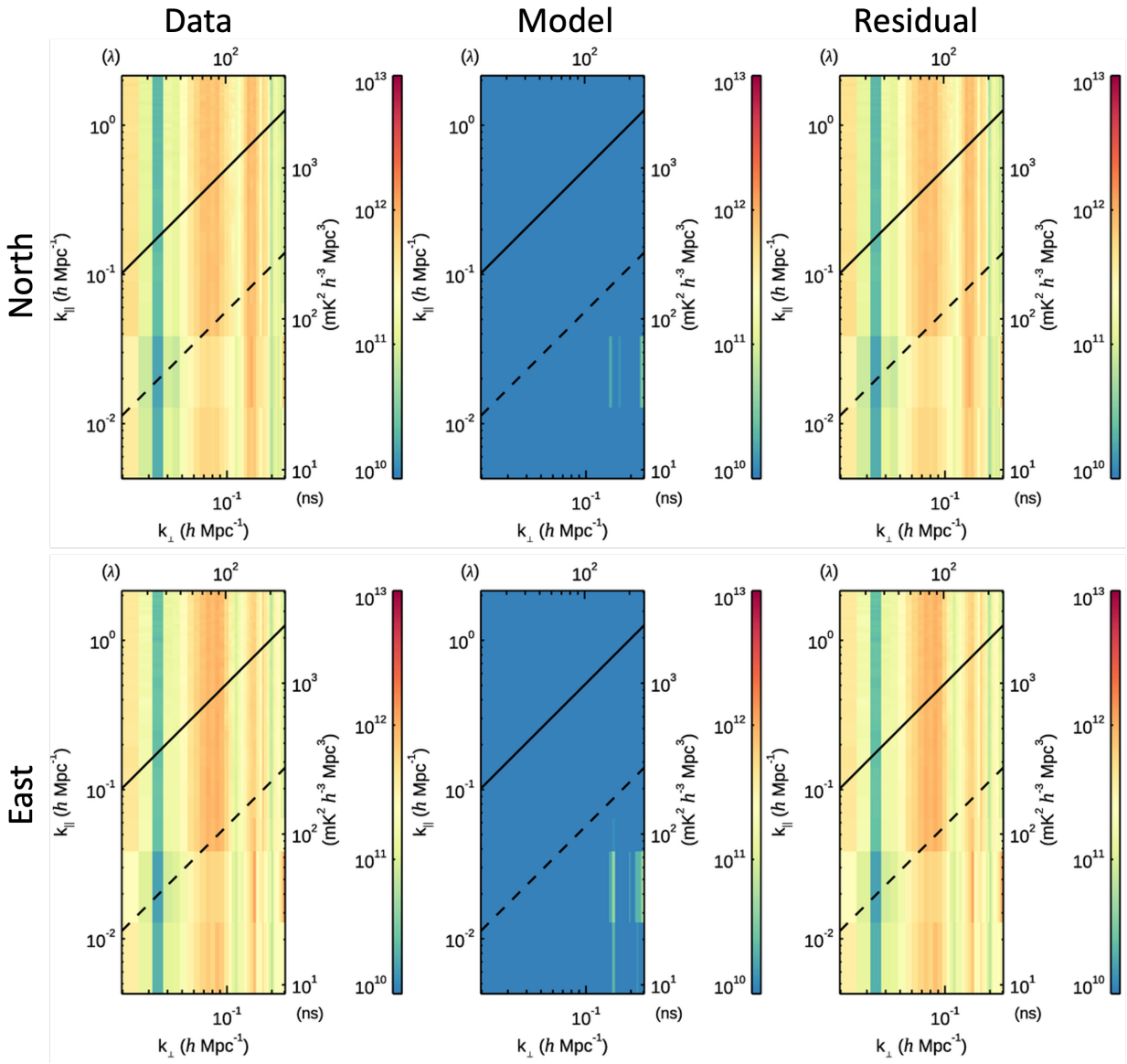


Figure 8.4: Data, model, and residual 2D difference power spectra for both polarizations. The difference PS is computed by taking the difference between the even and odd HEALPix cube sets, and is usually a good estimate of the thermal noise.

difference power spectra, which are calculated by taking the difference between the even and odd HEALPix cubes. In the case where all RFI and time-dependent systematics are properly flagged or removed this serves as a good estimate of the thermal noise. This figure shows that the same bin has very low noise. This is troublesome, as a true sampling gap would also appear in the model, but we see no evidence of the feature in the model power spectrum or difference power spectrum. Referring back to Figure 7.4, we cannot identify any weighting features that could explain either the low signal bin or the power roof.

To shed more light on these features we look at selected $k_{\parallel}$ slices in the $k_x - k_y$ plane. This space is helpful for identifying potential instrumental influences on the power spectrum as the $k_x - k_y$ plane is equivalent to the instrumental plane. Figure 8.5 shows a slice of the standard and difference power spectra for both polarizations at $k_{\parallel} = 0$, indicated by a green arrow in Figure 8.3. This slice is also equivalent to the front pane of panel b in Figure 8.1. The blue brackets indicate the location of the low power $k_{\perp}$ bin identified in the 2D power spectrum. The green bracket indicates a window of unexpected elevated power along the k_x axis. From this figure we see that the lower power bin is likely the due rather to a lack of systematic, and neighboring bins are actually exhibiting an excess of power.

Figures 8.6 and 8.7 show slices at a higher $k_{\parallel}$ bins, indicated by a blue and purple arrow in Figure 8.3, respectively. Examining slices at different $k_{\parallel}$ bins allows us to better understand the spectral nature of these features and to inspect regions of the 2D power spectrum where we expect different behavior. These features exhibit several noteworthy behaviors. First, the elevated power lump is not symmetric between k_x and k_y, but rather seems mostly isolated to k_x. Second, it appears at a similar k_x location in both polarizations, but its shape and extent are different between the two. Third, this elevated power is about two orders of magnitude brighter in the power spectrum than in the difference, but it is still clearly visible in the difference. This means that there is some subtraction, which indicates some level of time stability, but the lack of full subtraction indicates that the source of this power has some temporal fluctuation. In particular, along the k_y axis we see the difference slice roughly 3.5 orders of magnitude fainter than the power slice, compared to roughly 1.5-2 orders of magnitude fainter within the blob feature. This indicates that this feature has significant temporal fluctuations.

We also note that in Figure 8.6 there are noise-like pixels to both the right and left of the

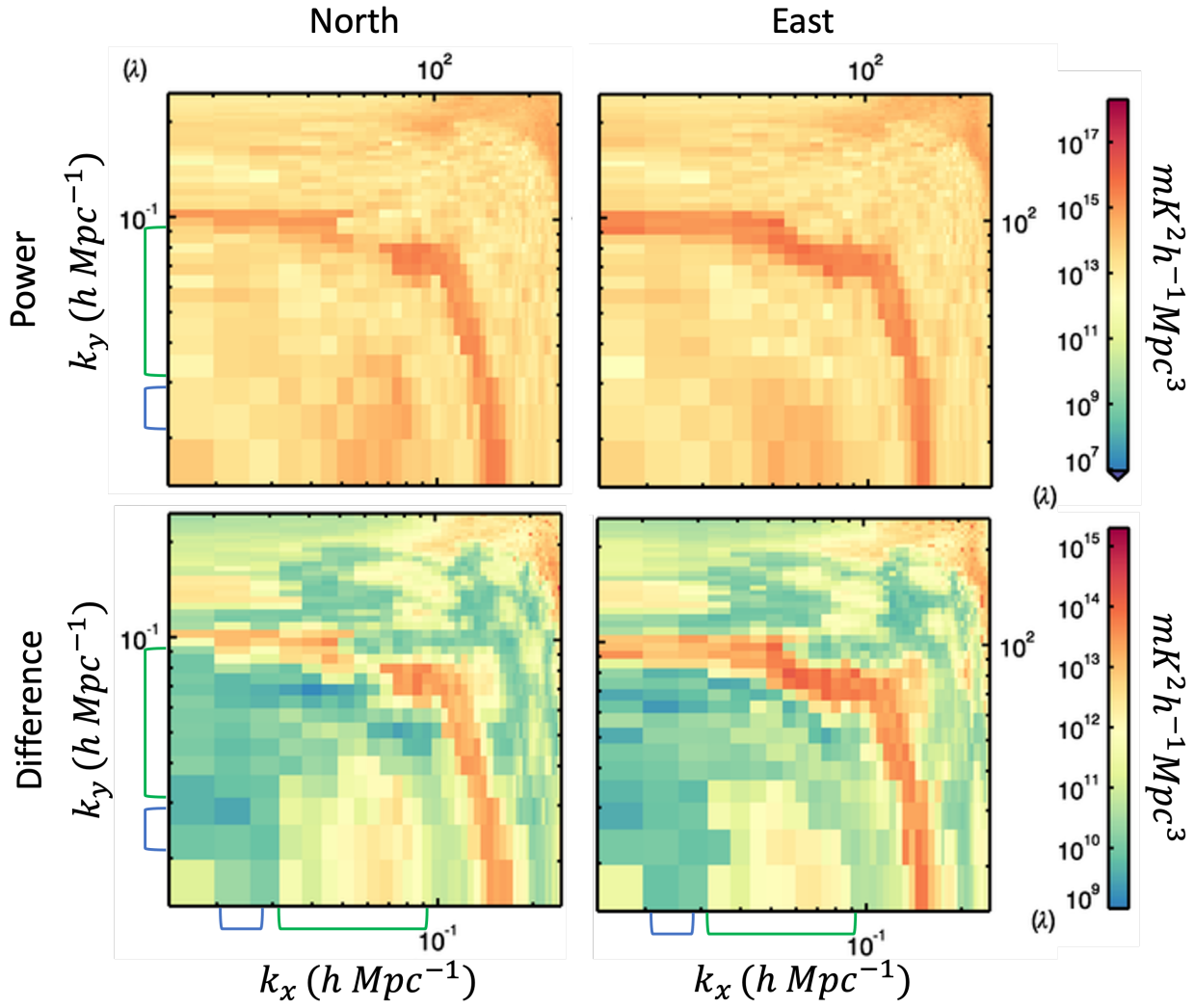


Figure 8.5: Slices of the power spectrum (top) and difference power spectrum (bottom) for both the North (left) and East (right) polarizations all at $k_{\parallel} = 0$. This bin is indicated by a green arrow in Figure 8.3. The blue brackets indicate the location of the low power $k_{\perp}$ bin, and the green bracket indicates the range of k_x where we see elevated power. Note that 2D power spectrum, the power slices and the difference slices are all on different color scales.

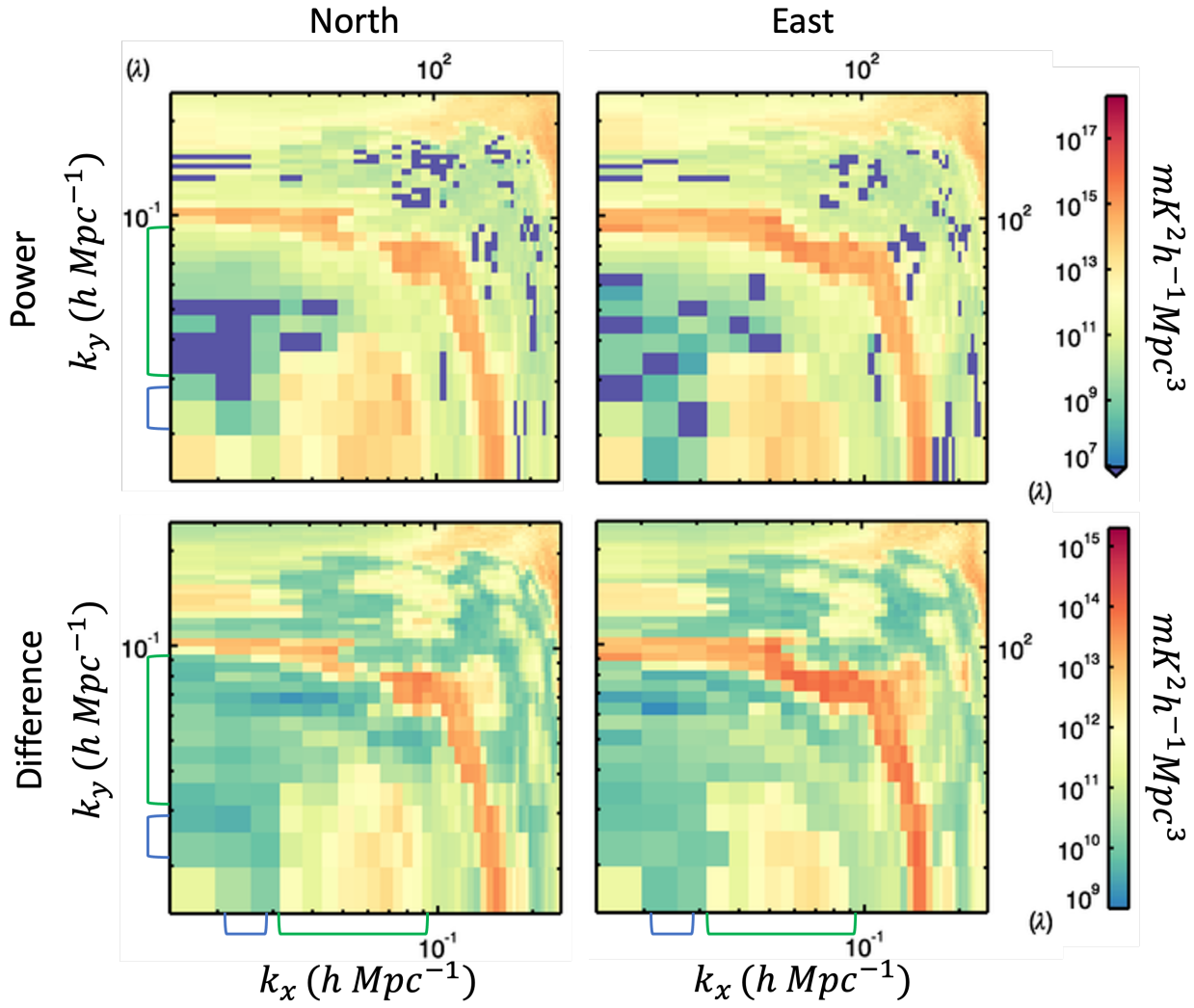


Figure 8.6: Slices of the power spectrum (top) and difference power spectrum (bottom) for both the North (left) and East (right) polarizations all at $k_{\parallel} = 0.2$. This bin is indicated by a blue arrow in Figure 8.3. The blue brackets indicate the location of the low power $k_{\perp}$ bin, and the green bracket indicates the range of k_x where we see elevated power. Note that 2D power spectrum, the power slices and the difference slices are all on different color scales.

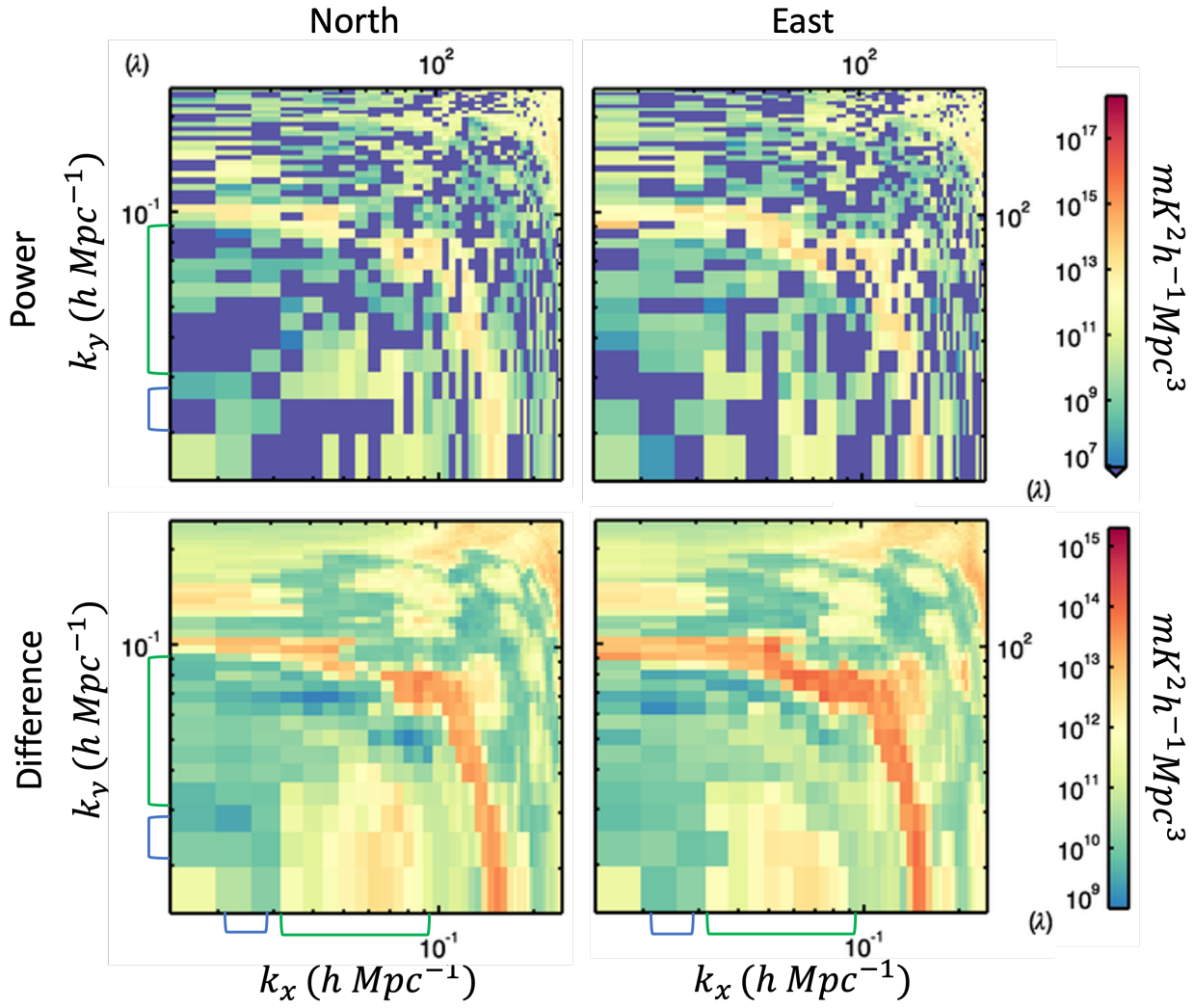


Figure 8.7: Slices of the power spectrum (top) and difference power spectrum (bottom) for both the North (left) and East (right) polarizations all at $k_{\parallel} = 0.78$. This bin is indicated by a purple arrow in Figure 8.3. The blue brackets indicate the location of the low power $k_{\perp}$ bin, and the green bracket indicates the range of k_x where we see elevated power. Note that 2D power spectrum, the power slices and the difference slices are all on different color scales.

elevated power lump in the power spectrum slice. This indicates that in absence of this feature, we would expect to see noise-like pixels in the 2D power spectrum at this value of $k_{\parallel}$. Referring back to Figure 8.3 we note that this bin is within the EoR window for $k_{\perp} < 0.6$, and is below the purple dashed line we used to indicate the roof of unexpected power leaking into the window. This implies that without this feature we would likely see the large majority of the EoR window as noise-dominated. Additionally, Figure 8.7 shows that while we are noise-dominated at higher $k_{\parallel}$ there is still evidence of this feature. This is a powerful finding, as it reveals that the source of this feature greatly effects the lower portion of the window, but also bleeds into even higher $k_{\parallel}$ modes. A feature of this magnitude severely restricts the science sensitivity of our power spectrum measurement.

There are many notable characteristics of this feature that make it difficult to identify the underlying cause. First, we see no evidence of this feature in our model, which indicates that it cannot be a true sampling issue (see Figure C.1). Second, it does not rotate with polarization. Our instrumental beam model is rotated ninety degrees between the North and East polarizations, and so we would expect issues arising from errors in our beam model to show this rotation, which our feature does not. Third, the feature is somewhat isolated to the k_x axis.

These three pieces together *could* indicate that this is a sky feature. We have seen evidence both in HERA and in other instruments that relatively short East-West oriented baselines are exceptionally sensitive to the galactic plane and can exhibit troubling behavior particularly when the galactic plane is in the poorly modeled sidelobes of the beam. However, there are several reasons to doubt this theory. First, we recall from Figure 5.4 that our observation window is as far from the galactic plane as possible. It is still feasible that we could see effects from the galaxy, but it would be surprising to see them so strong even when most of the galaxy is below the horizon. Second, if this feature were due to sky effects we would expect to see it evolve as the sky rotates. Appendix C shows that we see inconclusive differences in the nature of this feature between the first and last observations of our window. The differences are significant enough that we cannot rule out sky contributions, but subtle enough that the sky cannot explain the entirety of this feature. Finally, shorter baselines are the most sensitive to emission from the galactic plane, which makes it difficult to explain why this feature does not extend to smaller k_x modes.

Further study will be necessary to conclusively uncover the underlying cause of this excess power

feature. In particular, it will be useful to compare against HERA's mainline analysis efforts and fringe-rate filtering techniques to identify whether this feature is visible in delay space. Additionally, we could simulate the Global Sky Model (Zheng et al., 2016) to better explore how the galactic plane manifests in the power spectrum space. Both of these ideas will be the topic of future works. For now, we will proceed to show HERA's 1D power spectrum measurement.

8.5 1D Power Spectrum Results

Figure 8.8 shows the spherically averaged 1D power spectrum measurement and upper limit based on all data in the final analysis set. This data was coherently averaged across nights and incoherently averaged across LSTs. The k bins where the measured power drops to zero indicates regions where the power spectrum is noise dominated. With the current integration level this limit is significantly higher than other limits in the field. However, this is an encouraging result as we see that we do have noise dominated regions, which indicates that a deeper integration could yield a scientifically significant power spectrum result. Further, we see a hump at lower k bins that is due to the blob feature identified in the previous section. If this feature can be mitigated either through resolution in the instrument or through analysis techniques it is likely that the power spectrum would be noise

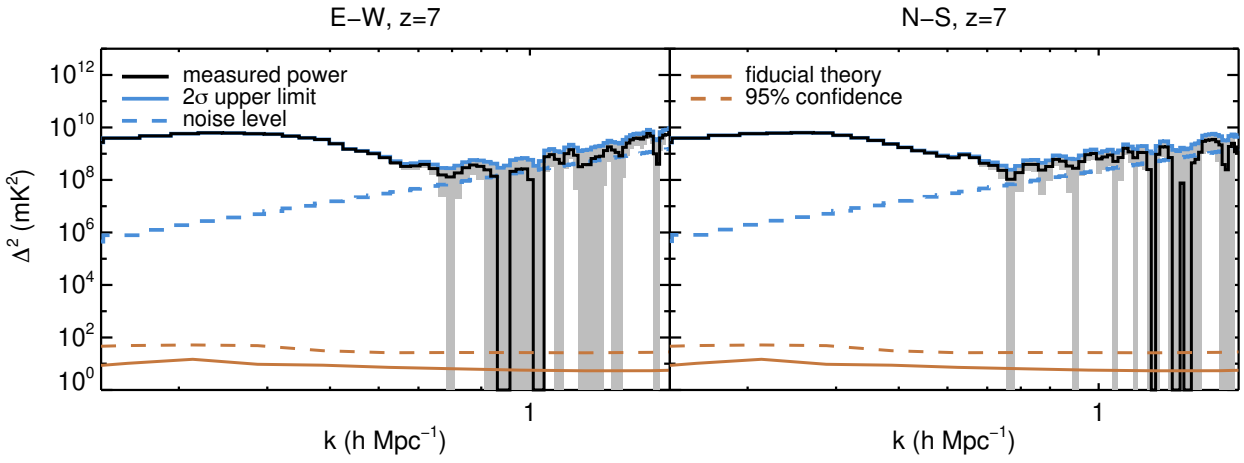


Figure 8.8: Spherically averaged 1D power spectrum measurements including all data from the final analysis set, along with a 2σ upper limit.

dominated at many more low k bins.

8.6 Conclusions

Throughout this thesis we have found that systematics mitigation is vital to successful analysis with HERA. With careful data monitoring we've identified and addressed many systematics within HERA, with excess correlation being the most critical. We've developed techniques for identifying and flagging problematic data, and through that we formulated the highest quality imaging data set feasible with current HERA observations. Despite tremendous progress and improvements, we have identified yet another systematic that is prevalent in the 2D power spectrum space and is the primary factor causing power leakage into the EoR window. The previous section showed that we do still have a small range of power spectrum space that is noise-limited, and therefore deeper integrations have the potential to produce a meaningful upper limit in that k range.

The best possible future prospects for imaging-based power spectrum measurements with HERA involve both increasing the set of available antennas (through bringing new antennas online and further systematics mitigation in existing antennas) and developing a deeper understanding of the underlying systematic that is causing this excess blob structure in the power spectrum. Regarding known systematics in HERA, mitigation of temporal structure in the instrument is the most vital next step in improving the systematics environment. Much analysis has already been done on this systematic, but more will be necessary to reach a true resolution. In addition, further investigation into the power spectrum blob and potential mitigation of this feature has the potential to yield significantly improved power spectrum results. In particular, this investigation may include more testing of the temporal evolution of this structure as it relates to sky rotation. Additionally, simulation of cross-coupling between antennas and its manifestation in the power spectrum would be enlightening. With effort in both of these areas, along with a deeper integration, HERA phase II has true potential to place a meaningful upper limit on the 21 cm power spectrum.

BIBLIOGRAPHY

Synthesis Imaging in Radio Astronomy II, volume 180 of *Astronomical Society of the Pacific Conference Series*, January 1999.

Zara Abdurashidova, James E. Aguirre, Paul Alexander, Zaki S. Ali, Yanga Balfour, Adam P. Beardsley, Gianni Bernardi, Tashalee S. Billings, Judd D. Bowman, Richard F. Bradley, Philip Bull, Jacob Burba, Steve Carey, Chris L. Carilli, Carina Cheng, David R. DeBoer, Matt Dexter, Eloy de Lera Acedo, Taylor Dibblee-Barkman, Joshua S. Dillon, John Ely, Aaron Ewall-Wice, Nicolas Fagnoni, Randall Fritz, Steven R. Furlanetto, Kingsley Gale-Sides, Brian Glendenning, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Bryna J. Hazelton, Jacqueline N. Hewitt, Jack Hickish, Daniel C. Jacobs, Austin Julius, Nicholas S. Kern, Joshua Kerrigan, Piyanat Kittiwisit, Saul A. Kohn, Matthew Kolopanis, Adam Lanman, Paul La Plante, Telalo Lekalake, David Lewis, Adrian Liu, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary E. Martinot, Eunice Matsetela, Andrei Mesinger, Mathakane Molewa, Miguel F. Morales, Tshegofalang Mosiane, Steven G. Murray, Abraham R. Neben, Bojan Nikolic, Chuneeta D. Nunhokee, Aaron R. Parsons, Nipanjana Patra, Robert Pascua, Samantha Pieterse, Jonathan C. Pober, Nima Razavi-Ghods, Jon Ringuette, James Robnett, Kathryn Rosie, Peter Sims, Saurabh Singh, Craig Smith, Angelo Syce, Nithyanandan Thyagarajan, Peter K. G. Williams, and Haoxuan Zheng. First results from hera phase i: Upper limits on the epoch of reionization 21 cm power spectrum. *The Astrophysical Journal*, 925(2):221, February 2022. ISSN 1538-4357. doi: 10.3847/1538-4357/ac1c78. URL http://dx.doi.org/10.3847/1538-4357/ac1c78.

James E Aguirre, Aaron R Parsons, Richard F Bradley, Christopher L Carilli, David R DeBoer, Zaki S Ali, Adrian Liu, Daniel C Jacobs, David F Moore, Jonathan C Pober, et al. The precision array for probing the epoch of reionization (paper): Overview and recent results. In *2014 United States National Committee of URSI National Radio Science Meeting (USNC-URSI NRSM)*, pages 1–1. IEEE, 2014.

N. Barry, B. Hazelton, I. Sullivan, M. F. Morales, and J. C. Pober. Calibration requirements for detecting the 21 cm epoch of reionization power spectrum and implications for the ska. *Monthly Notices of the Royal Astronomical Society*, 461(3):3135–3144, June 2016. ISSN 1365-2966. doi: 10.1093/mnras/stw1380. URL http://dx.doi.org/10.1093/mnras/stw1380.

N. Barry, A. P. Beardsley, R. Byrne, B. Hazelton, M. F. Morales, J. C. Pober, and I. Sullivan. The fhd/eppsilon epoch of reionisation power spectrum pipeline. *Publications of the Astronomical Society of Australia*, 36, 2019a. ISSN 1448-6083. doi: 10.1017/pasa.2019.21. URL http://dx.doi.org/10.1017/pasa.2019.21.

N. Barry, M. Wilensky, C. M. Trott, B. Pindor, A. P. Beardsley, B. J. Hazelton, I. S. Sullivan, M. F. Morales, J. C. Pober, J. Line, B. Greig, R. Byrne, A. Lanman, W. Li, C. H. Jordan, R. C. Joseph, B. McKinley, M. Rahimi, S. Yoshiura, J. D. Bowman, B. M. Gaensler, J. N. Hewitt, D. C. Jacobs, D. A. Mitchell, N. Udaya Shankar, S. K. Sethi, R. Subrahmanyan, S. J. Tingay, R. L. Webster, and J. S. B. Wyithe. Improving the Epoch of Reionization Power Spectrum Results from Murchison Widefield Array Season 1 Observations. *Astrophys. J.*, 884(1):1, 2019b. doi: 10.3847/1538-4357/ab40a8.

A. P. Beardsley, B. J. Hazelton, I. S. Sullivan, P. Carroll, N. Barry, M. Rahimi, B. Pindor, C. M. Trott, J. Line, Daniel C. Jacobs, and et al. First season mwa eor power spectrum results at redshift 7. *The Astrophysical Journal*, 833(1):102, Dec 2016a. ISSN 1538-4357. doi: 10.3847/1538-4357/833/1/102. URL http://dx.doi.org/10.3847/1538-4357/833/1/102.

A. P. Beardsley, B. J. Hazelton, I. S. Sullivan, P. Carroll, N. Barry, M. Rahimi, B. Pindor, C. M. Trott, J. Line, Daniel C. Jacobs, M. F. Morales, J. C. Pober, G. Bernardi, Judd D. Bowman, M. P. Busch, F. Briggs, R. J. Cappallo, B. E. Corey, A. de Oliveira-Costa, Joshua S. Dillon, D. Emrich, A. Ewall-Wice, L. Feng, B. M. Gaensler, R. Goeke, L. J. Greenhill, J. N. Hewitt, N. Hurley-Walker, M. Johnston-Hollitt, D. L. Kaplan, J. C. Kasper, H. S. Kim, E. Kratzenberg, E. Lenc, A. Loeb, C. J. Lonsdale, M. J. Lynch, B. McKinley, S. R. McWhirter, D. A. Mitchell, E. Morgan, A. R. Neben, Nithyanandan Thyagarajan, D. Oberoi, A. R. Offringa, S. M. Ord, S. Paul, T. Prabu, P. Procopio, J. Riding, A. E. E. Rogers, A. Roshi, N. Udaya Shankar, Shiv K. Sethi, K. S. Srivani, R. Subrahmanyan, M. Tegmark, S. J. Tingay, M. Waterson, R. B. Wayth,

R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams, C. Wu, and J. S. B. Wyithe. First Season MWA EoR Power Spectrum Results at Redshift 7. *Astrophys. J.*, 833(1):102, 2016b. doi: 10.3847/1538-4357/833/1/102.

L. V. Benkevitch, A. E. E. Rogers, C. J. Lonsdale, R. J. Cappallo, D. Oberoi, P. J. Erickson, and K. A. V. Baker. Van vleck correction generalization for complex correlators with multilevel quantization, 2016.

Philippe Berger, Laura B. Newburgh, Mandana Amiri, Kevin Bandura, Jean-François Cliche, Liam Connor, Meiling Deng, Nolan Denman, Matt Dobbs, Mateus Fandino, Adam J. Gilbert, Deborah Good, Mark Halpern, David Hanna, Adam D. Hincks, Gary Hinshaw, Carolin Höfer, Andre M. Johnson, Tom L. Landecker, Kiyoshi W. Masui, Juan Mena Parra, Niels Oppermann, Ue-Li Pen, Jeffrey B. Peterson, Andre Recnik, Timothy Robishaw, J. Richard Shaw, Seth Siegel, Kris Sigurdson, Kendrick Smith, Emilie Storer, Ian Tretyakov, Kwinten Van Gassen, Keith Vanderlinde, and Donald Wiebe. Holographic beam mapping of the chime pathfinder array. In Helen J. Hall, Roberto Gilmozzi, and Heather K. Marshall, editors, *Ground-based and Airborne Telescopes VI*. SPIE, August 2016. doi: 10.1117/12.2233782. URL http://dx.doi.org/10.1117/12.2233782.

Lindsay M. Berkhout, Daniel C. Jacobs, Zuhra Abdurashidova, Tyrone Adams, James E. Aguirre, Paul Alexander, Zaki S. Ali, Rushelle Baartman, Yanga Balfour, Adam P. Beardsley, Gianni Bernardi, Tashalee S. Billings, Judd D. Bowman, Richard F. Bradley, Philip Bull, Jacob Burba, Ruby Byrne, Steven Carey, Chris L. Carilli, Kai-Feng Chen, Carina Cheng, Samir Choudhuri, David R. DeBoer, Eloy de Lera Acedo, Matt Dexter, Joshua S. Dillon, Scott Dynes, Nico Eksteen, John Ely, Aaron Ewall-Wice, Nicolas Fagnoni, Randall Fritz, Steven R. Furlanetto, Kingsley Gale-Sides, Hugh Garsden, Bharat Kumar Gehlot, Abhik Ghosh, Brian Glendenning, Adelie Gorce, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Bryna J. Hazelton, Jacqueline N. Hewitt, Jack Hickish, Tian Huang, Alec Josaitis, Austin Julius, MacCalvin Kariseb, Nicholas S. Kern, Joshua Kerrigan, Honggeun Kim, Piyanat Kittiwisit, Saul A. Kohn, Matthew Kolopanis, Adam Lanman, Paul La Plante, Adrian Liu, Anita Loots, Yin-Zhe Ma, David Harold Edward MacMahon, Lourence Malan, Cresshim Malgas, Keith Malgas, Bradley Marero, Zachary E. Martinot, Andrei Mesinger, Mathakane Molewa, Miguel F. Morales, Tshego-

falang Mosiane, Steven G. Murray, Abraham R. Neben, Bojan Nikolic, Chuneeta Devi Nunhokee, Hans Nuwegeld, Aaron R. Parsons, Robert Pascua, Nipanjana Patra, Samantha Pieterse, Yuxiang Qin, Eleanor Rath, Nima Razavi-Ghods, Daniel Riley, James Robnett, Kathryn Rosie, Mario G. Santos, Peter Sims, Saurabh Singh, Dara Storer, Hilton Swarts, Jianrong Tan, Nithyanandan Thyagarajan, Pieter van Wyngaarden, Peter K. G. Williams, Haoxuan Zheng, and Zhilei Xu. Hydrogen epoch of reionization array (hera) phase ii deployment and commissioning. *Publications of the Astronomical Society of the Pacific*, 136(4):045002, April 2024. ISSN 1538-3873. doi: 10.1088/1538-3873/ad3122. URL http://dx.doi.org/10.1088/1538-3873/ad3122.

G. Bernardi, A. G. de Bruyn, G. Harker, M. A. Brentjens, B. Ciardi, V. Jelić, L. V. E. Koopmans, P. Labropoulos, A. Offringa, V. N. Pandey, and et al. Foregrounds for observations of the cosmological 21cm line. *Astronomy & Astrophysics*, 522:A67, Nov 2010. ISSN 1432-0746. doi: 10.1051/0004-6361/200913420. URL http://dx.doi.org/10.1051/0004-6361/200913420.

Sarah E I Bosman, Frederick B Davies, George D Becker, Laura C Keating, Rebecca L Davies, Yongda Zhu, Anna-Christina Eilers, Valentina D'Odorico, Fuyan Bian, Manuela Bischetti, Stefano V Cristiani, Xiaohui Fan, Emanuele P Farina, Martin G Haehnelt, Joseph F Hennawi, Girish Kulkarni, Andrei Mesinger, Romain A Meyer, Masafusa Onoue, Andrea Pallottini, Yuxiang Qin, Emma Ryan-Weber, Jan-Torge Schindler, Fabian Walter, Feige Wang, and Jinyi Yang. Hydrogen reionization ends by $z = 5.3$: Lyman-alpha optical depth measured by the xqr-30 sample. *Monthly Notices of the Royal Astronomical Society*, 514(1):55–76, June 2022. ISSN 1365-2966. doi: 10.1093/mnras/stac1046. URL http://dx.doi.org/10.1093/mnras/stac1046.

Ruby Byrne. *Enabling the Next Generation of 21 cm Cosmology: Precision Analysis Techniques for Radio Cosmology Observations*. PhD thesis, University of Washington, Seattle, January 2021.

Ruby Byrne, Miguel F. Morales, Bryna Hazelton, Wenyang Li, Nichole Barry, Adam P. Beardsley, Ronniy Joseph, Jonathan Pober, Ian Sullivan, and Cathryn Trott. Fundamental limitations on the calibration of redundant 21 cm cosmology instruments and implications for hera and the ska. *The Astrophysical Journal*, 875(1):70, April 2019. ISSN 1538-4357. doi: 10.3847/1538-4357/ab107d. URL http://dx.doi.org/10.3847/1538-4357/ab107d.

Ruby Byrne, Miguel F Morales, Bryna Hazelton, Ian Sullivan, Nichole Barry, Christene Lynch, Jack

L B Line, and Daniel C Jacobs. A map of diffuse radio emission at 182mhz to enhance epoch of reionization observations in the southern hemisphere. *Monthly Notices of the Royal Astronomical Society*, 510(2):2011–2024, November 2021. ISSN 1365-2966. doi: 10.1093/mnras/stab3276. URL http://dx.doi.org/10.1093/mnras/stab3276.

Ruby Byrne, Miguel F. Morales, Bryna J. Hazelton, and Michael Wilensky. A unified calibration framework for 21 cm cosmology. MNRAS, 503(2):2457–2477, May 2021. doi: 10.1093/mnras/stab647.

Thomas A. Callister, Marin M. Anderson, Gregg Hallinan, Larry R. D'addario, Jayce Dowell, Namir E. Kassim, T. Joseph W. Lazio, Danny C. Price, and Frank K. Schinzel. A first search for prompt radio emission from a gravitational-wave event. *The Astrophysical Journal*, 877(2): L39, Jun 2019. ISSN 2041-8213. doi: 10.3847/2041-8213/ab2248. URL http://dx.doi.org/10.3847/2041-8213/ab2248.

Emanuele Castorina, Simon Foreman, Dionysios Karagiannis, Adrian Liu, Kiyoshi W. Masui, Pieter D. Meerburg, Laura B. Newburgh, Paul O'Connor, Andrej Obuljen, Hamsa Padmanabhan, J. Richard Shaw, Anže Slosar, Paul Stankus, Peter T. Timbie, Benjamin Wallisch, and Martin White. Packed Ultra-wideband Mapping Array (PUMA): Astro2020 RFI Response. *arXiv e-prints*, art. arXiv:2002.05072, February 2020.

A Chokshi, J L B Line, N Barry, D Ung, D Kenney, A McPhail, A Williams, and R L Webster. Dual polarization measurements of mwa beampatterns at 137 mhz. *Monthly Notices of the Royal Astronomical Society*, 502(2):1990–2004, Jan 2021. ISSN 1365-2966. doi: 10.1093/mnras/stab156. URL http://dx.doi.org/10.1093/mnras/stab156.

F. de Gasperin, T. J. Dijkema, A. Drabent, M. Mevius, D. Rafferty, R. van Weeren, M. Brüggen, J. R. Callingham, K. L. Emig, G. Heald, H. T. Intema, L. K. Morabito, A. R. Offringa, R. Oonk, E. Orrù, H. Röttgering, J. Sabater, T. Shimwell, A. Shulevski, and W. Williams. Systematic effects in LOFAR data: A unified calibration strategy. A&A, 622:A5, February 2019. doi: 10.1051/0004-6361/201833867.

E. de Lera Acedo, P. Bolli, F. Paonessa, G. Virone, E. Colin-Beltran, N. Razavi-Ghods, I. Aicardi,

A. Lingua, P. Maschio, J. Monari, G. Naldi, M. Piras, and G. Pupillo. Ska aperture array verification system: electromagnetic modeling and beam pattern measurements using a micro uav. *Experimental Astronomy*, 45(1):1–20, December 2017. ISSN 1572-9508. doi: 10.1007/s10686-017-9566-x. URL http://dx.doi.org/10.1007/s10686-017-9566-x.

D. R. DeBoer, A. R. Parsons, J. E. Aguirre, P. Alexander, Z. S. Ali, A. P. Beardsley, G. Bernardi, J. D. Bowman, R. F. Bradley, C. L. Carilli, C. Cheng, E. de Lera Acedo, J. S. Dillon, A. Ewall-Wice, G. Fadana, N. Fagnoni, R. Fritz, S. R. Furlanetto, B. Glendenning, B. Greig, J. Grobbelaar, B. J. Hazelton, J. N. Hewitt, J. Hickish, D. C. Jacobs, A. Julius, M. Kariseb, S. A. Kohn, T. Lekalake, A. Liu, A. Loots, D. MacMahon, L. Malan, C. Malgas, M. Maree, Z. Martinot, N. Mathison, E. Matsetela, A. Mesinger, M. F. Morales, A. R. Neben, N. Patra, S. Pieterse, J. C. Pober, N. Razavi-Ghods, J. Ringuette, J. Robnett, K. Rosie, R. Sell, C. Smith, A. Syce, M. Tegmark, N. Thyagarajan, P. K. G. Williams, and H. Zheng. Hydrogen Epoch of Reionization Array (HERA). PASP, 129(4):045001, April 2017. doi: 10.1088/1538-3873/129/974/045001.

James Di Francesco, Dean Chalmers, Nolan Denman, Laura Fissel, Rachel Friesen, Bryan Gaensler, Julie Hlavacek-Larrondo, Helen Kirk, Brenda Matthews, Christopher O'Dea, Tim Robishaw, Erik Rosolowsky, Michael Rupen, Sarah Sadavoy, Samar Sa-Harb, Greg Sivakoff, Mehrnoosh Tahani, Nienke van der Marel, Jacob White, and Christine Wilson. The Next Generation Very Large Array. In *Canadian Long Range Plan for Astronomy and Astrophysics White Papers*, volume 2020, page 32, October 2019. doi: 10.5281/zenodo.3765763.

Joshua S. Dillon and Aaron R. Parsons. Redundant Array Configurations for 21 cm Cosmology. ApJ, 826(2):181, August 2016. doi: 10.3847/0004-637X/826/2/181.

Joshua S Dillon, Saul A Kohn, Aaron R Parsons, James E Aguirre, Zaki S Ali, Gianni Bernardi, Nicholas S Kern, Wenyang Li, Adrian Liu, Chuneeta D Nunhokee, and Jonathan C Pober. Polarized redundant-baseline calibration for 21 cm cosmology without adding spectral structure. *Monthly Notices of the Royal Astronomical Society*, April 2018. ISSN 1365-2966. doi: 10.1093/mnras/sty1060. URL http://dx.doi.org/10.1093/mnras/sty1060.

Joshua S Dillon, Max Lee, Zaki S Ali, Aaron R Parsons, Naomi Orosz, Chuneeta Devi Nunhokee, Paul La Plante, Adam P Beardsley, Nicholas S Kern, Zara Abdurashidova, James E

Aguirre, Paul Alexander, Yanga Balfour, Gianni Bernardi, Tashalee S Billings, Judd D Bowman, Richard F Bradley, Phil Bull, Jacob Burba, Steve Carey, Chris L Carilli, Carina Cheng, David R DeBoer, Matt Dexter, Eloy de Lera Acedo, John Ely, Aaron Ewall-Wice, Nicolas Fagnoni, Randall Fritz, Steven R Furlanetto, Kingsley Gale-Sides, Brian Glendenning, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Bryna J Hazelton, Jacqueline N Hewitt, Jack Hickish, Daniel C Jacobs, Austin Julius, Joshua Kerrigan, Piyanat Kittiwisit, Saul A Kohn, Matthew Kolopanis, Adam Lanman, Telalo Lekalake, David Lewis, Adrian Liu, Yin-Zhe Ma, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary E Martinot, Eunice Matsetela, Andrei Mesinger, Mathakane Molewa, Miguel F Morales, Tshegofalang Mosiane, Steven Murray, Abraham R Neben, Bojan Nikolic, Robert Pascua, Nipanjana Patra, Samantha Pieterse, Jonathan C Pober, Nima Razavi-Ghods, Jon Ringuette, James Robnett, Kathryn Rosie, Mario G Santos, Peter Sims, Craig Smith, Angelo Syce, Max Tegmark, Nithyanandan Thyagarajan, Peter K G Williams, and Haoxuan Zheng. Redundant-baseline calibration of the hydrogen epoch of reionization array. *Monthly Notices of the Royal Astronomical Society*, 499(4):5840–5861, October 2020a. ISSN 1365-2966. doi: 10.1093/mnras/staa3001. URL http://dx.doi.org/10.1093/mnras/staa3001.

Joshua S Dillon, Max Lee, Zaki S Ali, Aaron R Parsons, Naomi Orosz, Chuneeta Devi Nunhokee, Paul La Plante, Adam P Beardsley, Nicholas S Kern, Zara Abdurashidova, and et al. Redundant-baseline calibration of the hydrogen epoch of reionization array. *Monthly Notices of the Royal Astronomical Society*, 499(4):5840–5861, Oct 2020b. ISSN 1365-2966. doi: 10.1093/mnras/staa3001. URL http://dx.doi.org/10.1093/mnras/staa3001.

Michael W. Eastwood, Marin M. Anderson, Ryan M. Monroe, Gregg Hallinan, Morgan Catha, Jayce Dowell, Hugh Garsden, Lincoln J. Greenhill, Brian C. Hicks, Jonathon Kocz, Danny C. Price, Frank K. Schinzel, Harish Vedantham, and Yuankun Wang. The 21 cm power spectrum from the cosmic dawn: First results from the ovro-lwa. *The Astronomical Journal*, 158(2):84, July 2019. ISSN 1538-3881. doi: 10.3847/1538-3881/ab2629. URL http://dx.doi.org/10.3847/1538-3881/ab2629.

A. Ewall-Wice, Joshua S. Dillon, J. N. Hewitt, A. Loeb, A. Mesinger, A. R. Neben, A. R. Offringa,

M. Tegmark, N. Barry, A. P. Beardsley, and et al. First limits on the 21 cm power spectrum during the epoch of x-ray heating. *Monthly Notices of the Royal Astronomical Society*, 460(4): 4320–4347, May 2016. ISSN 1365-2966. doi: 10.1093/mnras/stw1022. URL http://dx.doi.org/10.1093/mnras/stw1022.

Aaron Ewall-Wice, Richard Bradley, David Deboer, Jacqueline Hewitt, Aaron Parsons, James Aguirre, Zaki S. Ali, Judd Bowman, Carina Cheng, Abraham R. Neben, Nipanjana Patra, Nithyanandan Thyagarajan, Mariet Venter, Eloy de Lera Acedo, Joshua S. Dillon, Roger Dickenson, Phillip Doolittle, Dennis Egan, Mike Hedrick, Patricia Klima, Saul Kohn, Patrick Schaffner, John Shelton, Benjamin Saliwanchik, H. A. Taylor, Rusty Taylor, Max Tegmark, and Butch Wirt. The Hydrogen Epoch of Reionization Array Dish. II. Characterization of Spectral Structure with Electromagnetic Simulations and Its Science Implications. ApJ, 831(2):196, November 2016. doi: 10.3847/0004-637X/831/2/196.

Nicolas Fagnoni, Eloy de Lera Acedo, David R DeBoer, Zara Abdurashidova, James E Aguirre, Paul Alexander, Zaki S Ali, Yanga Balfour, Adam P Beardsley, Gianni Bernardi, and et al. Understanding the hera phase i receiver system with simulations and its impact on the detectability of the eor delay power spectrum. *Monthly Notices of the Royal Astronomical Society*, 500(1):1232–1242, Oct 2020. ISSN 1365-2966. doi: 10.1093/mnras/staa3268. URL http://dx.doi.org/10.1093/mnras/staa3268.

Nicolas Fagnoni, Eloy de Lera Acedo, Nick Drought, David R. DeBoer, Daniel Riley, Nima Razavi-Ghods, Steven Carey, and Aaron R. Parsons. Design of the new wideband vivaldi feed for the hera radio-telescope phase ii. *IEEE Transactions on Antennas and Propagation*, 69(12):8143–8157, December 2021. ISSN 1558-2221. doi: 10.1109/tap.2021.3083788. URL http://dx.doi.org/10.1109/TAP.2021.3083788.

Steven R. Furlanetto, S. Peng Oh, and Frank H. Briggs. Cosmology at low frequencies: The 21cm transition and the high-redshift universe. *Physics Reports*, 433(4-6):181–301, Oct 2006. ISSN 0370-1573. doi: 10.1016/j.physrep.2006.08.002. URL http://dx.doi.org/10.1016/j.physrep.2006.08.002.

Hugh Garsden, Philip Bull, Mike Wilensky, Zuhra Abdurashidova, Tyrone Adams, James E.

Aguirre, Paul Alexander, Zaki S. Ali, Rushelle Baartman, Yanga Balfour, Adam P. Beardsley, Lindsay M. Berkhout, Gianni Bernardi, Tashalee S. Billings, Judd D. Bowman, Richard F. Bradley, Jacob Burba, Steven Carey, Chris L. Carilli, Kai-Feng Chen, Carina Cheng, Samir Choudhuri, David R. DeBoer, Eloy de Lera Acedo, Matt Dexter, Joshua S. Dillon, Scott Dynes, Nico Eksteen, John Ely, Aaron Ewall-Wice, Nicolas Fagnoni, Randall Fritz, Steven R. Furlanetto, Kingsley Gale-Sides, Bharat Kumar Gehlot, Abhik Ghosh, Brian Glendenning, Adelie Gorce, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Bryna J. Hazelton, Jacqueline N. Hewitt, Jack Hickish, Tian Huang, Daniel C. Jacobs, Alec Josaitis, Austin Julius, MacCalvin Kariseb, Nicholas S. Kern, Joshua Kerrigan, Honggeun Kim, Piyanat Kittiwisit, Saul A. Kohn, Matthew Kolopanis, Adam Lanman, Paul La Plante, Adrian Liu, Anita Loots, Yin-Zhe Ma, David H. E. MacMahon, Lourence Malan, Cresshim Malgas, Keith Malgas, Bradley Marero, Zachary E. Martinot, Andrei Mesinger, Mathakane Molewa, Miguel F. Morales, Tshegofalang Mosiane, Steven G. Murray, Abraham R. Neben, Bojan Nikolic, Chuneeta Devi Nunhokee, Hans Nuwegeld, Aaron R. Parsons, Robert Pascua, Nipanjana Patra, Samantha Pieterse, Yuxiang Qin, Eleanor Rath, Nima Razavi-Ghods, Daniel Riley, James Robnett, Kathryn Rosie, Mario G. Santos, Peter Sims, Saurabh Singh, Dara Storer, Hilton Swarts, Jianrong Tan, Nithyanandan Thyagarajan, Pieter van Wyngaarden, Peter K. G. Williams, Zhilei Xu, and Haoxuan Zheng. A demonstration of the effect of fringe-rate filtering in the hydrogen epoch of reionization array delay power spectrum pipeline. 2024.

Nickolay Y Gnedin and Piero Madau. Modeling cosmic reionization. *Living Reviews in Computational Astrophysics*, 8(1):3, 2022.

G. Hallinan, V. Ravi, and Deep Synoptic Array Team. The DSA-2000: A Radio Survey Camera. In *American Astronomical Society Meeting Abstracts*, volume 53 of *American Astronomical Society Meeting Abstracts*, page 316.05, January 2021.

Bryna J. Hazelton, Miguel F. Morales, and Ian S. Sullivan. The fundamental multi-baseline mode-mixing foreground in 21 cm epoch of reionization observations. *The Astrophysical Journal*, 770 (2):156, June 2013. ISSN 1538-4357. doi: 10.1088/0004-637x/770/2/156. URL http://dx.doi.org/10.1088/0004-637X/770/2/156.

HERA Collaboration, Zara Abdurashidova, James E. Aguirre, Paul Alexander, Zaki S. Ali, Yanga Balfour, Adam P. Beardsley, Gianni Bernardi, Tashalee S. Billings, Judd D. Bowman, Richard F. Bradley, Philip Bull, Jacob Burba, Steve Carey, Chris L. Carilli, Carina Cheng, David R. DeBoer, Matt Dexter, Eloy de Lera Acedo, Taylor Dibblee-Barkman, Joshua S. Dillon, John Ely, Aaron Ewall-Wice, Nicolas Fagnoni, Randall Fritz, Steven R. Furlanetto, Kingsley Gale-Sides, Brian Glendenning, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Bryna J. Hazelton, Jacqueline N. Hewitt, Jack Hickish, Daniel C. Jacobs, Austin Julius, Nicholas S. Kern, Joshua Kerrigan, Piyanat Kittiwisit, Saul A. Kohn, Matthew Kolopanis, Adam Lanman, Paul La Plante, Telalo Lekalake, David Lewis, Adrian Liu, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary E. Martinot, Eunice Matsetela, Andrei Mesinger, Mathakane Molewa, Miguel F. Morales, Tshegofalang Mosiane, Steven G. Murray, Abraham R. Neben, Bojan Nikolic, Chuneeta D. Nunhokee, Aaron R. Parsons, Nipanjana Patra, Robert Pascua, Samantha Pieterse, Jonathan C. Pober, Nima Razavi-Ghods, Jon Ringuette, James Robnett, Kathryn Rosie, Peter Sims, Saurabh Singh, Craig Smith, Angelo Syce, Nithyanandan Thyagarajan, Peter K. G. Williams, and Haoxuan Zheng. First results from hera phase i: Upper limits on the epoch of reionization 21 cm power spectrum, 2021.

David W. Hogg. Distance measures in cosmology, 2000.

Daniel Jacobs, Jacob Burba, Judd Bowman, Abraham R. Neben, Benjamin Stinnett, Lauren Turner, Kali Johnson, Michael Busch, Jay Allison, Marc Leatham, Victoria Serrano Rodriguez, Mason Denney, and David Nelson. First demonstration of echo: An external calibrator for hydrogen observatories. *Publications of the Astronomical Society of the Pacific*, 129(973), March 2017. ISSN 0004-6280. doi: 10.1088/1538-3873/aa56b9. Funding Information: ECHO development is supported by a grant from the National Science Foundation AST program through award 1407646. D.C.J. acknowledges support by the NSF Astronomy and Astrophysics Fellowship Program through award 1401708. We thank Rich Bradley and staff at the National Radio Astronomy Observatory, Green Bank and Andri Gretarsson and Embry Riddle Aeronautical Observatory for generously supporting this project with their time and equipment. This project makes use of the Astropy python library (Astropy Collaboration et al. 2013). Publisher Copyright: © 2017, The Astronomical Society of the Pacific. All rights reserved.

C. H. Jordan, S. Murray, C. M. Trott, R. B. Wayth, D. A. Mitchell, M. Rahimi, B. Pindor, P. Procopio, and J. Morgan. Characterization of the Ionosphere Above the Murchison Radio Observatory Using the Murchison Widefield Array. *Mon. Not. R. Astron. Soc.*, 471(4):3974–3987, 2017. doi: 10.1093/mnras/stx1797.

Ronniy C Joseph, C M Trott, R B Wayth, and A Nasirudin. Calibration and 21-cm power spectrum estimation in the presence of antenna beam variations. *Monthly Notices of the Royal Astronomical Society*, 492(2):2017–2028, 12 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz3375. URL https://doi.org/10.1093/mnras/stz3375.

Nicholas S. Kern, Aaron R. Parsons, Joshua S. Dillon, Adam E. Lanman, Nicolas Fagnoni, and Eloy de Lera Acedo. Mitigating Internal Instrument Coupling for 21 cm Cosmology. I. Temporal and Spectral Modeling in Simulations. ApJ, 884(2):105, October 2019. doi: 10.3847/1538-4357/ab3e73.

Nicholas S. Kern, Joshua S. Dillon, Aaron R. Parsons, Christopher L. Carilli, Gianni Bernardi, Zara Abdurashidova, James E. Aguirre, Paul Alexander, Zaki S. Ali, Yanga Balfour, Adam P. Beardsley, Tashalee S. Billings, Judd D. Bowman, Richard F. Bradley, Philip Bull, Jacob Burba, Steven Carey, Carina Cheng, David R. DeBoer, Matt Dexter, Eloy de Lera Acedo, John Ely, Aaron Ewall-Wice, Nicolas Fagnoni, Randall Fritz, Steve R. Furlanetto, Kingsley Gale-Sides, Brian Glendenning, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Bryna J. Hazelton, Jacqueline N. Hewitt, Jack Hickish, Daniel C. Jacobs, Austin Julius, Joshua Kerrigan, Piyanat Kittiwisit, Saul A. Kohn, Matthew Kolopanis, Adam Lanman, Paul La Plante, Telalo Lekalake, Adrian Liu, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary E. Martinot, Eunice Matsetela, Andrei Mesinger, Mathakane Molewa, Miguel F. Morales, Tshegofalang Mosiane, Steven G. Murray, Abraham R. Neben, Bojan Nikolic, Chuneeta D. Nunhokee, Nipanjana Patra, Samantha Pieterse, Jonathan C. Pober, Nima Razavi-Ghods, Jon Ringuette, James Robnett, Kathryn Rosie, Peter Sims, Craig Smith, Angelo Syce, Nithyanandan Thyagarajan, Peter K. G. Williams, and Haoxuan Zheng. Absolute Calibration Strategies for the Hydrogen Epoch of Reionization Array and Their Impact on the 21 cm Power Spectrum. ApJ, 890(2):122, February 2020a. doi: 10.3847/1538-4357/ab67bc.

Nicholas S. Kern, Aaron R. Parsons, Joshua S. Dillon, Adam E. Lanman, Adrian Liu, Philip Bull, Aaron Ewall-Wice, Zara Abdurashidova, James E. Aguirre, Paul Alexander, Zaki S. Ali, Yanga Balfour, Adam P. Beardsley, Gianni Bernardi, Judd D. Bowman, Richard F. Bradley, Jacob Burba, Chris L. Carilli, Carina Cheng, David R. DeBoer, Matt Dexter, Eloy de Lera Acedo, Nicolas Fagnoni, Randall Fritz, Steve R. Furlanetto, Brian Glendenning, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Bryna J. Hazelton, Jacqueline N. Hewitt, Jack Hickish, Daniel C. Jacobs, Austin Julius, Joshua Kerrigan, Piyanat Kittiwisit, Saul A. Kohn, Matthew Kolopanis, Paul La Plante, Telalo Lekalake, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary E. Martinot, Eunice Matsetela, Andrei Mesinger, Mathakane Molewa, Miguel F. Morales, Tshegofalang Mosiane, Steven G. Murray, Abraham R. Neben, Aaron R. Parsons, Nipanjana Patra, Samantha Pieterse, Jonathan C. Pober, Nima Razavi-Ghods, Jon Ringuette, James Robnett, Kathryn Rosie, Peter Sims, Craig Smith, Angelo Syce, Nithyanandan Thyagarajan, Peter K. G. Williams, and Haoxuan Zheng. Mitigating Internal Instrument Coupling for 21 cm Cosmology. II. A Method Demonstration with the Hydrogen Epoch of Reionization Array. ApJ, 888(2):70, January 2020b. doi: 10.3847/1538-4357/ab5e8a.

Matthew Kolopanis, Daniel C. Jacobs, Carina Cheng, Aaron R. Parsons, Saul A. Kohn, Jonathan C. Pober, James E. Aguirre, Zaki S. Ali, Gianni Bernardi, Richard F. Bradley, Chris L. Carilli, David R. DeBoer, Matthew R. Dexter, Joshua S. Dillon, Joshua Kerrigan, Pat Klima, Adrian Liu, David H. E. MacMahon, David F. Moore, Nithyanandan Thyagarajan, Chuneeta D. Nunhokee, William P. Walbrugh, and Andre Walker. A Simplified, Lossless Reanalysis of PAPER-64. *Astrophys. J.*, 883(2), 2019. doi: 10.3847/1538-4357/ab3e3a.

Matthew Kolopanis, Jonathan C Pober, Daniel C Jacobs, and Samantha McGraw. New eor power spectrum limits from mwa phase ii using the delay spectrum method and novel systematic rejection. *Monthly Notices of the Royal Astronomical Society*, 521(4):5120–5138, March 2023. ISSN 1365-2966. doi: 10.1093/mnras/stad845. URL http://dx.doi.org/10.1093/mnras/stad845.

Helge Kragh. Big bang cosmology. In *Cosmology*, pages 371–390. CRC Press, 2023.

W. Li, J. C. Pober, B. J. Hazelton, N. Barry, M. F. Morales, I. Sullivan, A. R. Parsons, Z. S. Ali, J. S. Dillon, A. P. Beardsley, J. D. Bowman, F. Briggs, R. Byrne, P. Carroll, B. Crosse,

D. Emrich, A. Ewall-Wice, L. Feng, T. M. O. Franzen, J. N. Hewitt, L. Horsley, D. C. Jacobs, M. Johnston-Hollitt, C. Jordan, R. C. Joseph, D. L. Kaplan, D. Kenney, H. Kim, P. Kittiwisit, A. Lanman, J. Line, B. McKinley, D. A. Mitchell, S. Murray, A. Neben, A. R. Offringa, D. Pallot, S. Paul, B. Pindor, P. Procopio, M. Rahimi, J. Riding, S. K. Sethi, N. Udaya Shankar, K. Steele, R. Subrahmanian, M. Tegmark, N. Thyagarajan, S. J. Tingay, C. Trott, M. Walker, R. B. Wayth, R. L. Webster, A. Williams, C. Wu, and S. Wyithe. Comparing redundant and sky-model-based interferometric calibration: A first look with phase ii of the mwa. *The Astrophysical Journal*, 863(2):170, August 2018. ISSN 1538-4357. doi: 10.3847/1538-4357/aad3c3. URL http://dx.doi.org/10.3847/1538-4357/aad3c3.

W. Li, J. C. Pober, N. Barry, B. J. Hazelton, M. F. Morales, C. M. Trott, A. Lanman, M. Wilensky, I. Sullivan, A. P. Beardsley, T. Booler, J. D. Bowman, R. Byrne, B. Crosse, D. Emrich, T. M. O. Franzen, K. Hasegawa, L. Horsley, M. Johnston-Hollitt, D. C. Jacobs, C. H. Jordan, R. C. Joseph, T. Kaneuji, D. L. Kaplan, D. Kenney, K. Kubota, J. Line, C. Lynch, B. McKinley, D. A. Mitchell, S. Murray, D. Pallot, B. Pindor, M. Rahimi, J. Riding, G. Sleap, K. Steele, K. Takahashi, S. J. Tingay, M. Walker, R. B. Wayth, R. L. Webster, A. Williams, C. Wu, J. S. B. Wyithe, S. Yoshiura, and Q. Zheng. First Season MWA Phase II Epoch of Reionization Power Spectrum Results at Redshift 7. *The Astrophysical Journal*, 887(2):141, December 2019. doi: 10.3847/1538-4357/ab55e4.

J. L. B. Line, B. McKinley, J. Rasti, M. Bhardwaj, R. B. Wayth, R. L. Webster, D. Ung, D. Emrich, L. Horsley, A. Beardsley, B. Crosse, T. M. O. Franzen, B. M. Gaensler, M. Johnston-Hollitt, D. L. Kaplan, D. Kenney, M. F. Morales, D. Pallot, K. Steele, S. J. Tingay, C. M. Trott, M. Walker, A. Williams, and C. Wu. In situ measurement of MWA primary beam variation using ORBCOMM. PASA, 35:e045, December 2018. doi: 10.1017/pasa.2018.30.

Adrian Liu and J. Richard Shaw. Data analysis for precision 21 cm cosmology. *Publications of the Astronomical Society of the Pacific*, 132(1012):062001, Apr 2020. ISSN 1538-3873. doi: 10.1088/1538-3873/ab5bfd. URL http://dx.doi.org/10.1088/1538-3873/ab5bfd.

Adrian Liu, Max Tegmark, Scott Morrison, Andrew Lutomirski, and Matias Zaldarriaga. Precision calibration of radio interferometers using redundant baselines: Redundant baseline calibration

of arrays. *Monthly Notices of the Royal Astronomical Society*, 408(2):1029–1050, July 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2010.17174.x. URL http://dx.doi.org/10.1111/j.1365-2966.2010.17174.x.

Adrian Liu, Aaron R. Parsons, and Cathryn M. Trott. Epoch of reionization window. i. mathematical formalism. *Physical Review D*, 90(2), July 2014a. ISSN 1550-2368. doi: 10.1103/physrevd.90.023018. URL http://dx.doi.org/10.1103/PhysRevD.90.023018.

Adrian Liu, Aaron R. Parsons, and Cathryn M. Trott. Epoch of reionization window. ii. statistical methods for foreground wedge reduction. *Physical Review D*, 90(2), July 2014b. ISSN 1550-2368. doi: 10.1103/physrevd.90.023019. URL http://dx.doi.org/10.1103/PhysRevD.90.023019.

Grant J. Mathews, Motohiko Kusakabe, and Toshitaka Kajino. Introduction to big bang nucleosynthesis and modern cosmology. *International Journal of Modern Physics E*, 26(08):1741001, August 2017. ISSN 1793-6608. doi: 10.1142/s0218301317410014. URL http://dx.doi.org/10.1142/S0218301317410014.

Garrelt Mellema, Léon V. E. Koopmans, Filipe A. Abdalla, Gianni Bernardi, Benedetta Ciardi, Soobash Daiboo, A. G. de Bruyn, Kanan K. Datta, Heino Falcke, Andrea Ferrara, Ilian T. Iliev, Fabio Iocco, Vibor Jelić, Hannes Jensen, Ronniy Joseph, Panos Labroupoulos, Avery Meiksin, Andrei Mesinger, André R. Offringa, V. N. Pandey, Jonathan R. Pritchard, Mario G. Santos, Dominik J. Schwarz, Benoit Semelin, Harish Vedantham, Sarod Yatawatta, and Saleem Zaroubi. Reionization and the Cosmic Dawn with the Square Kilometre Array. *Experimental Astronomy*, 36(1-2):235–318, August 2013. doi: 10.1007/s10686-013-9334-5.

F G Mertens, M Mevius, L V E Koopmans, A R Offringa, G Mellema, S Zaroubi, M A Brentjens, H Gan, B K Gehlot, V N Pandey, A M Sardarabadi, H K Vedantham, S Yatawatta, K M B Asad, B Ciardi, E Chapman, S Gazagnes, R Ghara, A Ghosh, S K Giri, I T Iliev, V Jelić, R Kooistra, R Mondal, J Schaye, and M B Silva. Improved upper limits on the 21 cm signal power spectrum of neutral hydrogen at z = 9.1 from lofar. *Monthly Notices of the Royal Astronomical Society*, 493(2):1662–1685, February 2020. ISSN 1365-2966. doi: 10.1093/mnras/staa327. URL http://dx.doi.org/10.1093/mnras/staa327.

Miguel F. Morales and Jacqueline Hewitt. Toward epoch of reionization measurements with wide-field radio observations. *The Astrophysical Journal*, 615(1):7–18, November 2004. ISSN 1538-4357. doi: 10.1086/424437. URL http://dx.doi.org/10.1086/424437.

Miguel F. Morales and Michael Matejek. Software holography: interferometric data analysis for the challenges of next generation observatories. *Monthly Notices of the Royal Astronomical Society*, 400(4):1814–1820, December 2009. doi: 10.1111/j.1365-2966.2009.15537.x.

Miguel F. Morales and J. Stuart B. Wyithe. Reionization and Cosmology with 21-cm Fluctuations. ARA&A, 48:127–171, September 2010. doi: 10.1146/annurev-astro-081309-130936.

Miguel F. Morales, Bryna Hazelton, Ian Sullivan, and Adam Beardsley. Four fundamental foreground power spectrum shapes for 21 cm cosmology observations. *The Astrophysical Journal*, 752(2):137, June 2012. ISSN 1538-4357. doi: 10.1088/0004-637x/752/2/137. URL http://dx.doi.org/10.1088/0004-637X/752/2/137.

Miguel F Morales, Adam Beardsley, Jonathan Pober, Nichole Barry, Bryna Hazelton, Daniel Jacobs, and Ian Sullivan. Understanding the diversity of 21 cm cosmology analyses. *Monthly Notices of the Royal Astronomical Society*, 483(2):2207–2216, October 2018. ISSN 1365-2966. doi: 10.1093/mnras/sty2844. URL http://dx.doi.org/10.1093/mnras/sty2844.

P. Moreira, J. Serrano, T. Wlostowski, P. Loschmidt, and G. Gaderer. White rabbit: Sub-nanosecond timing distribution over ethernet. *2009 International Symposium on Precision Clock Synchronization for Measurement, Control and Communication*, pages 1–5, 2009.

Steven G. Murray and C. M. Trott. The Effect of Baseline Layouts on the Epoch of Reionization Foreground Wedge: A Semianalytical Approach. ApJ, 869(1):25, December 2018. doi: 10.3847/1538-4357/aaebfa.

Abraham R. Neben, Richard F. Bradley, Jacqueline N. Hewitt, David R. DeBoer, Aaron R. Parsons, James E. Aguirre, Zaki S. Ali, Carina Cheng, Aaron Ewall-Wice, Nipanjana Patra, Nithyanandan Thyagarajan, Judd Bowman, Roger Dickenson, Joshua S. Dillon, Phillip Doolittle, Dennis Egan, Mike Hedrick, Daniel C. Jacobs, Saul A. Kohn, Patricia J. Klima, Kavilan Moodley, Benjamin R. B. Saliwanchik, Patrick Schaffner, John Shelton, H. A. Taylor, Rusty Taylor, Max Tegmark,

Butch Wirt, and Haoxuan Zheng. The Hydrogen Epoch of Reionization Array Dish. I. Beam Pattern Measurements and Science Implications. ApJ, 826(2):199, August 2016. doi: 10.3847/0004-637X/826/2/199.

Laura B. Newburgh, Graeme E. Addison, Mandana Amiri, Kevin Bandura, J. Richard Bond, Liam Connor, Jean-François Cliche, Greg Davis, Meiling Deng, Nolan Denman, and et al. Calibrating chime: a new radio interferometer to probe dark energy. *Ground-based and Airborne Telescopes V*, Jul 2014. doi: 10.1117/12.2056962. URL http://dx.doi.org/10.1117/12.2056962.

Chuneeta D. Nunhokee, Aaron R. Parsons, Nicholas S. Kern, Bojan Nikolic, Jonathan C. Pober, Gianni Bernardi, Chris L. Carilli, Zara Abdurashidova, James E. Aguirre, Paul Alexander, Zaki S. Ali, Yanga Balfour, Adam P. Beardsley, Tashalee S. Billings, Judd D. Bowman, Richard F. Bradley, Jacob Burba, Carina Cheng, David R. DeBoer, Matt Dexter, Eloy de Lera Acedo, Joshua S. Dillon, Aaron Ewall-Wice, Nicolas Fagnoni, Randall Fritz, Steve R. Furlanetto, Kingsley Gale-Sides, Brian Glendenning, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Bryna J. Hazelton, Jacqueline N. Hewitt, Daniel C. Jacobs, Austin Julius, Joshua Kerrigan, Piyanat Kittiwisit, Saul A. Kohn, Matthew Kolopanis, Adam Lanman, Paul La Plante, Telalo Lekalake, Adrian Liu, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary E. Martinot, Eunice Matsetela, Andrei Mesinger, Mathakane Molewa, Miguel F. Morales, Tshegofalang Mosiane, Abraham R. Neben, Nipanjana Patra, Samantha Pieterse, Nima Razavi-Ghods, Jon Ringuette, James Robnett, Kathryn Rosie, Peter Sims, Craig Smith, Angelo Syce, Nithyanandan Thyagarajan, Peter K. G. Williams, and Haoxuan Zheng. Measuring hera's primary beam in situ: Methodology and first results. *The Astrophysical Journal*, 897(1):5, June 2020. ISSN 1538-4357. doi: 10.3847/1538-4357/ab9634. URL http://dx.doi.org/10.3847/1538-4357/ab9634.

Gregory Paciga, Tzu-Ching Chang, Yashwant Gupta, Rajaram Nityanada, Julia Odegova, Ue-Li Pen, Jeffrey B. Peterson, Jayanta Roy, and Kris Sigurdson. The GMRT Epoch of Reionization experiment: a new upper limit on the neutral hydrogen power spectrum at z$\sim$ 8.6. *Monthly Notices of the Royal Astronomical Society*, 413(2):1174–1183, 04 2011. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2011.18208.x. URL https://doi.org/10.1111/j.1365-2966.2011.18208.x.

Gregory Paciga, Joshua G. Albert, Kevin Bandura, Tzu-Ching Chang, Yashwant Gupta, Christopher Hirata, Julia Odegova, Ue-Li Pen, Jeffrey B. Peterson, Jayanta Roy, J. Richard Shaw, Kris Sigurdson, and Tabitha Voytek. A simulation-calibrated limit on the H I power spectrum from the GMRT Epoch of Reionization experiment. MNRAS, 433(1):639–647, July 2013. doi: 10.1093/mnras/stt753.

Aaron R. Parsons, Donald C. Backer, Griffin S. Foster, Melvyn C. H. Wright, Richard F. Bradley, Nicole E. Gugliucci, Chaitali R. Parashare, Erin E. Benoit, James E. Aguirre, Daniel C. Jacobs, Chris L. Carilli, David Herne, Mervyn J. Lynch, Jason R. Manley, and Daniel J. Werthimer. The Precision Array for Probing the Epoch of Re-ionization: Eight Station Results. AJ, 139(4): 1468–1480, April 2010a. doi: 10.1088/0004-6256/139/4/1468.

Aaron R. Parsons, Donald C. Backer, Griffin S. Foster, Melvyn C. H. Wright, Richard F. Bradley, Nicole E. Gugliucci, Chaitali R. Parashare, Erin E. Benoit, James E. Aguirre, Daniel C. Jacobs, Chris L. Carilli, David Herne, Mervyn J. Lynch, Jason R. Manley, and Daniel J. Werthimer. The Precision Array for Probing the Epoch of Re-ionization: Eight Station Results. AJ, 139(4): 1468–1480, April 2010b. doi: 10.1088/0004-6256/139/4/1468.

Aaron R. Parsons, Jonathan C. Pober, James E. Aguirre, Christopher L. Carilli, Daniel C. Jacobs, and David F. Moore. A per-baseline, delay-spectrum technique for accessing the 21 cm cosmic reionization signature. *The Astrophysical Journal*, 756(2):165, August 2012. ISSN 1538-4357. doi: 10.1088/0004-637x/756/2/165. URL http://dx.doi.org/10.1088/0004-637X/756/2/165.

A. H. Patil, S. Yatawatta, L. V. E. Koopmans, A. G. de Bruyn, M. A. Brentjens, S. Zaroubi, K. M. B. Asad, M. Hatef, V. Jelić, M. Mevius, A. R. Offringa, V. N. Pandey, H. Vedantham, F. B. Abdalla, W. N. Brouw, E. Chapman, B. Ciardi, B. K. Gehlot, A. Ghosh, G. Harker, I. T. Iliev, K. Kakiichi, S. Majumdar, G. Mellema, M. B. Silva, J. Schaye, D. Vrbanec, and S. J. Wijnholds. Upper limits on the 21 cm epoch of reionization power spectrum from one night with lofar. *The Astrophysical Journal*, 838(1):65, March 2017. ISSN 1538-4357. doi: 10.3847/1538-4357/aa63e7. URL http://dx.doi.org/10.3847/1538-4357/aa63e7.

Ue-Li Pen, Tzu-Ching Chang, Christopher M. Hirata, Jeffrey B. Peterson, Jayanta Roy, Yashwant Gupta, Julia Odegova, and Kris Sigurdson. The gmrt eor experiment: limits on polarized sky

brightness at 150 mhz. *Monthly Notices of the Royal Astronomical Society*, 399(1):181–194, October 2009. ISSN 1365-2966. doi: 10.1111/j.1365-2966.2009.14980.x. URL http://dx.doi.org/10.1111/j.1365-2966.2009.14980.x.

Jonathan C. Pober, Aaron R. Parsons, Daniel C. Jacobs, James E. Aguirre, Richard F. Bradley, Chris L. Carilli, Nicole E. Gugliucci, David F. Moore, and Chaitali R. Parashare. A technique for primary beam calibration of drift-scanning, wide-field antenna elements. *The Astronomical Journal*, 143(2):53, January 2012. ISSN 1538-3881. doi: 10.1088/0004-6256/143/2/53. URL http://dx.doi.org/10.1088/0004-6256/143/2/53.

D C Price, L J Greenhill, A Fialkov, G Bernardi, H Garsden, B R Barsdell, J Kocz, M M Anderson, S A Bourke, J Craig, and et al. Design and characterization of the large-aperture experiment to detect the dark age (leda) radiometer systems. *Monthly Notices of the Royal Astronomical Society*, May 2018. ISSN 1365-2966. doi: 10.1093/mnras/sty1244. URL http://dx.doi.org/10.1093/mnras/sty1244.

Jonathan R Pritchard and Abraham Loeb. 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901, Jul 2012. ISSN 1361-6633. doi: 10.1088/0034-4885/75/8/086901. URL http://dx.doi.org/10.1088/0034-4885/75/8/086901.

M Rahimi, B Pindor, J L B Line, N Barry, C M Trott, R L Webster, C H Jordan, M Wilensky, S Yoshiura, A Beardsley, and et al. Epoch of reionization power spectrum limits from murchison widefield array data targeted at eor1 field. *Monthly Notices of the Royal Astronomical Society*, 508(4):5954–5971, Oct 2021. ISSN 1365-2966. doi: 10.1093/mnras/stab2918. URL http://dx.doi.org/10.1093/mnras/stab2918.

N. Razavi-Ghods, S. H. Carey, J. A. Ely, and P. F Scott. Hera rf and calibration system design. In *2017 Progress In Electromagnetics Research Symposium - Spring (PIERS)*, pages 1544–1548, 2017. doi: 10.1109/PIERS.2017.8261992.

M. Remazeilles, C. Dickinson, A. J. Banday, M. A. Bigot-Sazy, and T. Ghosh. An improved source-subtracted and destriped 408 mhz all-sky map, 2015a.

M. Remazeilles, C. Dickinson, A. J. Banday, M. A. Bigot-Sazy, and T. Ghosh. An improved source-subtracted and destriped 408 mhz all-sky map, 2015b.

Benjamin R. B. Saliwanchik, Aaron Ewall-Wice, Devin Crichton, Emily R. Kuhn, Deniz Ölçek, Kevin Bandura, Martin Bucher, Tzu-Ching Chang, H. Cynthia Chiang, Kit Gerodias, Kabelo Kesebonye, Vincent MacKay, Kavilan Moodley, Laura B. Newburgh, Viraj Nistane, Jeffrey B. Peterson, Elizabeth Pieters, Carla Pieterse, Keith Vanderlinde, Jonathan L. Sievers, Amanda Weltman, and Dallas Wulf. Mechanical and optical design of the hirax radio telescope, 2021.

Mario G. Santos, Asantha Cooray, and Lloyd Knox. Multifrequency analysis of 21 centimeter fluctuations from the era of reionization. *The Astrophysical Journal*, 625(2):575–587, Jun 2005. ISSN 1538-4357. doi: 10.1086/429857. URL http://dx.doi.org/10.1086/429857.

Robert J. Selina, Eric J. Murphy, Mark McKinnon, Anthony Beasley, Bryan Butler, Chris Carilli, Barry Clark, Alan Erickson, Wes Grammer, James Jackson, Brian Kent, Brian Mason, Matthew Morgan, Omar Ojeda, William Shillue, Silver Sturgis, and Denis Urbain. The next generation very large array: A technical overview, 2018.

Hayato Shimabukuro, Kenji Hasegawa, Akira Kuchinomachi, Hidenobu Yajima, and Shintaro Yoshiura. Exploring the cosmic dawn and epoch of reionization with the 21 cm line. *Publications of the Astronomical Society of Japan*, 75(Supplement 1), 2021. doi: 10.1093/pasj/psac042.

Pyxie Star. Implementation of van vleck correction for the mwa. Memo, 2020. URL https://github.com/EoRImaging/Memos/blob/master/PDFs/007_Van_Vleck_A.pdf.

Dara Storer, Joshua S. Dillon, Daniel C. Jacobs, Miguel F. Morales, Bryna J. Hazelton, Aaron Ewall-Wice, Zara Abdurashidova, James E. Aguirre, Paul Alexander, Zaki S. Ali, Yanga Balfour, Adam P. Beardsley, Gianni Bernardi, Tashalee S. Billings, Judd D. Bowman, Richard F. Bradley, Philip Bull, Jacob Burba, Steven Carey, Chris L. Carilli, Carina Cheng, David R. DeBoer, Eloy de Lera Acedo, Matt Dexter, Scott Dynes, John Ely, Nicolas Fagnoni, Randall Fritz, Steven R. Furlanetto, Kingsley Gale-Sides, Brian Glendenning, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Jacqueline N. Hewitt, Jack Hickish, Tian Huang, Alec Josaitis, Austin Julius, MacCalvin Kariseb, Nicholas S. Kern, Joshua Kerrigan, Piyanat

Kittiwisit, Saul A. Kohn, Matthew Kolopanis, Adam Lanman, Paul La Plante, Adrian Liu, Anita Loots, David MacMahon, Lourence Malan, Cresshim Malgas, Zachary E. Martinot, Andrei Mesinger, Mathakane Molewa, Tshegofalang Mosiane, Steven G. Murray, Abraham R. Neben, Bojan Nikolic, Chuneeta Devi Nunhokee, Aaron R. Parsons, Robert Pascua, Nipanjana Patra, Samantha Pieterse, Jonathan C. Pober, Nima Razavi-Ghods, Daniel Riley, James Robnett, Kathryn Rosie, Mario G. Santos, Peter Sims, Saurabh Singh, Craig Smith, Jianrong Tan, Nithyanandan Thyagarajan, Peter K. G. Williams, and Haoxuan Zheng. Automated detection of antenna malfunctions in large-n interferometers: A case study with the hydrogen epoch of reionization array. *Radio Science*, 57(1), January 2022. ISSN 1944-799X. doi: 10.1029/2021rs007376. URL http://dx.doi.org/10.1029/2021RS007376.

I. S. Sullivan, M. F. Morales, B. J. Hazelton, W. Arcus, D. Barnes, G. Bernardi, F. H. Briggs, J. D. Bowman, J. D. Bunton, R. J. Cappallo, B. E. Corey, A. Deshpande, L. Desouza, D. Emrich, B. M. Gaensler, R. Goeke, L. J. Greenhill, D. Herne, J. N. Hewitt, M. Johnston-Hollitt, D. L. Kaplan, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, C. J. Lonsdale, M. J. Lynch, S. R. McWhirter, D. A. Mitchell, E. Morgan, D. Oberoi, S. M. Ord, J. Pathikulangara, T. Prabu, R. A. Remillard, A. E.E. Rogers, A. Roshi, J. E. Salah, R. J. Sault, N. Udaya Shankar, K. S. Srivani, J. Stevens, R. Subrahmanyan, S. J. Tingay, R. B. Wayth, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams, and J. S.B. Wyithe. Fast Holographic Deconvolution: A New Technique for Precision Radio Interferometry. *Astrophys. J.*, 759(1):17, 2012. doi: 10.1088/0004-637X/759/1/17.

Greg B Taylor, Chris Luke Carilli, and Richard A Perley. Synthesis imaging in radio astronomy ii. *Synthesis Imaging in Radio Astronomy II*, 180, 1999.

A Richard Thompson, James M Moran, and George W Swenson. *Interferometry and synthesis in radio astronomy*. Springer Nature, 2017.

S. J. Tingay, R. Goeke, J. D. Bowman, D. Emrich, S. M. Ord, D. A. Mitchell, M. F. Morales, T. Booler, B. Crosse, R. B. Wayth, C. J. Lonsdale, S. Tremblay, D. Pallot, T. Colegate, A. Wicenec, N. Kudryavtseva, W. Arcus, D. Barnes, G. Bernardi, F. Briggs, S. Burns, J. D. Bunton, R. J. Cappallo, B. E. Corey, A. Deshpande, L. Desouza, B. M. Gaensler, L. J. Greenhill,

P. J. Hall, B. J. Hazelton, D. Herne, J. N. Hewitt, M. Johnston-Hollitt, D. L. Kaplan, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, M. J. Lynch, B. Mckinley, S. R. Mcwhirter, E. Morgan, D. Oberoi, J. Pathikulangara, T. Prabu, R. A. Remillard, A. E. E. Rogers, A. Roshi, J. E. Salah, R. J. Sault, N. Udaya-Shankar, F. Schlagenhaufer, K. S. Srivani, J. Stevens, R. Subrahmanyan, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams, and J. S. B. Wyithe. The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies. PASA, 30:e007, January 2013. doi: 10.1017/pasa.2012.007.

Cathryn M. Trott, Randall B. Wayth, and Steven J. Tingay. The impact of point-source subtraction residuals on 21 cm epoch of reionization estimation. *The Astrophysical Journal*, 757(1):101, September 2012. ISSN 1538-4357. doi: 10.1088/0004-637x/757/1/101. URL http://dx.doi.org/10.1088/0004-637X/757/1/101.

Cathryn M Trott, C H Jordan, S Midgley, N Barry, B Greig, B Pindor, J H Cook, G Sleap, S J Tingay, D Ung, P Hancock, A Williams, J Bowman, R Byrne, A Chokshi, B J Hazelton, K Hasegawa, D Jacobs, R C Joseph, W Li, J L B Line, C Lynch, B McKinley, D A Mitchell, M F Morales, M Ouchi, J C Pober, M Rahimi, K Takahashi, R B Wayth, R L Webster, M Wilensky, J S B Wyithe, S Yoshiura, Z Zhang, and Q Zheng. Deep Multiredshift Limits on Epoch of Reionization 21 cm Power Spectra from Four Seasons of Murchison Widefield Array Observations. *Mon. Not. R. Astron. Soc.*, 493(4), 2020. doi: 10.1093/mnras/staa414.

M. P. van Haarlem, M. W. Wise, A. W. Gunst, G. Heald, J. P. McKean, J. W. T. Hessels, A. G. de Bruyn, R. Nijboer, J. Swinbank, R. Fallows, M. Brentjens, A. Nelles, R. Beck, H. Falcke, R. Fender, J. Hörandel, L. V. E. Koopmans, G. Mann, G. Miley, H. Röttgering, B. W. Stappers, R. A. M. J. Wijers, S. Zaroubi, M. van den Akker, A. Alexov, J. Anderson, K. Anderson, A. van Ardenne, M. Arts, A. Asgekar, I. M. Avruch, F. Batejat, L. Bähren, M. E. Bell, M. R. Bell, I. van Bemmel, P. Bennema, M. J. Bentum, G. Bernardi, P. Best, L. Bîrzan, A. Bonafede, A. J. Boonstra, R. Braun, J. Bregman, F. Breitling, R. H. van de Brink, J. Broderick, P. C. Broekema, W. N. Brouw, M. Brüggen, H. R. Butcher, W. van Cappellen, B. Ciardi, T. Coenen, J. Conway, A. Coolen, A. Corstanje, S. Damstra, O. Davies, A. T. Deller, R. J. Dettmar, G. van Diepen, K. Dijkstra, P. Donker, A. Doorduin, J. Dromer, M. Drost, A. van Duin, J. Eislöffel,

J. van Enst, C. Ferrari, W. Frieswijk, H. Gankema, M. A. Garrett, F. de Gasperin, M. Gerbers, E. de Geus, J. M. Grießmeier, T. Grit, P. Gruppen, J. P. Hamaker, T. Hassall, M. Hoeft, H. A. Holties, A. Horneffer, A. van der Horst, A. van Houwelingen, A. Huijgen, M. Iacobelli, H. Intema, N. Jackson, V. Jelic, A. de Jong, E. Juette, D. Kant, A. Karastergiou, A. Koers, H. Kollen, V. I. Kondratiev, E. Kooistra, Y. Koopman, A. Koster, M. Kuniyoshi, M. Kramer, G. Kuper, P. Lambropoulos, C. Law, J. van Leeuwen, J. Lemaitre, M. Loose, P. Maat, G. Macario, S. Markoff, J. Masters, R. A. McFadden, D. McKay-Bukowski, H. Meijering, H. Meulman, M. Mevius, E. Middelberg, R. Millenaar, J. C. A. Miller-Jones, R. N. Mohan, J. D. Mol, J. Morawietz, R. Morganti, D. D. Mulcahy, E. Mulder, H. Munk, L. Nieuwenhuis, R. van Nieuwpoort, J. E. Noordam, M. Norden, A. Noutsos, A. R. Offringa, H. Olofsson, A. Omar, E. Orrú, R. Overeem, H. Paas, M. Pandey-Pommier, V. N. Pandey, R. Pizzo, A. Polatidis, D. Rafferty, S. Rawlings, W. Reich, J. P. de Reijer, J. Reitsma, G. A. Renting, P. Riemers, E. Rol, J. W. Romein, J. Roosjen, M. Ruiter, A. Scaife, K. van der Schaaf, B. Scheers, P. Schellart, A. Schoenmakers, G. Schoonderbeek, M. Serylak, A. Shulevski, J. Sluman, O. Smirnov, C. Sobey, H. Spreeuw, M. Steinmetz, C. G. M. Sterks, H. J. Stiepel, K. Stuurwold, M. Tagger, Y. Tang, C. Tasse, I. Thomas, S. Thoudam, M. C. Toribio, B. van der Tol, O. Usov, M. van Veelen, A. J. van der Veen, S. ter Veen, J. P. W. Verbiest, R. Vermeulen, N. Vermaas, C. Vocks, C. Vogt, M. de Vos, E. van der Wal, R. van Weeren, H. Weggemans, P. Weltevrede, S. White, S. J. Wijnholds, T. Wilhelmsson, O. Wucknitz, S. Yatawatta, P. Zarka, A. Zensus, and J. van Zwieten. LOFAR: The LOw-Frequency ARray. A&A, 556:A2, August 2013a. doi: 10.1051/0004-6361/201220873.

M. P. van Haarlem, M. W. Wise, A. W. Gunst, G. Heald, J. P. McKean, J. W. T. Hessels, A. G. de Bruyn, R. Nijboer, J. Swinbank, R. Fallows, M. Brentjens, A. Nelles, R. Beck, H. Falcke, R. Fender, J. Hörandel, L. V. E. Koopmans, G. Mann, G. Miley, H. Röttgering, B. W. Stappers, R. A. M. J. Wijers, S. Zaroubi, M. van den Akker, A. Alexov, J. Anderson, K. Anderson, A. van Ardenne, M. Arts, A. Asgekar, I. M. Avruch, F. Batejat, L. Bähren, M. E. Bell, M. R. Bell, I. van Bemmel, P. Bennema, M. J. Bentum, G. Bernardi, P. Best, L. Bîrzan, A. Bonafede, A. J. Boonstra, R. Braun, J. Bregman, F. Breitling, R. H. van de Brink, J. Broderick, P. C. Broekema, W. N. Brouw, M. Brüggen, H. R. Butcher, W. van Cappellen, B. Ciardi, T. Coenen, J. Conway, A. Coolen, A. Corstanje, S. Damstra, O. Davies, A. T. Deller, R. J. Dettmar, G. van

Diepen, K. Dijkstra, P. Donker, A. Doorduin, J. Dromer, M. Drost, A. van Duin, J. Eislöffel, J. van Enst, C. Ferrari, W. Frieswijk, H. Gankema, M. A. Garrett, F. de Gasperin, M. Gerbers, E. de Geus, J. M. Grießmeier, T. Grit, P. Gruppen, J. P. Hamaker, T. Hassall, M. Hoeft, H. A. Holties, A. Horneffer, A. van der Horst, A. van Houwelingen, A. Huijgen, M. Iacobelli, H. Intema, N. Jackson, V. Jelic, A. de Jong, E. Juette, D. Kant, A. Karastergiou, A. Koers, H. Kollen, V. I. Kondratiev, E. Kooistra, Y. Koopman, A. Koster, M. Kuniyoshi, M. Kramer, G. Kuper, P. Lambropoulos, C. Law, J. van Leeuwen, J. Lemaitre, M. Loose, P. Maat, G. Macario, S. Markoff, J. Masters, R. A. McFadden, D. McKay-Bukowski, H. Meijering, H. Meulman, M. Mevius, E. Middelberg, R. Millenaar, J. C. A. Miller-Jones, R. N. Mohan, J. D. Mol, J. Morawietz, R. Morganti, D. D. Mulcahy, E. Mulder, H. Munk, L. Nieuwenhuis, R. van Nieuwpoort, J. E. Noordam, M. Norden, A. Noutsos, A. R. Offringa, H. Olofsson, A. Omar, E. Orrú, R. Overeem, H. Paas, M. Pandey-Pommier, V. N. Pandey, R. Pizzo, A. Polatidis, D. Rafferty, S. Rawlings, W. Reich, J. P. de Reijer, J. Reitsma, G. A. Renting, P. Riemers, E. Rol, J. W. Romein, J. Roosjen, M. Ruiter, A. Scaife, K. van der Schaaf, B. Scheers, P. Schellart, A. Schoenmakers, G. Schoonderbeek, M. Serylak, A. Shulevski, J. Sluman, O. Smirnov, C. Sobey, H. Spreeuw, M. Steinmetz, C. G. M. Sterks, H. J. Stiepel, K. Stuurwold, M. Tagger, Y. Tang, C. Tasse, I. Thomas, S. Thoudam, M. C. Toribio, B. van der Tol, O. Usov, M. van Veelen, A. J. van der Veen, S. ter Veen, J. P. W. Verbiest, R. Vermeulen, N. Vermaas, C. Vocks, C. Vogt, M. de Vos, E. van der Wal, R. van Weeren, H. Weggemans, P. Weltevrede, S. White, S. J. Wijnholds, T. Wilhelmsson, O. Wucknitz, S. Yatawatta, P. Zarka, A. Zensus, and J. van Zwieten. LOFAR: The LOw-Frequency ARray. A&A, 556:A2, August 2013b. doi: 10.1051/0004-6361/201220873.

Keith Vanderlinde, Adrian Liu, Bryan Gaensler, Dick Bond, Gary Hinshaw, Cherry Ng, Cynthia Chiang, Ingrid Stairs, Jo-Anne Brown, Jonathan Sievers, Juan Mena, Kendrick Smith, Kevin Bandura, Kiyoshi Masui, Kristine Spekkens, Leo Belostotski, Matt Dobbs, Neil Turok, Patrick Boyle, Michael Rupen, Tom Landecker, Ue-Li Pen, and Victoria Kaspi. The Canadian Hydrogen Observatory and Radio-transient Detector (CHORD). In *Canadian Long Range Plan for Astronomy and Astrophysics White Papers*, volume 2020, page 28, October 2019. doi: 10.5281/zenodo.3765414.

Harish Vedantham, N. Udaya Shankar, and Ravi Subrahmanyan. Imaging the Epoch of Reion-

ization: Limitations from Foreground Confusion and Imaging Algorithms. ApJ, 745(2):176, February 2012. doi: 10.1088/0004-637X/745/2/176.

R. B. Wayth, E. Lenc, M. E. Bell, J. R. Callingham, K. S. Dwarakanath, T. M.O. Franzen, B. Q. For, B. Gaensler, P. Hancock, L. Hindson, N. Hurley-Walker, C. A. Jackson, M. Johnston-Hollitt, A. D. Kapińska, B. McKinley, J. Morgan, A. R. Offringa, P. Procopio, L. Staveley-Smith, C. Wu, Q. Zheng, C. M. Trott, G. Bernardi, J. D. Bowman, F. Briggs, R. J. Cappallo, B. E. Corey, A. A. Deshpande, D. Emrich, R. Goeke, L. J. Greenhill, B. J. Hazelton, D. L. Kaplan, J. C. Kasper, E. Kratzenberg, C. J. Lonsdale, M. J. Lynch, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, D. Oberoi, S. M. Ord, T. Prabu, A. E.E. Rogers, A. Roshi, N. Udaya Shankar, K. S. Srivani, R. Subrahmanyan, S. J. Tingay, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, and C. L. Williams. GLEAM: The GaLactic and Extragalactic All-Sky MWA Survey. *Publ. Astron. Soc. Aust.*, 32, 2015. doi: 10.1017/pasa.2015.26.

Sarah V. White, Thomas M. O Franzen, Chris J. Riseley, O. Ivy Wong, Anna D. Kapińska, Natasha Hurley-Walker, Joseph R. Callingham, Kshitij Thorat, Chen Wu, Paul Hancock, and et al. The gleam 4-jy (g4jy) sample: I. definition and the catalogue. *Publications of the Astronomical Society of Australia*, 37, 2020. ISSN 1448-6083. doi: 10.1017/pasa.2020.9. URL http://dx.doi.org/10.1017/pasa.2020.9.

Lily R. Whitler, Adam Beardsley, and Daniel Jacobs. The Effects of RFI on 21-cm Measurements of the Epoch of Reionization. In *American Astronomical Society Meeting Abstracts #233*, volume 233 of *American Astronomical Society Meeting Abstracts*, page 349.17, January 2019.

Mark H. Wieringa. An investigation of the telescope based calibration methods 'redundancy' and 'self-cal'. *Experimental Astronomy*, 2:203–225, 1992. URL https://api.semanticscholar.org/CorpusID:121356049.

Michael J. Wilensky, Miguel F. Morales, Bryna J. Hazelton, Nichole Barry, Ruby Byrne, and Sumit Roy. Absolving the SSINS of Precision Interferometric Radio Data: A New Technique for Mitigating Faint Radio Frequency Interference. *Publ. Astron. Soc. Pacific*, 131(1005):114507, 2019. doi: 10.1088/1538-3873/ab3cad.

Michael J Wilensky, Nichole Barry, Miguel F Morales, Bryna J Hazelton, and Ruby Byrne. Quantifying excess power from radio frequency interference in Epoch of Reionization measurements. *Monthly Notices of the Royal Astronomical Society*, 498(1):265–275, 08 2020. ISSN 0035-8711. doi: 10.1093/mnras/staa2442. URL https://doi.org/10.1093/mnras/staa2442.

Michael J. Wilensky, Miguel F. Morales, Bryna J. Hazelton, Pyxie L. Star, Nichole Barry, Ruby Byrne, C. H. Jordan, Daniel C. Jacobs, Jonathan C. Pober, and C. M. Trott. Evidence of ultrafaint radio frequency interference in deep 21 cm epoch of reionization power spectra with the murchison wide-field array. *The Astrophysical Journal*, 957(2):78, November 2023. ISSN 1538-4357. doi: 10.3847/1538-4357/acffbd. URL http://dx.doi.org/10.3847/1538-4357/acffbd.

Michael J. Wilensky, Jacob Burba, Philip Bull, Hugh Garsden, Katrine A. Glasscock, Nicolas Fagnoni, Eloy de Lera Acedo, David R. DeBoer, and Nima Razavi-Ghods. High-dimensional inference of radio interferometer beam patterns i: Parametric model of the hera beams, 2024.

Jacqueline N. Xu, Zhilei nd Hewitt, Kai-Feng Chen, Honggeun Kim, Joshua S. Dillon, Nicholas S. Kern, Miguel F. Morales, Bryna J. Hazelton, Ruby Byrne, Nicolas Fagnoni, Eloy de Lera Acedo, Zara Abdurashidova, Tyrone Adams, James E. Aguirre, Paul Alexander, Zaki S. Ali, Rushelle Baartman, Yanga Balfour, Adam P. Beardsley, Gianni Bernardi, Tashalee S. Billings, Judd D. Bowman, Richard F. Bradley, Philip Bull, Jacob Burba, Steven Carey, Chris L. Carilli, Carina Cheng, David R. DeBoer, Matt Dexter, Nico Eksteen, John Ely, Aaron Ewall-Wice, Randall Fritz, Steven R. Furlanetto, Kingsley Gale-Sides, Brian Glendenning, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Jack Hickish, Daniel C. Jacobs, Austin Julius, MacCalvin Kariseb, Joshua Kerrigan, Piyanat Kittiwisit, Saul A. Kohn, Matthew Kolopanis, Adam Lanman, Paul La Plante, Adrian Liu, Anita Loots, Yin-Zhe Ma, David Harold Edward MacMahon, Lourence Malan, Cresshim Malgas, Keith Malgas, Bradley Marero, Zachary E. Martinot, Andrei Mesinger, Mathakane Molewa, Tshegofalang Mosiane, Steven G. Murray, Abraham R. Neben, Bojan Nikolic, Hans Nuwegeld, Aaron R. Parsons, Nipanjana Patra, Samantha Pieterse, Jonathan C. Pober, Nima Razavi-Ghods, James Robnett, Kathryn Rosie, Peter Sims, Craig Smith, Hilton Swarts, Nithyanandan Thyagarajan, Pieter Van Van Wyngaarden, Peter K. G. Williams, and Haoxuan Zheng. Direct optimal mapping for 21 cm cosmology: A demonstration with the hydro-

gen epoch of reionization array. *The Astrophysical Journal*, 938(2):128, October 2022. ISSN 1538-4357. doi: 10.3847/1538-4357/ac9053. URL http://dx.doi.org/10.3847/1538-4357/ac9053.

Saleem Zaroubi. *The Epoch of Reionization*, page 45–101. Springer Berlin Heidelberg, September 2012. ISBN 9783642323621. doi: 10.1007/978-3-642-32362-1_2. URL http://dx.doi.org/10.1007/978-3-642-32362-1_2.

Zheng Zhang, Jonathan C. Pober, Wenyang Li, Bryna J. Hazelton, Miguel F. Morales, Cathryn M. Trott, Christopher H. Jordan, Ronniy C. Joseph, Adam Beardsley, Nichole Barry, Ruby Byrne, Steven J. Tingay, Aman Chokshi, Kenji Hasegawa, Daniel C. Jacobs, Adam Lanman, Jack L. B. Line, Christene Lynch, Benjamin McKinley, Daniel A. Mitchell, Steven Murray, Bart Pindor, Mahsa Rahimi, Keitaro Takahashi, Randall B. Wayth, Rachel L. Webster, Michael Wilensky, Shintaro Yoshiura, and Qian Zheng. The impact of tandem redundant/sky-based calibration in mwa phase ii data analysis. *Publications of the Astronomical Society of Australia*, 37, 2020. ISSN 1448-6083. doi: 10.1017/pasa.2020.37. URL http://dx.doi.org/10.1017/pasa.2020.37.

H. Zheng, M. Tegmark, J. S. Dillon, D. A. Kim, A. Liu, A. R. Neben, J. Jonas, P. Reich, and W. Reich. An improved model of diffuse galactic radio emission from 10 mhz to 5 thz. *Monthly Notices of the Royal Astronomical Society*, 464(3):3486–3497, October 2016. ISSN 1365-2966. doi: 10.1093/mnras/stw2525. URL http://dx.doi.org/10.1093/mnras/stw2525.

Appendix A

PAM REFIT COMPARISON HISTOGRAMS

Here we show histogram representations of the correlation metric both before and after the PAM refit. Figure A.1 shows spectra (top) and histograms (middle and bottom) of the real and imaginary components of the correlation metric for the entire array before the PAM refit. In the spectra we see an excess particularly at high frequencies and on baselines within the same SNAP. Similarly, Figure A.2 shows the real and imaginary components of the correlation metric as spectra (top) and as histograms averaged over frequency (bottom), now colored by node connection. Again we see an excess in the real values particularly at high frequencies for all nodes. Conversely, Figures A.3 and A.4 show the same representations of the metric after the PAM refit, and we see that the excess in intrasnap baselines and real values are almost entirely mitigated.

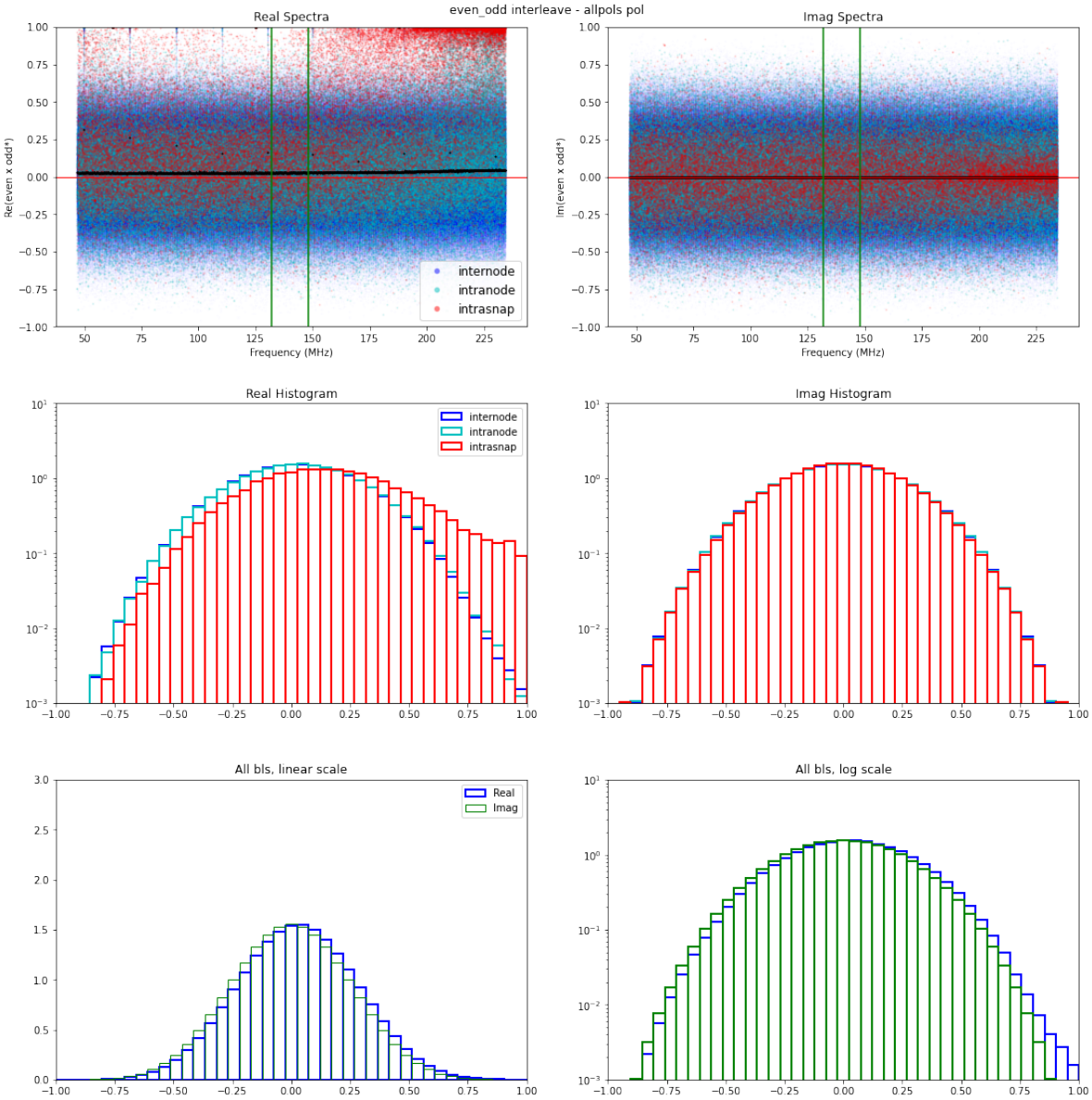


Figure A.1: Real and imaginary values of the correlation metric before the PAM refit. The top row shows the metric value per baseline and per frequency, colored by whether the baseline is between nodes, within a node but between SNAPs, or within a SNAP. The middle row shows real and imaginary histograms across the frequency band indicated by vertical green lines in the top row, again colored by baseline type. The bottom row shows the same but with all baselines combined. We see a substantial elevation of real values within a SNAP, and a more subtle elevation within a node.

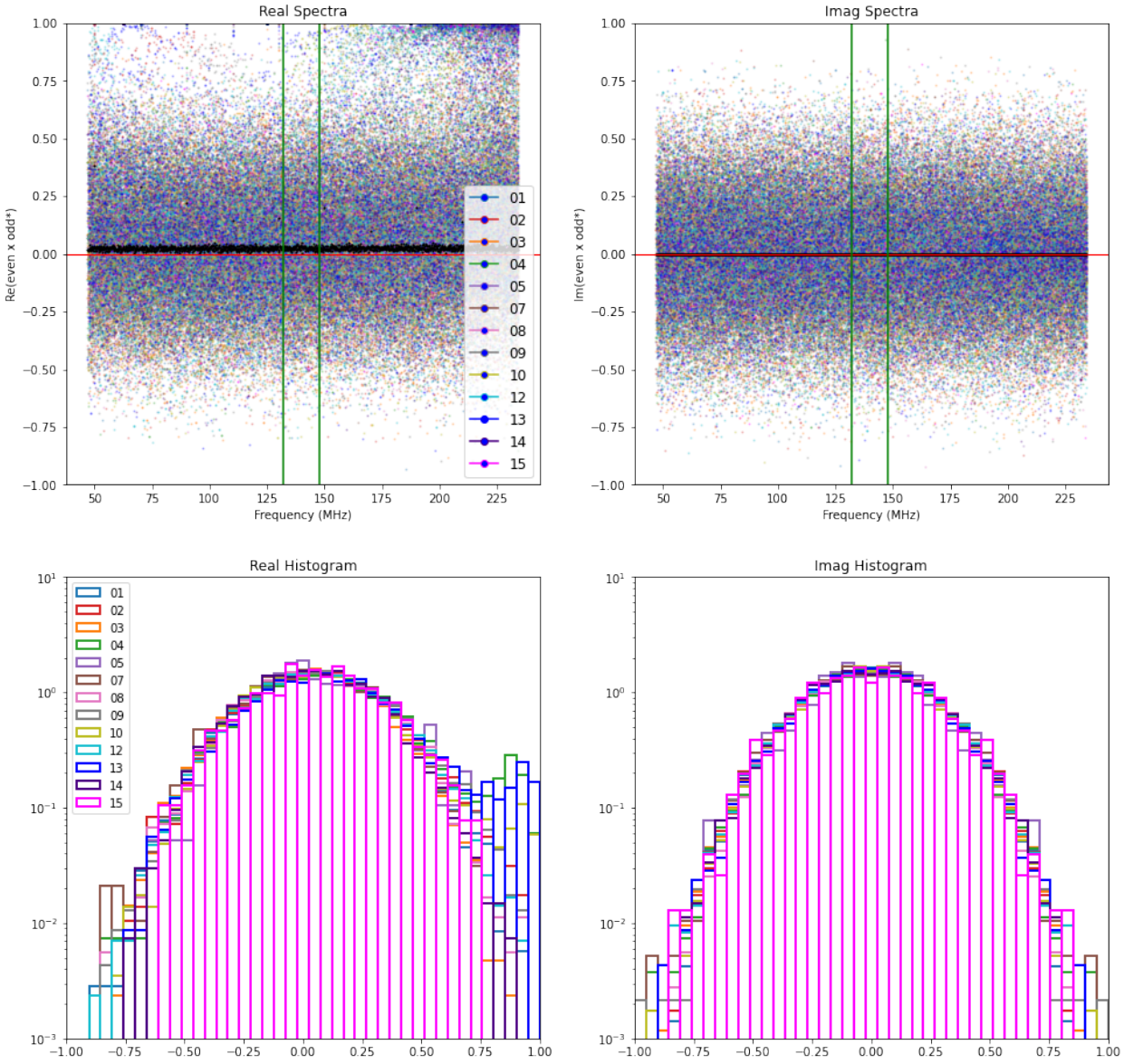


Figure A.2: Real and imaginary values of the correlation metric before the PAM refit. The top row shows the metric value per baseline and per frequency, colored by node. The black line shows the average across all baselines. The bottom row shows real and imaginary histograms across the frequency band indicated by vertical green lines in the top row, again colored by node. We see an excess in the real values, particularly at high frequencies.

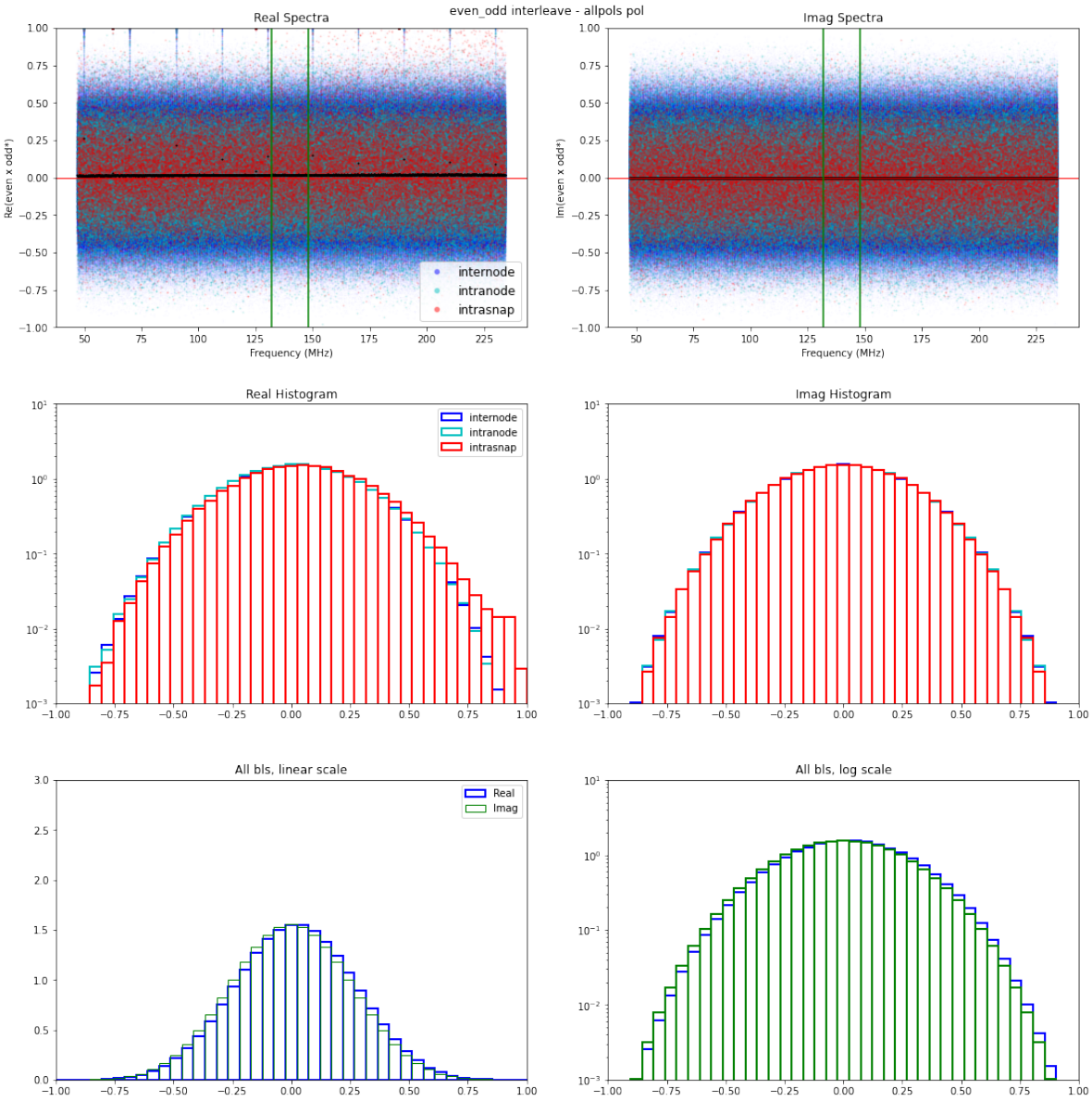


Figure A.3: Real and imaginary values of the correlation metric after the PAM refit. The top row shows the metric value per baseline and per frequency, colored by whether the baseline is between nodes, within a node but between SNAPs, or within a SNAP. The middle row shows real and imaginary histograms across the frequency band indicated by vertical green lines in the top row, again colored by baseline type. The bottom row shows the same but with all baselines combined. We see a much smaller elevation of real values compared to Figure A.1.

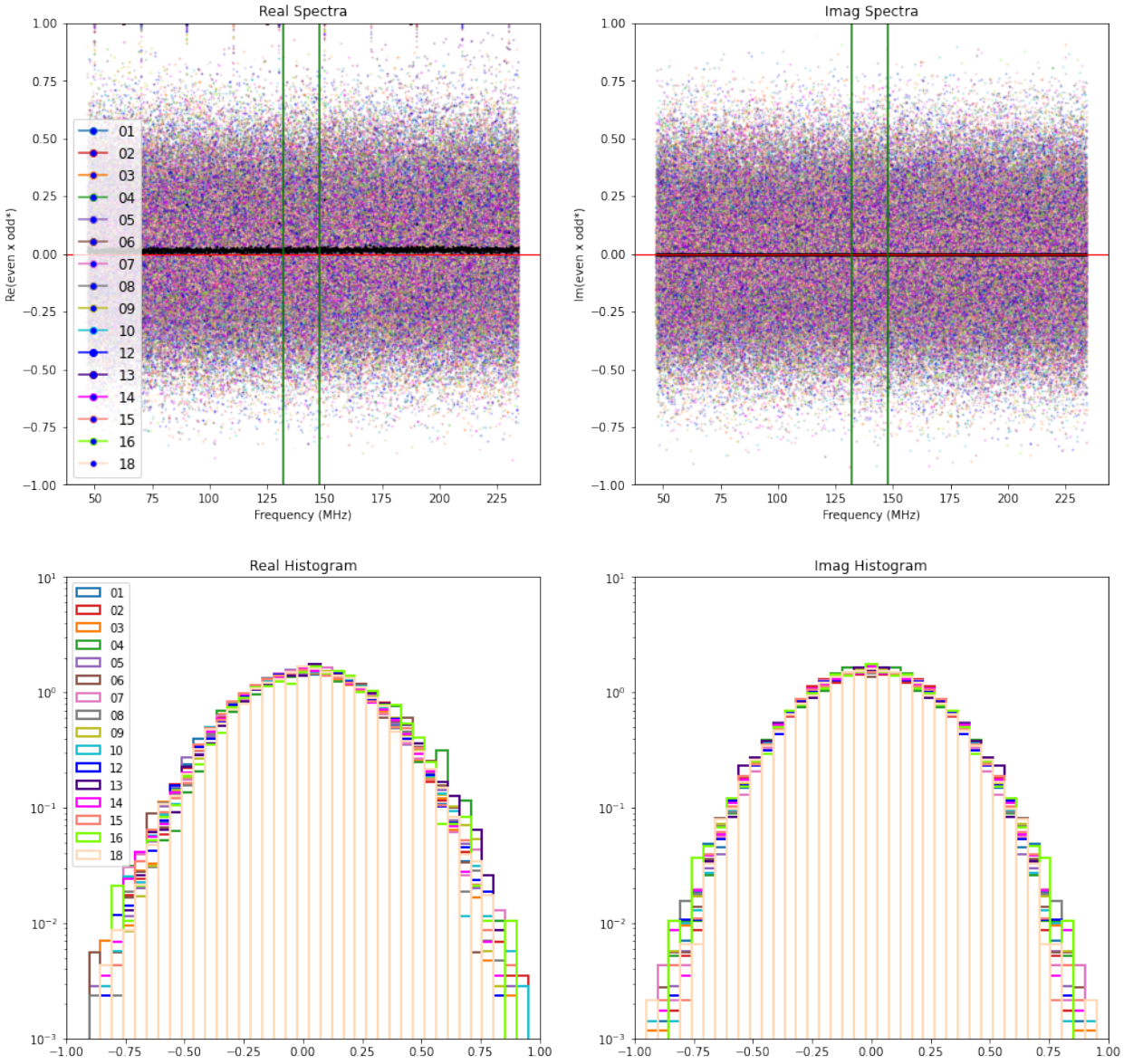


Figure A.4: Real and imaginary values of the correlation metric after the PAM refit. The top row shows the metric value per baseline and per frequency, colored by node. The black line shows the average across all baselines. The bottom row shows real and imaginary histograms across the frequency band indicated by vertical green lines in the top row, again colored by node. We see no clear excess in the real values.

Appendix B

TIME INTEGRATION WITH HERA

We find that integrating across LSTs with HERA causes decoherence in the uv weights. Figure B.1 shows the uv weights after integrating between one and sixteen observations (or between 40 seconds and 10.5 minutes). We see that with no time integration (top left panel) the weights are smooth within the compact core and have the expected outrigger wings. However, as we integrate over longer windows we see gaps in the weights with progressively tighter spacing. Eventually, as evident in the bottom right panel, this effect causes the bins centered around $k_x = 0$ to be significantly up-weighted relatve to other bins, and creates a lot of un-physical structure in the weights.

This feature is suspicious, as we see no evidence of this evolution in either the data or the model, regardless how long we integrate. However, once the weights are applied only data near $k_x = 0$ has substantial weight and most of the data is functionally discarded, leading to very low sensitivity in our power spectrum. We do not yet understand the origin of this feature, and therefore we choose to avoid it by not coherently averaging across LSTs. Instead, we coherently average each LST bin across all 17 nights of observation. Then, after creating the power spectrum cubes we incoherently average across all LSTs.

We suspect this feature is related to HERA's nature as a drift-scan telescope (all other telescopes used with the `FHD-`ϵ`ppsilon` pipeline have been pointing telescopes). However, more investigation is necessary to fully understand and more thoroughly address this issue. For the sake of future thought on this topic, Figures B.2, B.3, and B.4 show slices of the Healpix image cubes when we coherently average across 64 observations ($\sim$40 minutes), coherently average across sets of four observations and then incoherently average these sets, and incoherently average across all observations, respectively. In all cases we show the Data, model, weights, and variance cubes. In the cases where some times are incoherently averaged we show cubes from the beginning, middle, and end of the night to show the evolution.

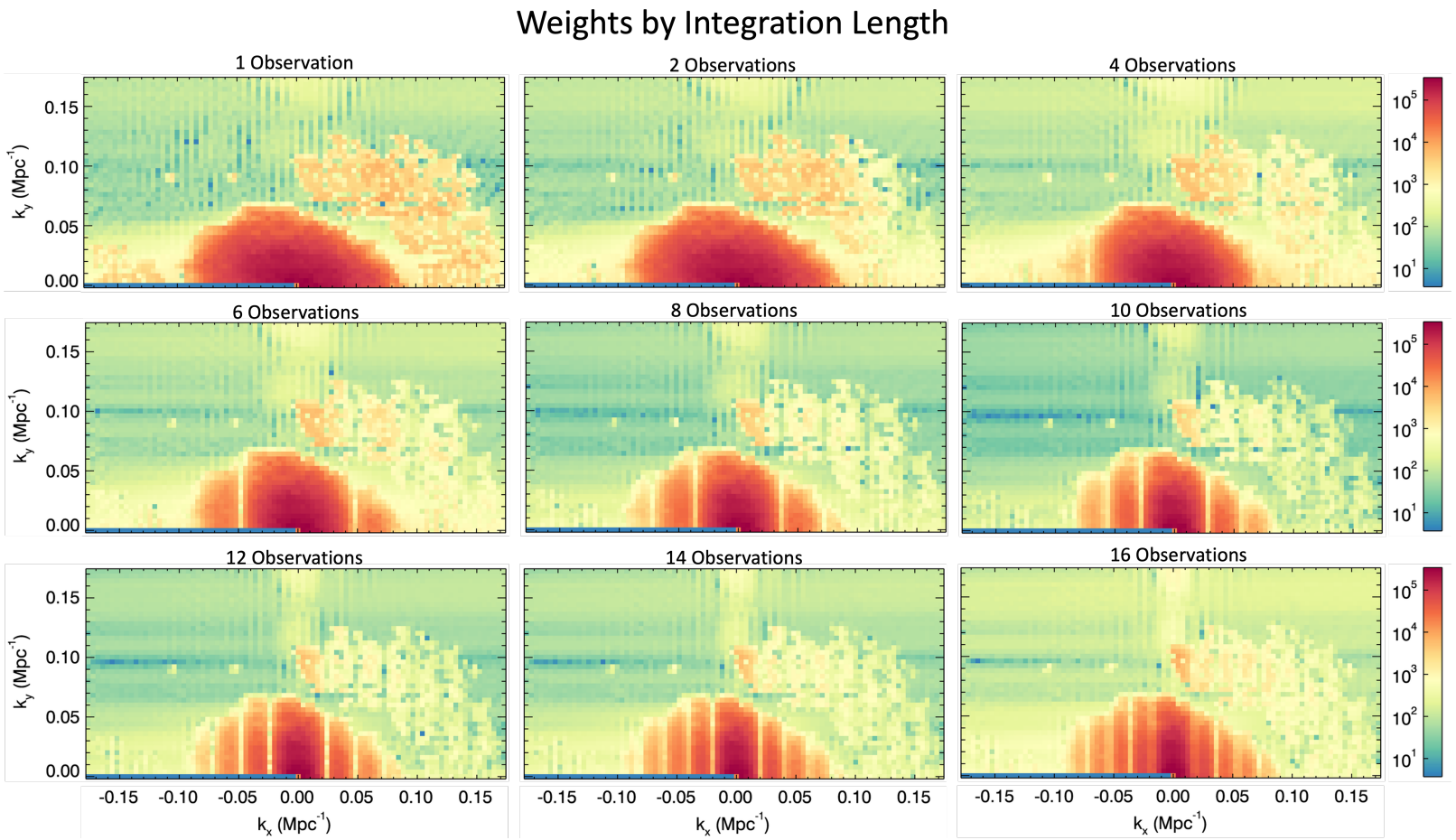


Figure B.1: Power spectrum weights used by `ϵppsilon` for 9 different integration lengths. We see that integrating over more than four observations starts to introduce structure to the weights, and causes a significant over-weighting in a central stripe at $k_x = 0$.

Similarly, Figures B.5, B.6, B.7 show the 2D power spectra for the same three integration schemes. In all cases, we show the dirty, model, and residual power spectra for both polarizations. While coherent averaging provides enhanced sensitivity in general, we see that in HERA's case it causes substantial power leakage beyond the noise and is a prohibitive systematic.

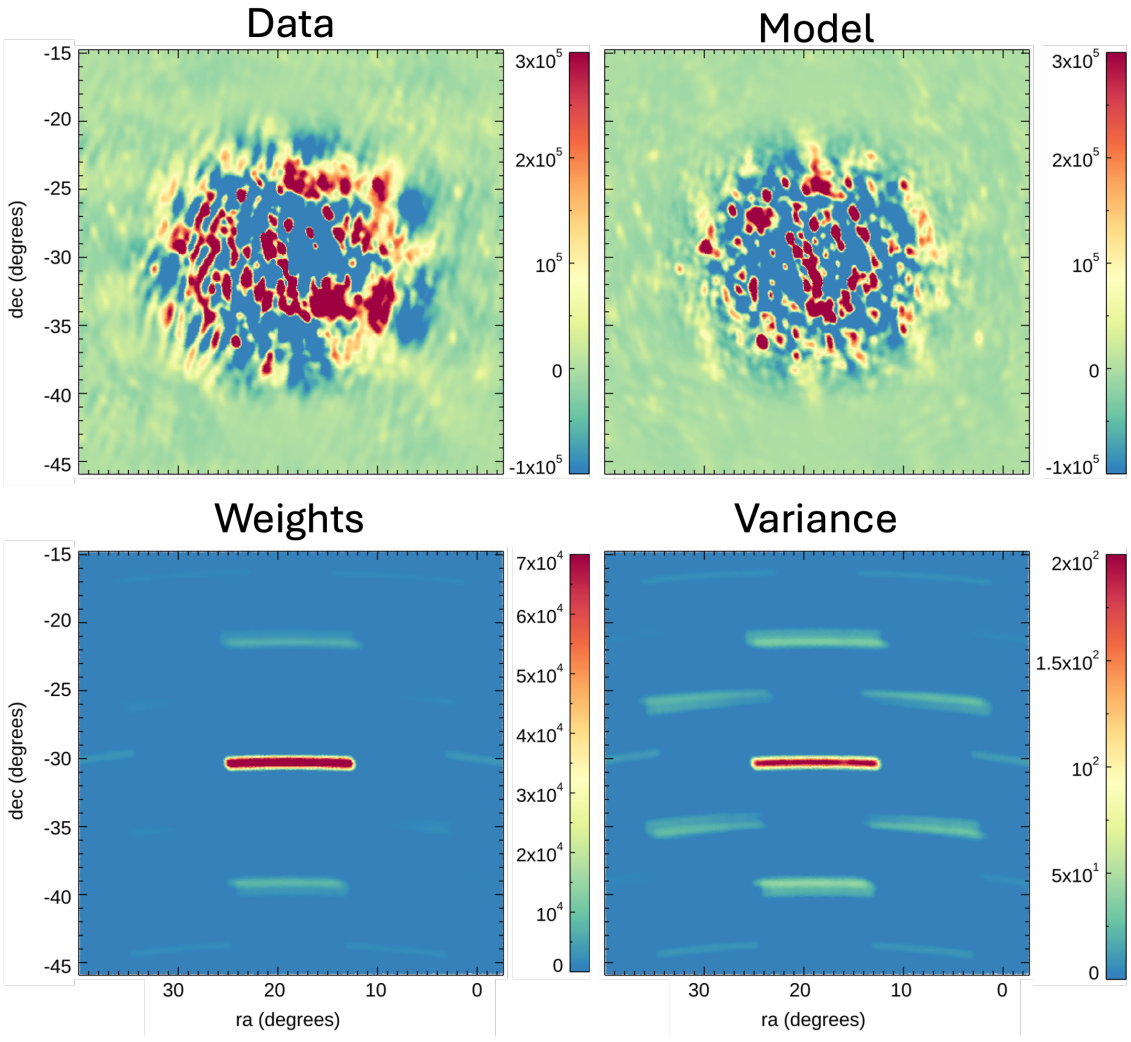


Figure B.2: Slices of the data, model, weights, and variance cubes at the first frequency bin in the case where all 64 LST bins are coherently averaged.

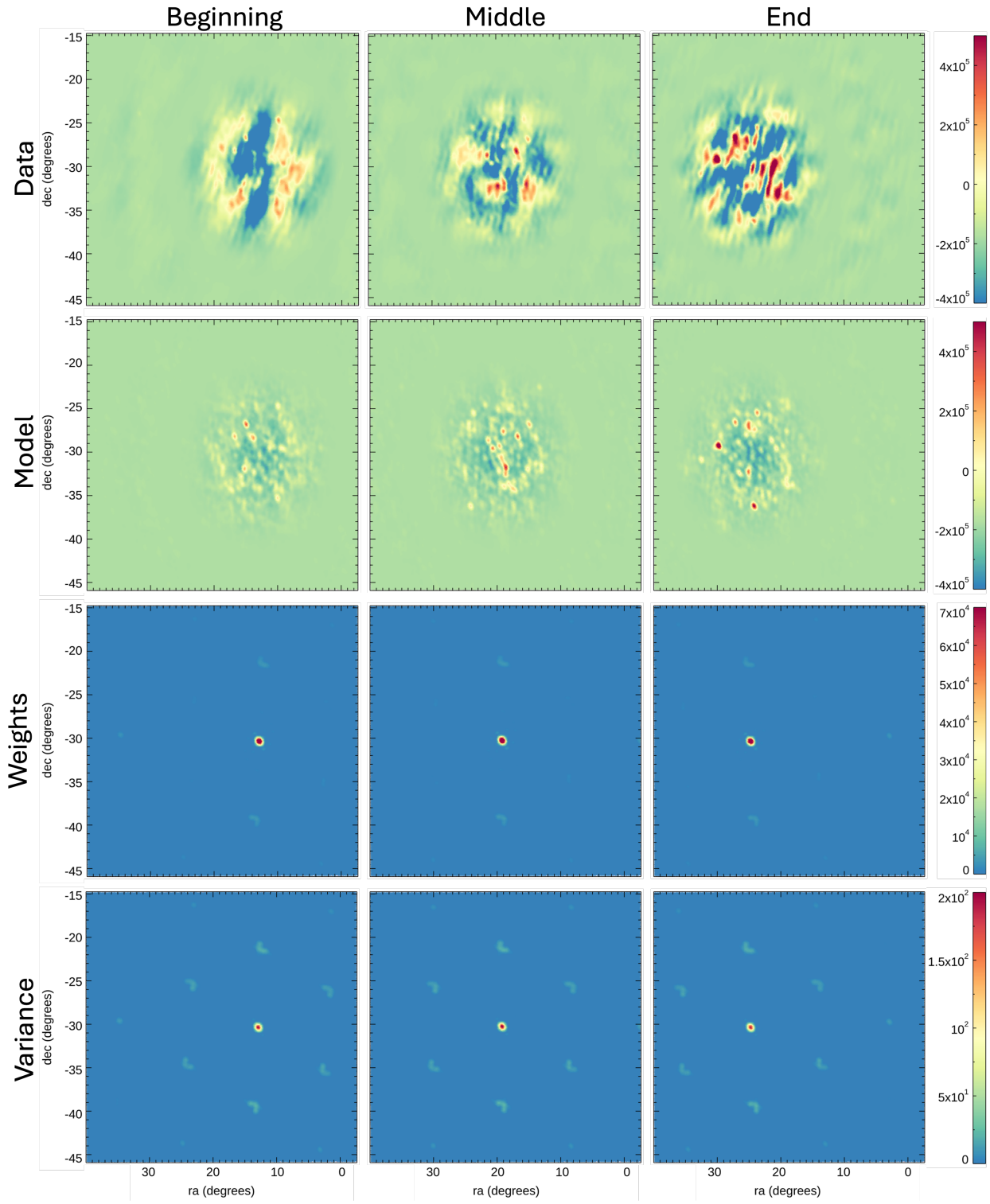


Figure B.3: Slices of the data, model, weights, and variance cubes at the first frequency bin in the case where LST bins are coherently averaged in sets of four.

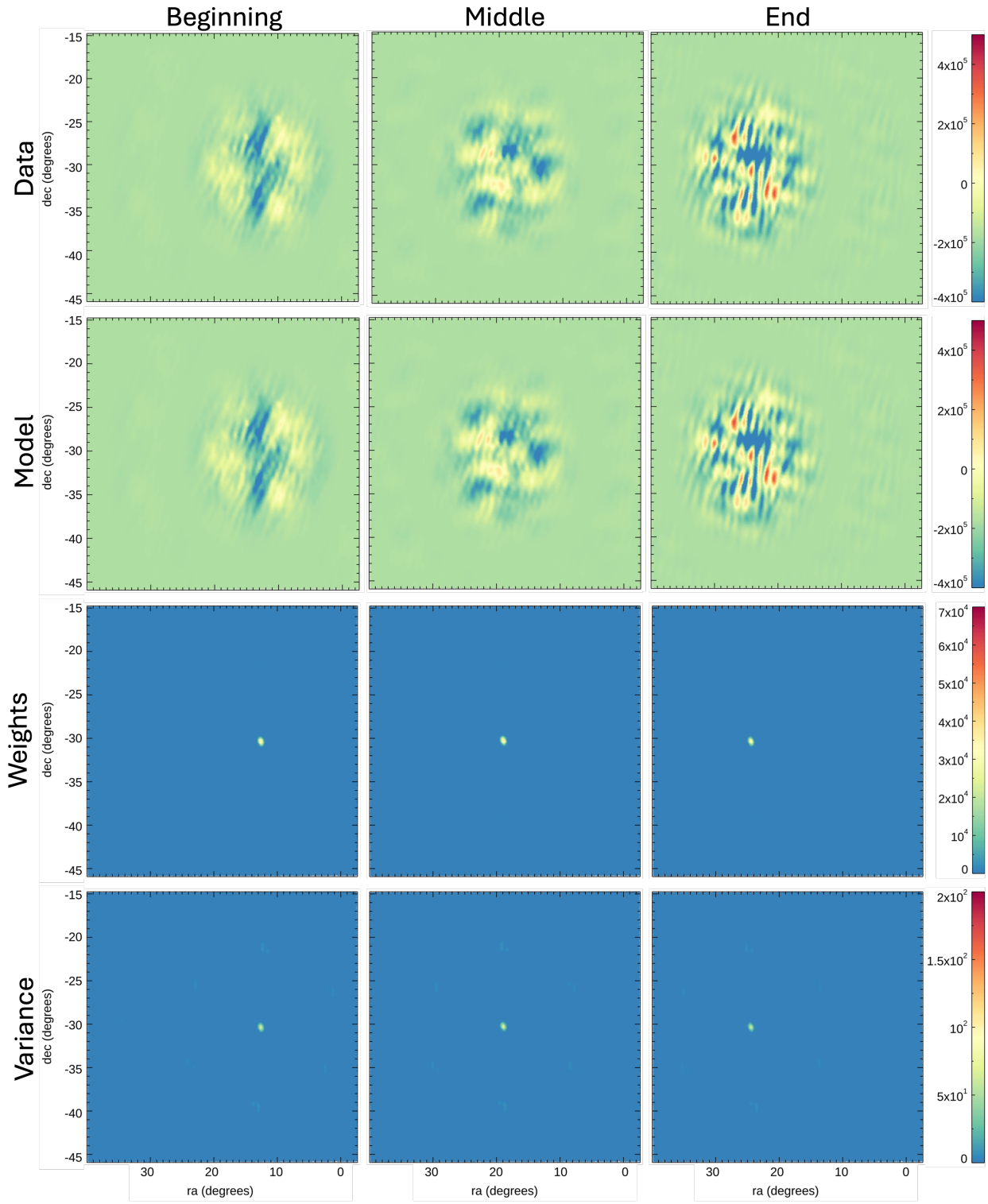


Figure B.4: Slices of the data, model, weights, and variance cubes at the first frequency bin in the case where there is no averaging across LSTs.

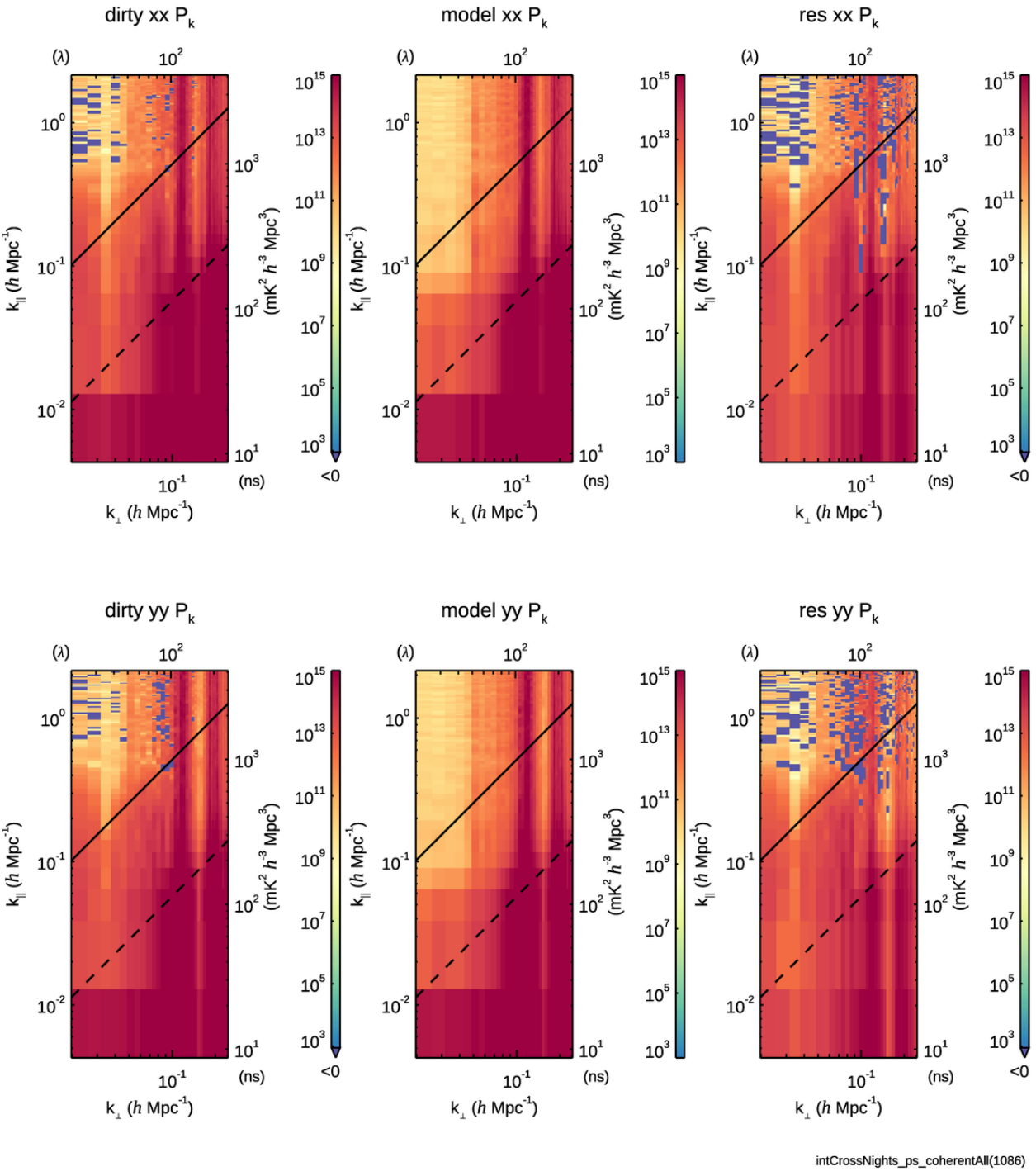


Figure B.5: Dirty, model, and residual 2D power spectra for both polarizations in the case where all 64 LST bins are coherently averaged.

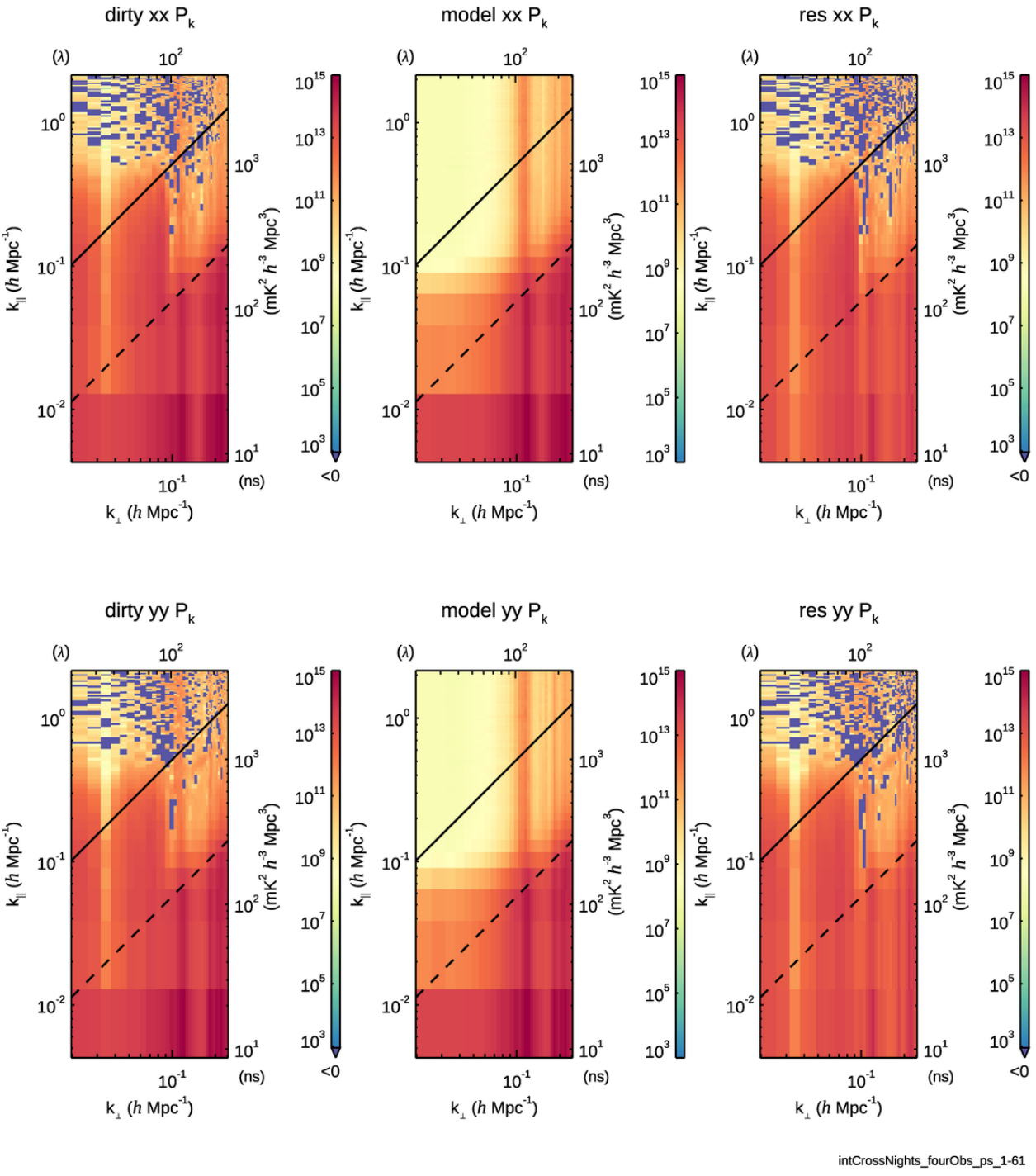


Figure B.6: Dirty, model, and residual 2D power spectra for both polarizations in the case where LST bins are coherently averaged in sets of four, and then the sets are all incoherently averaged.

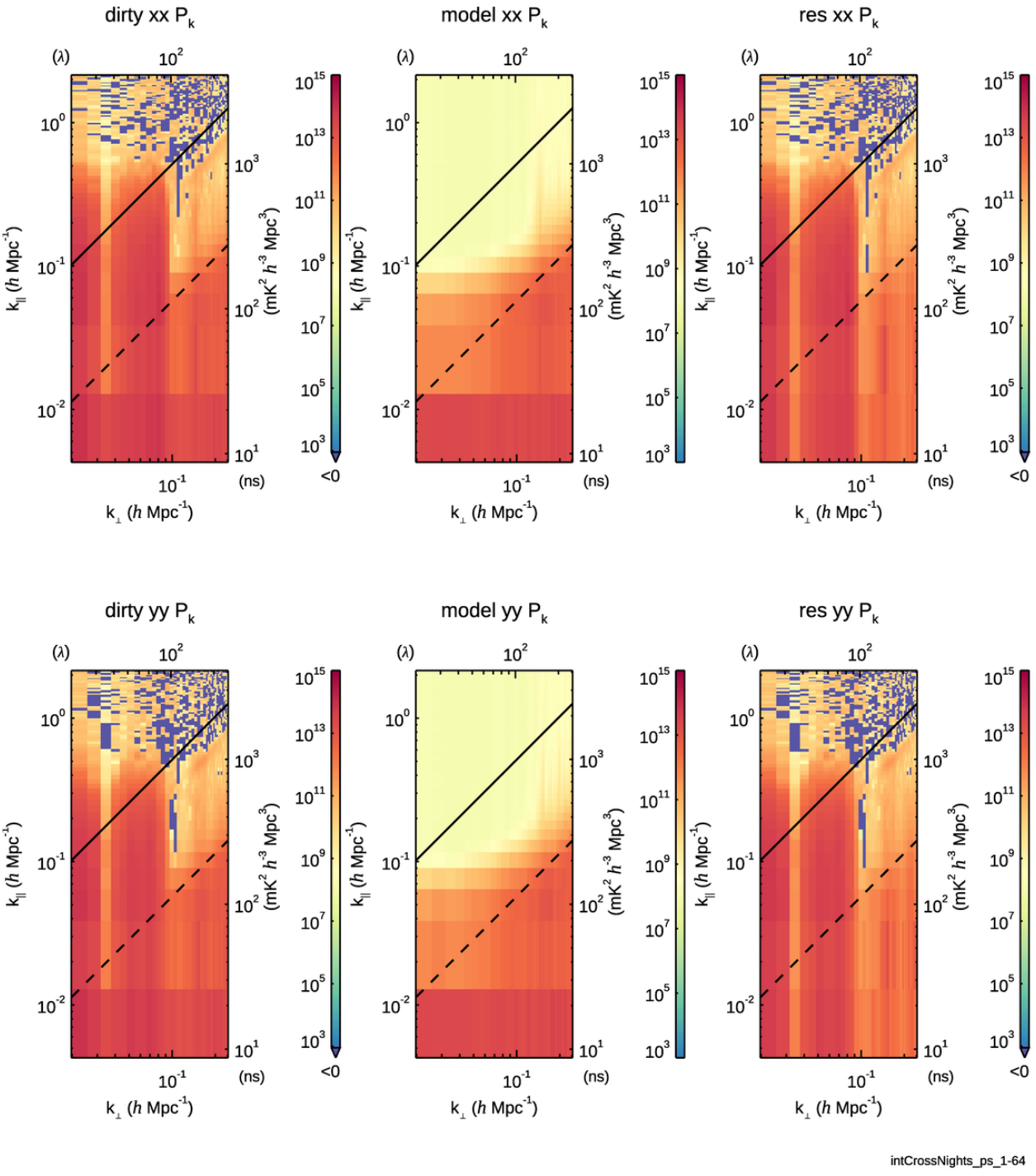


Figure B.7: Dirty, model, and residual 2D power spectra for both polarizations in the case where all 64 LST bins are incoherently averaged.

Appendix C

INVESTIGATING THE $k_x - k_y$ PLANE

This appendix elaborates on the discussion in Section 8.4 by showing some additional slices in the $k_x - k_y$ plane. First, Figure C.1 shows a slice of the model power spectrum at $k_\parallel = 0.2$. The purpose of this figure is simply to validate that we do not see any evidence of the blob feature in the model power spectra. The green brackets indicated the range of k_x and k_y values where we see the blob feature in the data power spectra, and we clearly see that it is not evident in the model power spectra.

Next, we present slices at the beginning and end of our LST range to investigate whether the blob feature exhibits temporal evolution. In particular, we are interested in identifying whether this feature could be an effect related to sources on the sky. Figures C.2 and C.3 show power spectrum slices at the first time in the final analysis range. This data has been coherently averaged across all 17 nights, but has not been averaged across LSTs. In both cases we see evidence of the blob, although it is somewhat difficult to see in the $k_\parallel = 0$ power slice. In comparison, Figures C.4 and C.5 are taken at the last time in the analysis set, approximately 40 minutes after the previous slices. We see that the blob is somewhat brighter in the $k_\parallel = 0$ power slice compared to the first time, and there are slight changes to the shape of the blob. However, these changes are not significant enough to clearly indicate they are due to sky rotation. More analysis is needed in order to more clearly establish the relationship between the blob and sources on the sky. Particularly, it would be quite informative to do single-LST power spectrum measurements at times when the galactic plane is higher on the sky.

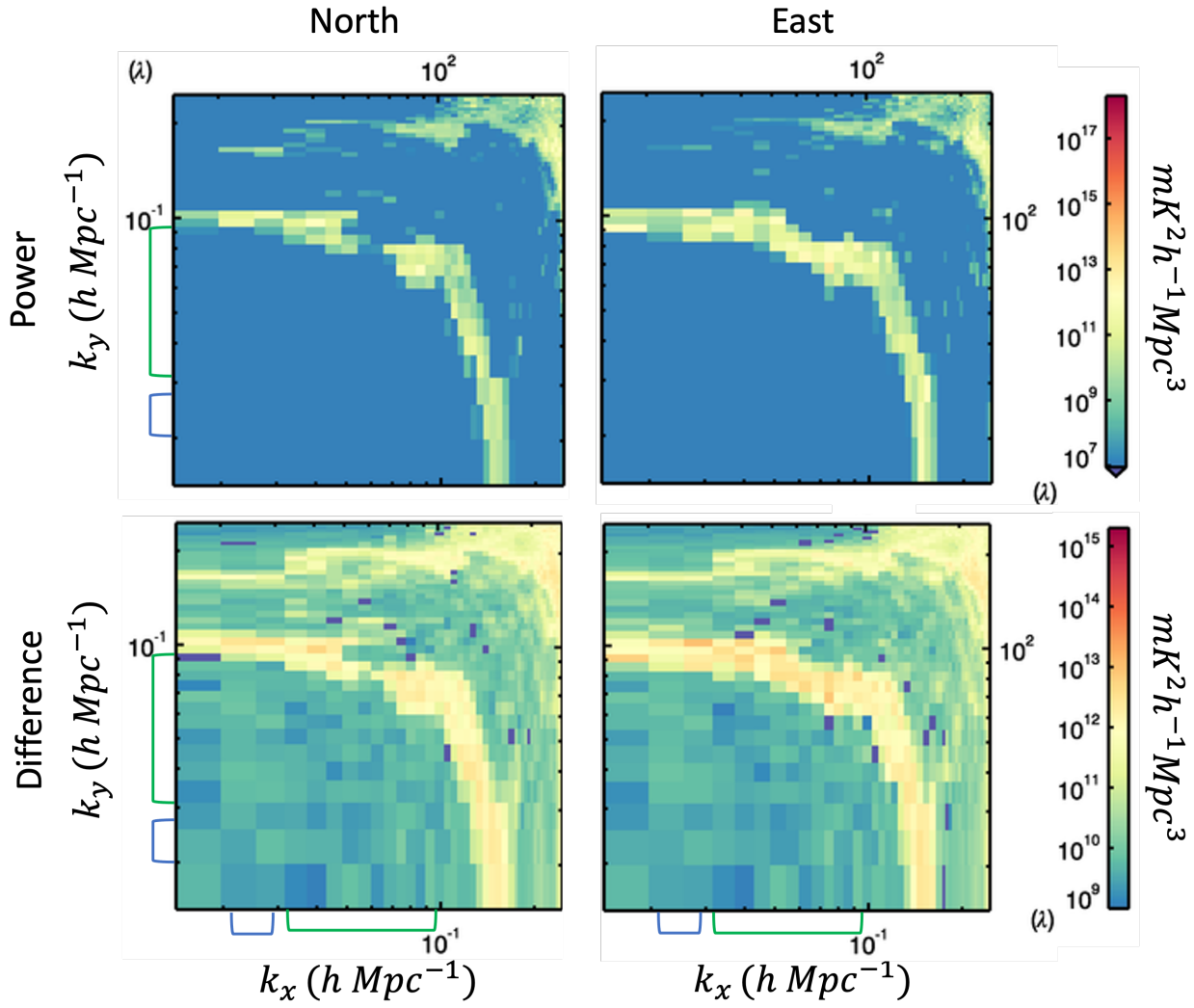


Figure C.1: Slices of the model power spectrum (top) and difference model power spectrum (bottom) for both the North (left) and East (right) polarizations all at $k_{\parallel} = 0.2$. This bin is indicated by a blue arrow in Figure 8.3. Note that the power and difference slices are on different color scales. The blue brackets indicate the location of the low power $k_{\perp}$ bin, and the green bracket indicates the range of k_x where we see elevated power.

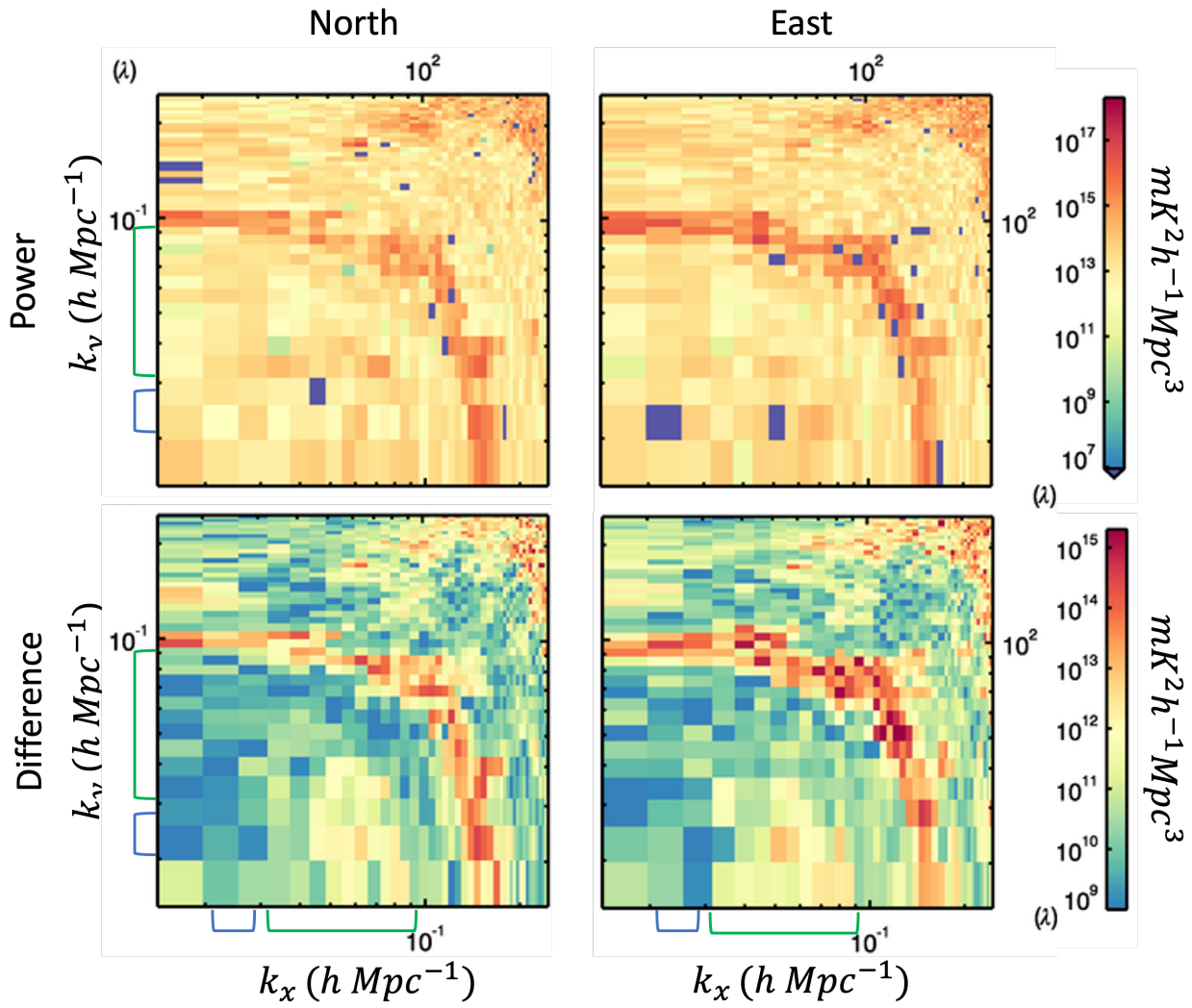


Figure C.2: Slices of the power spectrum (top) and difference power spectrum (bottom) for both the North (left) and East (right) polarizations all at $k_{\parallel} = 0$ for the first observation of our analysis set. This data comprises a single LST that has been coherently averaged across all 17 nights, but has not been integrated across LSTs. This bin is indicated by a green arrow in Figure 8.3. The blue brackets indicate the location of the low power $k_{\perp}$ bin, and the green bracket indicates the range of k_x where we see elevated power. Note that the power and difference slices are on different color scales.

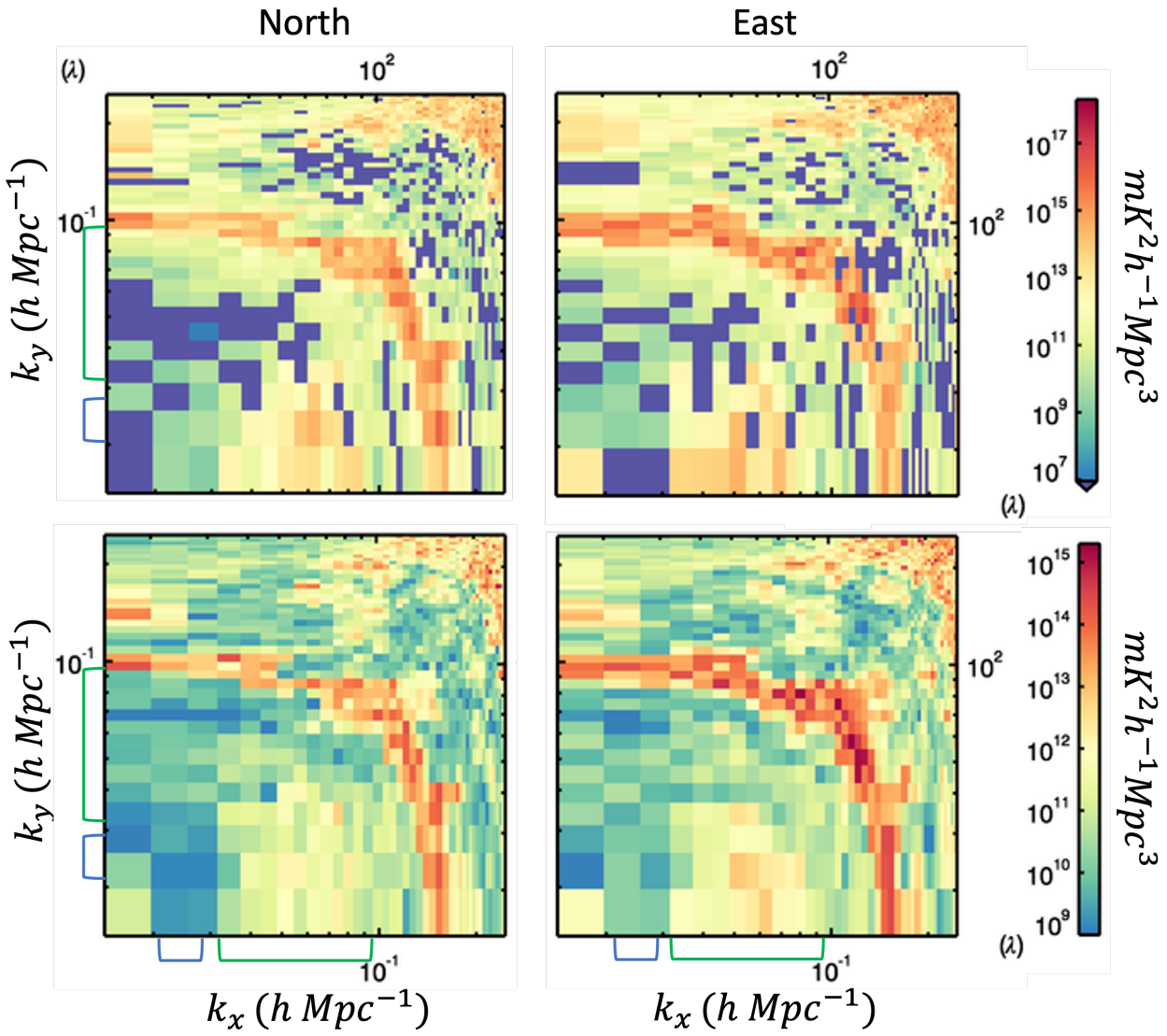


Figure C.3: Slices of the power spectrum (top) and difference power spectrum (bottom) for both the North (left) and East (right) polarizations all at $k_{\parallel} = 0.2$ for the first observation of our analysis set. This data comprises a single LST that has been coherently averaged across all 17 nights, but has not been integrated across LSTs. This bin is indicated by a blue arrow in Figure 8.3. The blue brackets indicate the location of the low power $k_{\perp}$ bin, and the green bracket indicates the range of k_x where we see elevated power. Note that the power and difference slices are on different color scales.

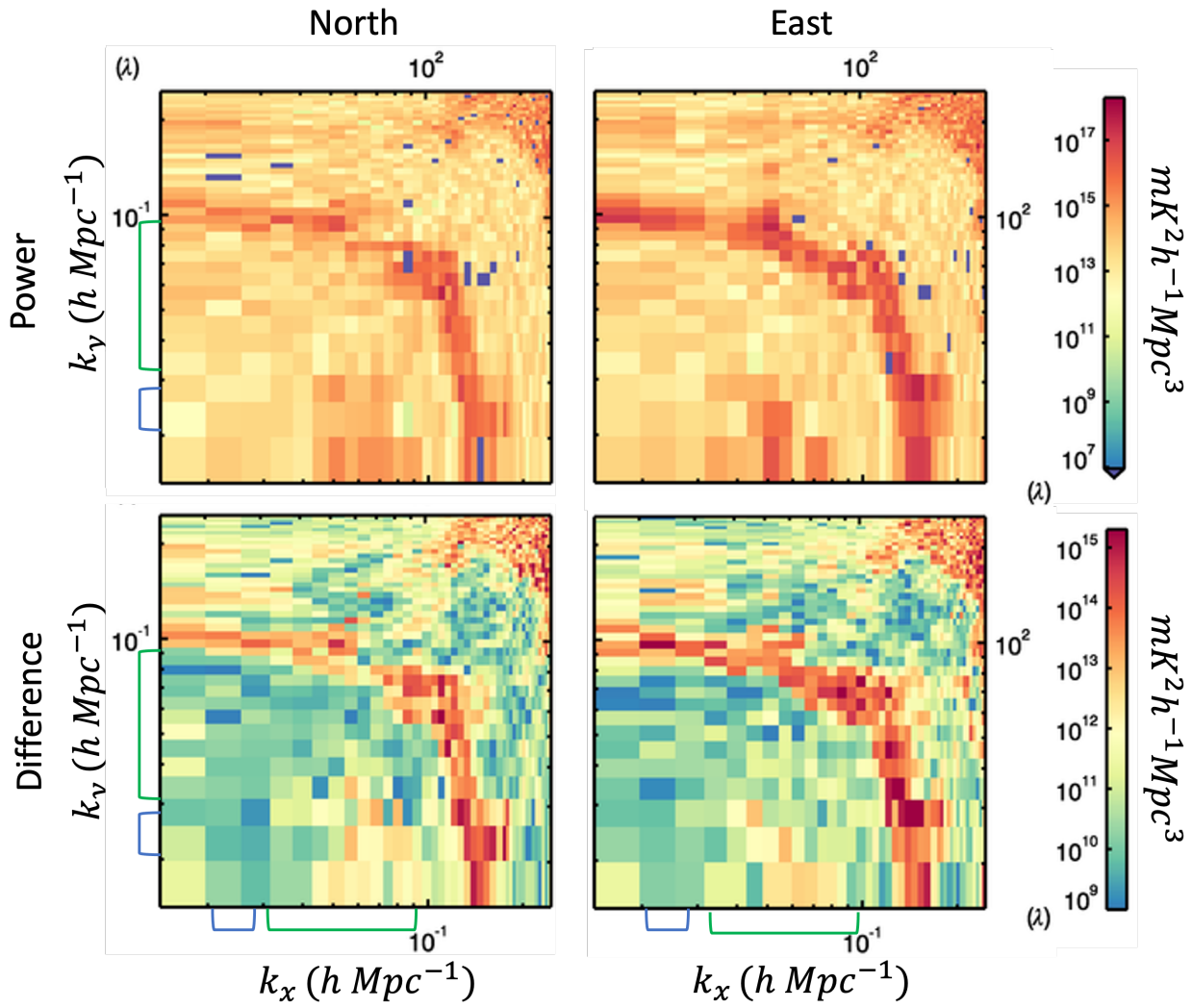


Figure C.4: Slices of the power spectrum (top) and difference power spectrum (bottom) for both the North (left) and East (right) polarizations all at $k_{\parallel} = 0$ for the last observation of our analysis set. This data comprises a single LST that has been coherently averaged across all 17 nights, but has not been integrated across LSTs. This bin is indicated by a green arrow in Figure 8.3. The blue brackets indicate the location of the low power $k_{\perp}$ bin, and the green bracket indicates the range of k_x where we see elevated power. Note that the power and difference slices are on different color scales.

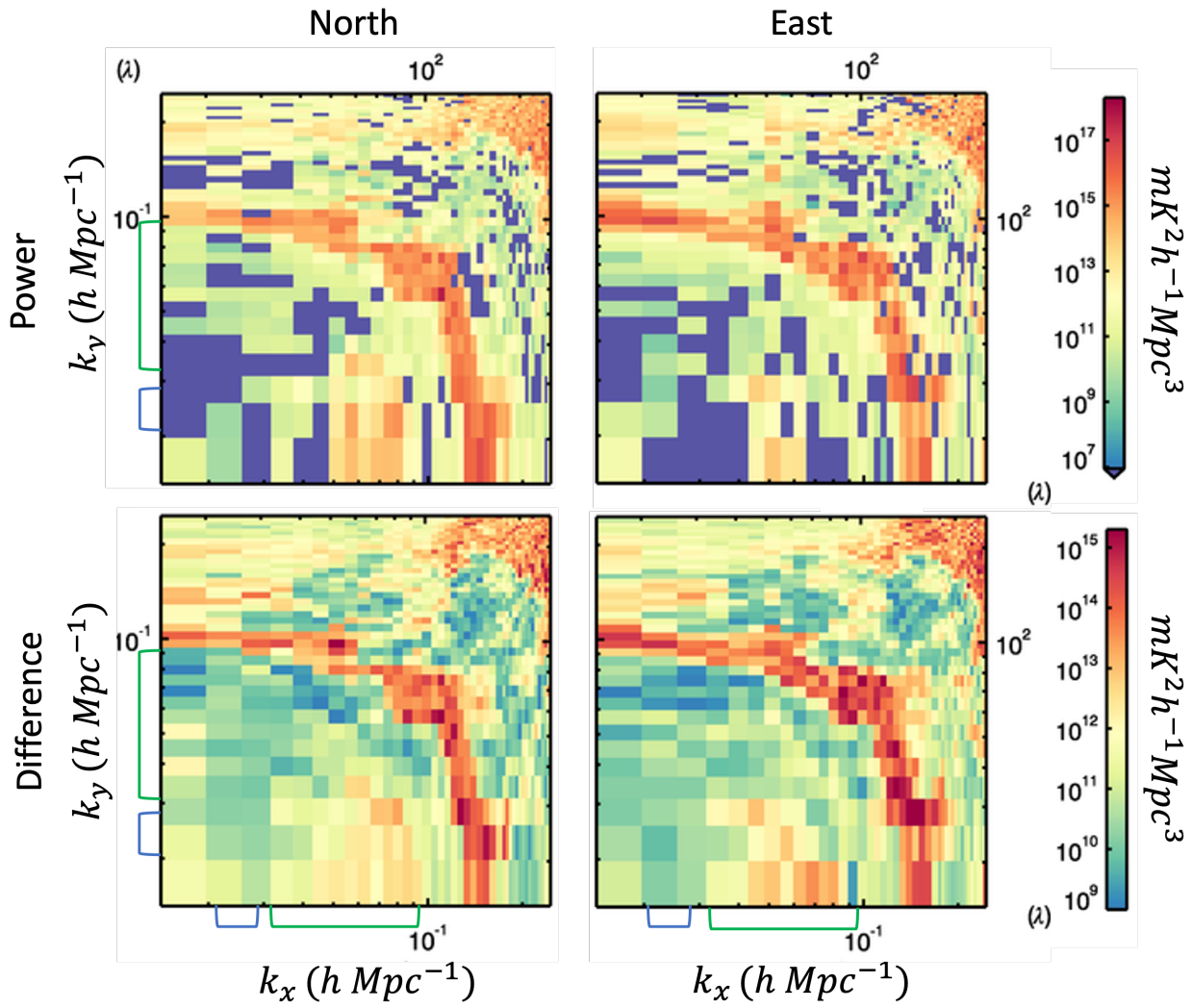


Figure C.5: Slices of the power spectrum (top) and difference power spectrum (bottom) for both the North (left) and East (right) polarizations all at $k_{\parallel} = 0.2$ for the last observation of our analysis set. This data comprises a single LST that has been coherently averaged across all 17 nights, but has not been integrated across LSTs. This bin is indicated by a blue arrow in Figure 8.3. The blue brackets indicate the location of the low power $k_{\perp}$ bin, and the green bracket indicates the range of k_x where we see elevated power. Note that the power and difference slices are on different color scales.