Enhancing EoR limits through improved instrumental calibration of the MWA

Nichole Barry

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

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Miguel F. Morales, Chair

Leslie Rosenberg

Matthew McQuinn

Program Authorized to Offer Degree:
Physics

University of Washington

Abstract

Enhancing EoR limits through improved
instrumental calibration of the MWA

Nichole Barry

Chair of the Supervisory Committee:
Professor Miguel F. Morales
Physics

Epoch of Reionization observations have the potential to be transformative in the field of cosmology, but this is impossible without unprecedented levels of precision in calibration. We enhance EoR upper limits from the Murchison Widefield Array through the improvement of instrumental calibration with in situ simulations. The reduced limit is a direct result of development in our pipeline, consisting of FHD and εppsilon, and highlights the precision and accuracy that we must achieve in calibration.

We describe our pipeline in detail, including analysis of instrumental effects like pointings, cables, and polyphase filter banks. We perform experimental assessment of various sky-calibration approaches to the bandpass, and we characterize bit errors, phase response, and new techniques. An in situ calibration simulation verifies our experimental findings, and identifies the fundamental limits of sky-based calibration. We set a precedent for required bandpass accuracy to one part in 10^5. Various improvements also contribute to our final limit, including beam kernel corrections and anti-aliasing filters found through signal loss simulations using our in situ framework.

Our improved EoR upper limit is $6.75{\times}10^3\,\mathrm{mK}^2$ at $k = 0.2\,h\,\mathrm{Mpc}^{-1}$ for $z = 7$, an improvement of a factor of 4.0 over a previous analysis using the exact same data set. Techniques and infrastructure created in this work will influence future pipeline advances.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY

BASELINE: the distance between measuring elements, usually measured in wavelengths.

BEAM: the measurement collecting area of the instrument.

CALIBRATION DELAY SPACE: spectra of calibration modes, or the Fourier transform of each bandpass calibration solution along frequency.

COARSE CHANNELS: a group of frequency channels totaling 1.28 MHz that are created during the first stage of the polyphase filter banks.

COHERENCY DOMAIN: a domain that describes the instrumental polarization between two dipoles in XX, XY, YX, YY.

COPLANAR: the assumption that all of the antennas effectively lie in the same 2D plane, allowing for averaging in image space.

DIPOLES: combinations of the arms of the MWA antenna, where the N and S arms constitute the Y dipole and the W and E arms constitute the X dipole.

DIGITAL GAIN JUMP: data multiplication to maintain proper bit selection in the digital system.

EOR: the Epoch of Reionization, or the period of time in the history of the universe where radiation ionized neutral hydrogen.

EOR WINDOW: foreground-free region in 2D power spectrum space outside of the foreground wedge.

FOREGROUND WEDGE: chromatic instrumental response that couples intrinsic foregrounds to $k_{||}$ and $k_{\perp}$ modes within the horizon in 2D power spectrum space.

GLOBAL BANDPASS: the normalized amplitude average over all antennas as a function of frequency for XX and YY.

GRID/GRIDDING: the process of de-integrating a visibility by convolving with the beam in uv-space at the baseline location.

INTRINSIC FOREGROUNDS: spectrally smooth foregrounds restricted to low $k_{||}$ modes in 2D power spectrum space.

JONES MATRIX: matrix operators which represent the projection of light onto the instrument's measuring elements.

LEAKAGE: artificially coupling power to other modes after performing a Fourier transform on finite data.

MAXIMUM LIKELIHOOD ESTIMATES: estimates of distribution parameters which maximizes the likelihood function of the data given knowledge of the type of underlying distribution.

MODEL VISIBILITIES: visibilities generated with a direct Fourier transform from a sky catalog and integrated with the beam in the uv-plane.

MURCHISON WIDEFIELD ARRAY (MWA): an interferometer located in Australia.

POINTING: a discrete set of instrument delay lines that point the antenna to measure a certain area of the sky.

POINT SPREAD FUNCTION (PSF): the response of the instrument, given its layout, to a point source on the sky.

POLYPHASE FILTER BANK (PFB): an array of bandpass filters which transforms a time series into a frequency series with reduced spectral leakage.

RFI: radio frequency interference, usually in the form of unwanted FM, AM, or TV signals.

SAMPLING MAP: map of how much of each measurement went into each pixel, created by gridding visibilities of value 1 with the beam gridding kernel.

SEASON 1: MWA observing season from August 2013 through December 2013 for the EoR0 field.

SIGNAL LOSS: accidental removal of signal due to processing techniques which could lead to a false lower limit.

TILE: a beamformed element of the MWA, composed of 16 antennas in a 4×4 grid.

UV-PLANE/SPACE: the measurement plane of visibilities, where $\boldsymbol{u} = \{u, v, w\}$ is the spatial Fourier dual of the sky coordinates $\boldsymbol{\theta} = \{\theta_x, \theta_y, \theta_z\}$.

VARIANCE MAP: map of how much variance of each measurement went into each pixel, created by gridding visibilities of value 1 with the square of the beam gridding kernel.

VISIBILITY: the Fourier transform of a measurement correlation between two elements of an interferometer.

WINDOW FUNCTION: a multiplicative function that reduces leakage caused by taking a Fourier transform, at the cost of effective bandwidth.

ACKNOWLEDGMENTS

I am deeply grateful for the community that has supported me throughout graduate school and life. The collective encouragement and kindness from all of you has guided me to where I am today.

My family has been a pillar of fortitude in my life. My parents Lisa and David Barry have made sacrifices that no ordinary human should have to bear, and for that I am eternally grateful. For Justin, thank you for showing me what success looks like. Of course, thank you to Marshall for all the love that one could possibly give. You are my forever.

As for my graduate education, I express profound appreciation to Miguel—I have learned much from your quick wit and apparently inexhaustible knowledge! The postdocs Bryna, Ian, and Jonnie were integral to my progress and had infinite patience, and they deserve all the prosperity they have found. I follow in the footsteps of past graduate students, Adam and Patti. The current graduate students, including Ruby and Mike, have been my partners in exploration, both in science and in Australia.

The EoR community has been extremely supportive. I look forward to working with Bart, Jack, and Rachel, who I already know will be fantastic cohorts. Chris, Danny, Wenyang, Adam, Josh, and Matt: working with you all has been a blast! Special thanks to Cath for being my advisor away from home—I hope to be as successful as you one day.

Thank you to my committee, UW Physics, and UW Astrobiology in guiding me through these final steps in my education. This work has been funded by National Science Foundation grants AST-1410484, AST-1506024, AST-1613855, and AST-1643011. I acknowledge the Duwamish as traditional owners of the land of Seattle, and the Wajarri Yamatji as the traditional owners of the land where the MWA resides.

DEDICATION

To Marshall. We will go to the ends of the Earth.

Chapter 1

INTRODUCTION

The birth of the universe was hot, dense, and sudden. In the first fractions of a second, the four fundamental forces of physics separated, causing exponential inflation. Over the course of twenty minutes, simple nuclei in the form of hydrogen, helium, and lithium were forged through nuclear fusion as the universe continued to cool and expand.

If the birth of the universe was sudden, then the following life stages were extremely slow. The simple, electron-less nuclei were coupled to photons, and this dense, opaque soup dominated the universe for hundreds of thousands of years. Finally, as the universe cooled to about the temperature of the surface of the sun, the simple nuclei captured electrons. This allowed the photons to flow freely, creating what is now the Cosmic Microwave Background.

The universe then went dark for hundreds of millions of years. In contrast to the exciting beginning, this epoch was cold, sparse, and long-lasting. There were no regular sources of light; no stars or galaxies. However, gravitation started to pull the neutral hydrogen, helium, and lithium into regions of dark matter over-densities. Stars and galaxies began to form.

Ionizing radiation from the first stars and galaxies streamed outward, encountering neutral nuclei. Electrons were once again stripped from nearby hydrogen and other nuclei in a period called the Epoch of Reionization (**EoR**). The ionized regions around stars and galaxies grew as more and more sources of ionizing light formed, and the regions began to overlap until there were only islands of neutral hydrogen. As these regions were ionized, the Epoch of Reionization came to a close, nearly one billion years into the lifetime of the universe.

The bulk universe after this point resembles what we see today: free-flowing radiation, stars and galaxies, and an ionized intergalactic medium. Therefore, studying the early stages is crucial to our understanding of the universe. The numerous phase changes in the first

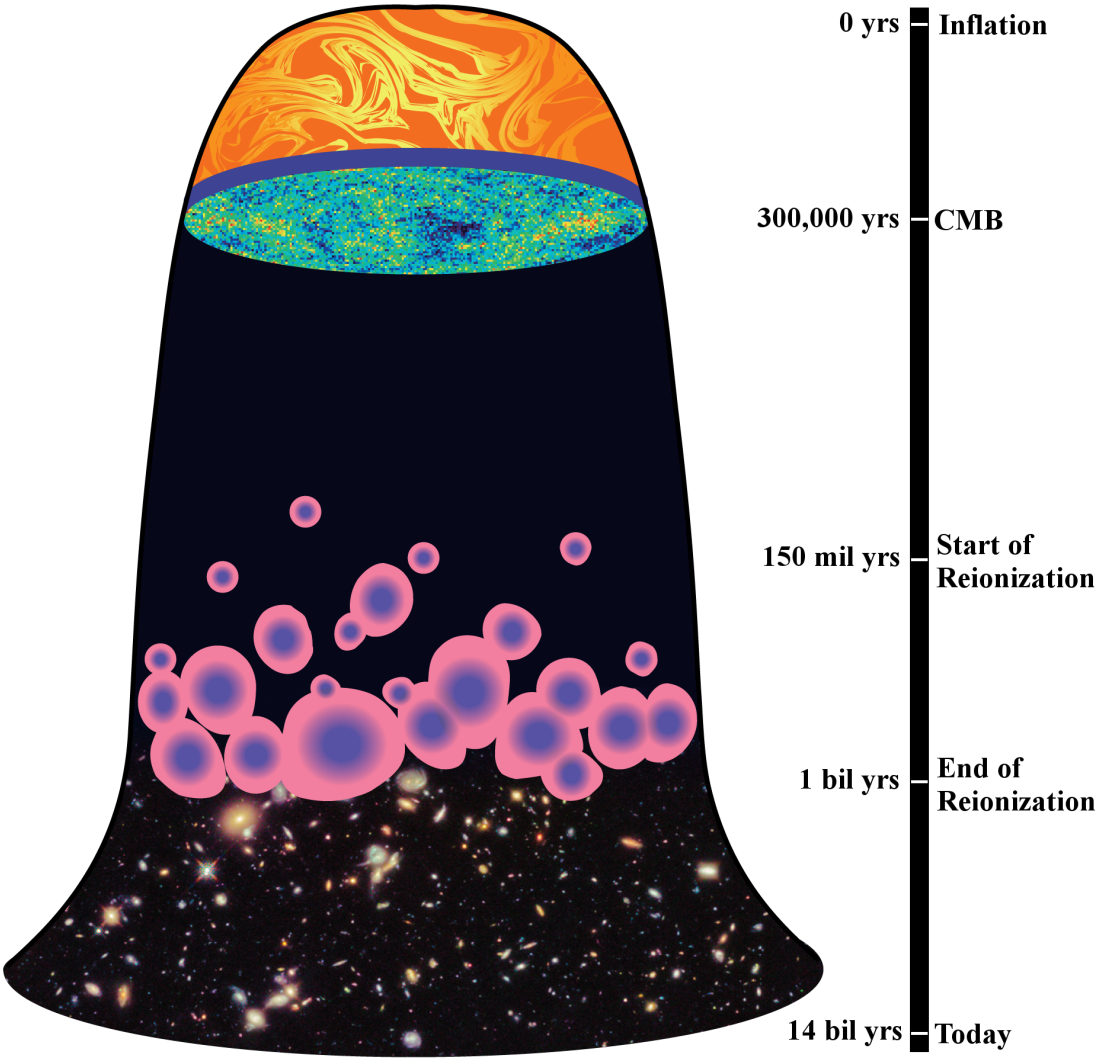


Figure 1.1: Timeline of the universe. Rapid inflation and cooling led to the decoupling of photons, creating the Cosmic Microwave Background (CMB). The Dark Ages then dominated with abundant amounts of neutral hydrogen, until the first galaxies formed. Thus began the Epoch of Reionization, where neutral hydrogen was ionized by radiation from the first galaxies in the universe. When there were enough galaxies to ionize all the hydrogen in the surrounding medium, the Epoch of Reionization ended.

billion years can tell us a great deal about the underlying physics.

The earliest stages of the universe before the Cosmic Microwave Background cannot be probed with electromagnetic radiation because photons were not decoupled, and hence the time period was opaque. The Cosmic Microwave Background itself has been studied in detail since the 1960s [84]. The Dark Ages and the Epoch of Reionization, while observable, have not been measured directly. It is no surprise that the Dark Ages are hard to observe since there is little to see, but the Epoch of Reionization is both accessible and a mystery.

In fact, the Epoch of Reionization can be directly measured. The ionizing light from some early sources can be observed, which can help to probe the medium between the light's emission and our measurements [19]. However, there is an even more powerful observation.

Neutral hydrogen emits light [26]. A spontaneous transition can occur in the properties of the bound electron, where the intrinsic magnetic dipole moment of the electron flips from symmetric to anti-symmetric compared to the proton's intrinsic magnetic dipole moment. This is called the hyperfine transition, in part because the difference in the energy states is very small. A photon is emitted to account for this change in energy at a wavelength of approximately 21 cm.

However, this occurs rarely. The rate of this spontaneous transition is close to once every 10 million years. This would mean there would be very few 21 cm photons, except for the fact that about 75% of baryonic matter is neutral hydrogen and that collisions speed up the

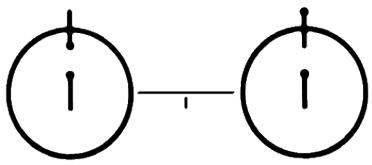


Figure 1.2: The hyperfine transition of hydrogen, engraved on the Pioneer plaque and Voyager Golden Record as a standard unit of measure (credit: NASA). Light is emitted of wavelength 21 cm when the electron of neutral hydrogen experiences a spin-flip.

rate. Therefore, there is a nontrivial glow coming from regions of neutral hydrogen. This creates a negative image of the formation of sources within the universe; 21 cm light will highlight the regions *between* ionizing sources.

Why would we look for this faint glow, rather than the first stars and galaxies themselves? The reason is due to the properties of the emitted light. Higher energy light, or light with smaller wavelengths, is easily absorbed by interstellar gas and galaxies; for example, if the light has enough energy, a neutral hydrogen atom will absorb the photon and release its electron. The 21 cm glow is very low in energy, and thus not readily absorbed.

In addition, the universe is expanding over time [89]. As the universe expands, so does the light within it. The 21 cm photons emitted during the first billion years of the universe will then be redshifted, or expand in wavelength, by about a factor of 10 by the time they reach Earth 13 billion years later. This light flows relatively unimpeded and unabsorbed.

The emitted 21 cm light is redshifted into low-frequency radio wavelengths by the time it reaches Earth. Logistically, this is difficult to measure, given that we have been emitting our own radio wavelengths in the form of AM radio, FM radio, and TV broadcasting since the 1920s. Even more problematic, though, is the fact that there are other sources on the sky that emit in radio wavelengths. Synchrotron emission from electrons traveling at relativistic speeds around galaxies is a huge radio source. Not only does our own galaxy emit this, but large amounts of radio light is seen from giant radio galaxies. In total, this signal is about 5 orders of magnitude brighter than the signal we are trying to measure [60].

Luckily, the glow from neutral hydrogen is separable from other sources. Each snapshot in time during the Epoch of Reionization was different, as ionized regions grew and morphed over time. Therefore the intensity of the glow, or the power, changes rapidly as a function of time. We can determine the time the light was emitted by using the measured wavelength, since the expansion rate of the universe and the emitted wavelength are known. Therefore, the power of the Epoch of Reionization varies rapidly as a function of *measured wavelength*. In contrast, the power from synchrotron emission varies smoothly over wavelength.

The Epoch of Reionization signal will stand out if we look at the measurement in a space

that highlights the change as a function of wavelength. Attempts at a first measurement are therefore done with the power spectrum in Fourier space. Rather than try to image the sky to see regions of neutral hydrogen, we try to measure how much power there is in Fourier modes. In other words, how much does the brightness of the Epoch of Reionization signal change, both spatially and in time, as a function of wavelength?

The power spectrum is a statistical measurement. While we won't be able to see the neutral hydrogen directly, we can still extract the underlying physics. Knowing the sizes of neutral hydrogen regions throughout time will help us to constrain cosmological models and understand our universe.

We try to measure this signal with a very specialized instrument. Typical radio astronomy instruments come in the form of dishes, but these instruments usually look at specific objects on the sky, and thus have small measurement areas. In contrast, to measure the Epoch of Reionization signal, we need as much measurement area as possible. Integrating the faint neutral hydrogen regions will make their power spectrum brighter, and hopefully easier to see. One way to have a wide field of view with the required sensitivity and resolution is to build an interferometer.

An interferometer consists of multiple elements that are spatially separated. An electric field is measured at each element from incoming plane waves, and the electric fields are combined together between each pair of elements. There will be points where the waves add to zero or where they add to be large since the electric fields propagate as waves. The resulting pattern is called a fringe. It contains information about the direction on the sky where the plane wave originated, the brightness of the plane wave, and the distance between the elements, or **baseline**.

Mathematically, this means that an interferometer contains the same information as a dish with holes: the more baselines of varying length, the more filled the dish. Logistically, spreading elements of an interferometer out many kilometers is much easier and less expensive than building an entire dish of the same size.

A collaboration of over 15 universities from across the world came together to build such

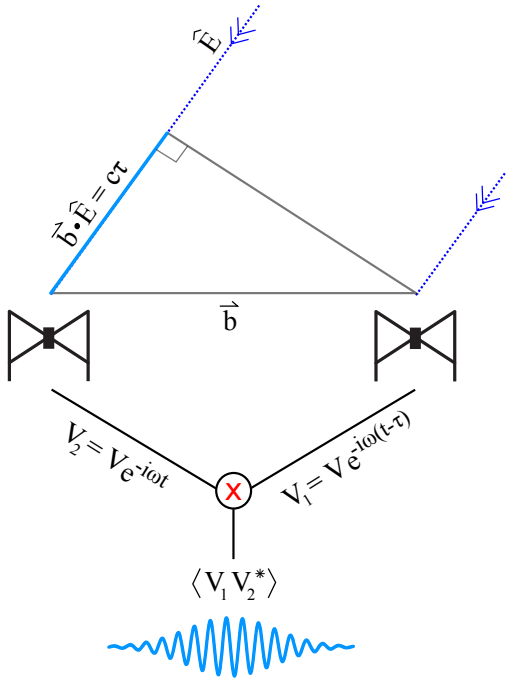


Figure 1.3: Schematic of interferometry between a pair of elements and the resulting fringe pattern. The direction of incoming light ($\hat{E}$) is delayed by time τ depending on the distance between elements $|\vec{b}|$. The two voltage signals are cross-correlated and time-averaged to get a visibility, which adds constructively and destructively as the source moves across the sky.

an interferometer to search for the Epoch of Reionization. The Murchison Widefield Array (**MWA**) is located in the Australian outback, and has 128 elements with over 8000 baselines. While it is used for a wide variety of science goals, its main mission is to look for Epoch of Reionization signals [9].

One drawback in using interferometers is that they are computationally intensive. The interference between two antennas creates a fringe. A time-averaged complex fringe, or **visibility**, is created between each pair of antennas for each frequency channel and for each time integration. For a typical night of observing with the MWA, over ten billion visibilities are generated. This poses a huge undertaking for real-time computation hardware, disk storage, and analysis pipelines.

Figure 1.4: The MWA, a radio interferometer in the Australian outback.

The theme of this dissertation is that of software, and the challenges and progress made towards measuring the Epoch of Reionization. We will describe current constraints in the following section to provide a solid background on the subject. Then, we will explain the measurement theory and the instrument itself in fine detail to finish out §2, which will lay necessary groundwork for later chapters. In §3, we will outline our data analysis pipeline. Modifications to the calibration and the resulting improvements are discussed in §4. To prove the importance of calibration and how it can affect the power spectrum, we built a calibration simulation pipeline, outlined in §5. §6 presents the results from the simulation. Given these outcomes, we delve into more advanced calibration techniques in §7. We describe other improvements in §8 as well as our newest limit using all of our advancements in §9.

Chapter 2

MOTIVATION, INSTRUMENT, AND FOURIER SPACE

Searching for the EoR signal is a promising field of study. Before we can proceed further, there remains two unanswered questions: why should we measure the EoR directly, and how do we accomplish that?

There exists a multitude of indirect limits using a variety of methods. However, the current state of the field is rather inconclusive, which motivates a direct model-independent measurement. This can be accomplished with radio interferometers, and greater detail will be provided on the MWA and its basic measurement theory in this chapter.

2.1 Cosmological importance and current constraints

Little is known about the EoR. It was the epoch of the first stars and galaxies, the phase change from neutral to ionized, and the beginning of the bulk universe as we see it today. However, we don't know to any certainty *when* it happened or *what* made it happen.

If we could measure the EoR directly, we would be able to constrain galaxy formation models. There are many unanswered questions about the sources that contributed to the EoR. The population statistics (e.g. mass, heavy element production, formation rate, and interaction and feedback) are relatively unknown. These astrophysical questions contribute directly to simulation models of the universe [53].

If we could measure the EoR across time, we would be able to understand the rate of reionization. We know approximately when reionization ended, but we don't understand how fast reionization happened or how the rate changed throughout time. These particular questions hold cosmological importance; they hold physical understandings of the formation of the universe [28].

Experimental constraints from various collaborations and studies have helped place estimates on reionization. The multitude of experiments have provided context—but they also have created confusion. Some constraints appear almost contradictory to others, while other constraints have changed over time beyond significant confidence levels. Most of these issues arise from results being model-dependent in some fashion. A measure of the ionization fraction is possible, though usually only indirectly, and thus depends on the model parameters used in the estimate.

A few of the experiments that have contributed to estimations of the ionization fraction are listed here for reference. Recent results are compiled into Figure 2.1 (from [72]).

CMB: The Cosmic Microwave Background has constrained many cosmological parameters from the power spectra of the temperature anisotropies embedded into the surface of last scattering [6, 36, 86]. Measuring the optical depth τ gives estimates on the history of reionization through the use of simulation and models. Calculations of this parameter have varied greatly since it was first measured. It relies heavily on an accurate dust model and on faint, large modes in the power spectrum [85].

Lyman α forest: Neutral hydrogen is very effective at absorbing light from quasars. As the universe expands, the quasar's light will redshift, causing different portions of the spectrum to be absorbed from neutral gas clouds, whether dense or sparse.

***Gunn-Peterson troughs*:** Absorption features will be present in a quasar's spectrum from neutral hydrogen, and as the quasar's light redshifts, these absorption features will populate different regions of the spectrum. The distribution and depths of these absorption features give limits to the neutral hydrogen along that line of sight. Considering population statistics from multiple quasars give estimates of bulk limits, yet so far has been limited to lower redshifts due to available quasars [27].

***Dark pixels*:** Lower limits on the timing of reionization can be made by consider-

ing the fraction of completely absorbed pixels in the spectrum to pixels which have some unabsorbed light [52]. This cannot measure the density of neutral hydrogen—trace amounts will quickly absorb the light to levels below measure. However, this is less model-dependent than other measures since no models are used to separate absorption features [55].

***Near zone*:** If the medium was sufficiently neutral around a quasar, a damping wing in the spectrum can occur. This is caused by multiple-scattering absorption in wavelengths redder than the Lyman α line. While rarely seen in spectra, models can be fit to the damping wing to gain estimates on the medium around the quasar, and thus estimates for reionization timing [7].

Lyman α emitter: Some young, low mass galaxies emit significant Lyman α light. This is a specific wavelength, so their distance can be estimated given the redshifted light measured.

***Luminosity function*:** A luminosity function, or number of emitters versus brightness of emitters, can be made for a variety of redshifts. The evolution of the function over time can tell us how much neutral hydrogen obscured the emitters, and thus is a probe of reionization history [50].

***Clustering*:** It is theorized that Lyman α emitters will be measured to be more clustered during the EoR [29]. The spectra can be redshifted by the time it leaves the ionized bubble, and thus is not readily absorbed by neutral hydrogen. Only if the bubble is large enough will this occur, and hence multiple emitters must be nearby to make a bubble large enough for this to happen.

Lyman break galaxies: Young galaxies will emit a range of wavelengths, and if surrounded by neutral gas, most wavelengths shorter than Lyman α will be absorbed. However, the longer wavelengths will flow relatively unimpeded and redshift on the way to Earth.

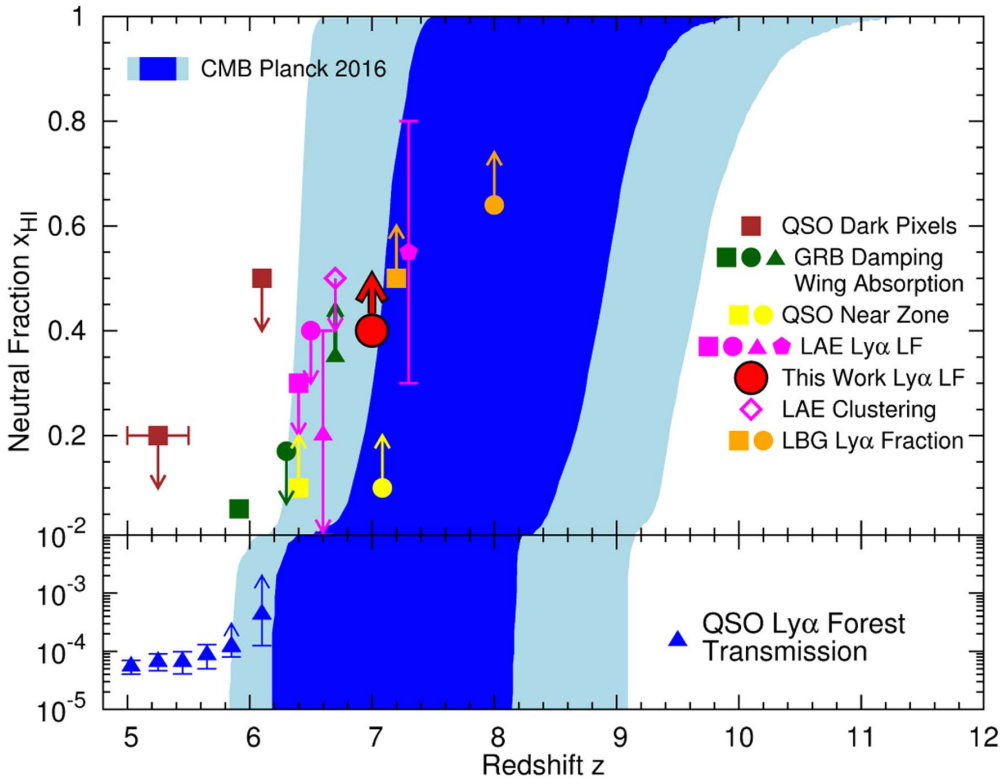


Figure 2.1: Experimental constraints on the history of the EoR (figure from [72]). Dark pixel constraints from quasars (dark red squares) are from [52]. Gamma-ray burst damping wing absorption (green square, circle, and triangle) are from [107], [108], and [32]. Near zone constraints from quasars (yellow square and circle) are from [94] and [7]. Lyman α emitter luminosity functions (pink square, circle, triangle, and pentagon) are from [50], [42], [73], and [43]. Clustering of these sources (pink open diamond) was also investigated by [73]. The Lyman α luminosity function constraint (red circle) from the work that generated this figure [72] is highlighted. Lyman break galaxies (orange square and circle) are a joint constraint from [96], [82], [83], [93], [92], [69], [109], [13], [14], and [30] at redshift 7 and [92] at redshift 8. Lyman α forest transmission spectra (blue triangles) are from [27]. The blue and light blue regions are constraints from the Planck collaboration [85] with 68% confidence and 95% confidence, respectively.

Population statistics on these galaxies put lower-limit constraints on the timing of reionization [96].

Gamma-ray burst damping: Gamma-ray bursts can probe reionization if their progenitors exist at high redshift. They are extremely bright, and much easier to see than most Lyman α emitters or quasar spectra. In addition, they have a simple power-law burst spectrum that can experience deviations (e.g. absorption features or damping) in the reddened tail due to the intergalactic medium [108]. However, the results depend greatly on the fitting techniques [107].

Given the variety of constraints, we have a rough estimate for when reionization happened. The epoch was essentially finished by a redshift of 5, and was active after about a redshift of 6. When reionization began is a mystery, and is only constrained by the Planck collaboration to have happened after a redshift of about 10 [85]. However, the first CMB results presented reionization as having *ended* by redshift 11 [95]. Their results are dependent on foreground models, which makes them difficult to constrain.

A direct measurement of the EoR is needed in order to constrain when it happened. Many of the current estimates are model-dependent, which has caused some conflicting results. In order to study this period of time, we need to measure the hyperfine transition of hydrogen, and this predicament has motivated the construction of radio interferometers.

2.2 Instrument description

The Murchison Widefield Array, or MWA, is a radio interferometer in the Shire of Murchison, Australia [105]. At any given time, only about 100 people live in the shire, which is larger in size than the Netherlands. The area is extremely flat, punctuated by small mesas that are sacred to the Wajarri Yamatji people, the traditional owners of the land. Nearby is the Australian Square Kilometre Array Pathfinder (ASKAP) telescope, and together the MWA and ASKAP share a double-Faraday-cage correlator building.

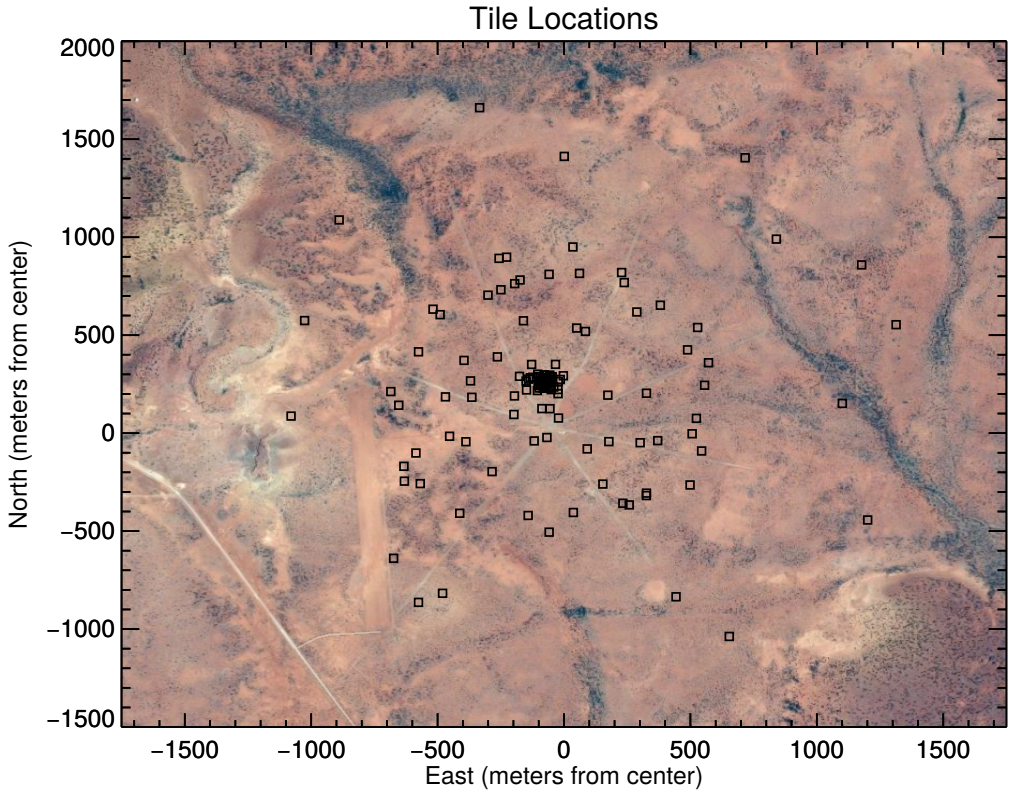


Figure 2.2: Configuration of the MWA. There is a dense core of tiles, and a sparse, long baseline spread. The dense core provides sensitivity of EoR modes, and the long baselines provide high angular resolution. Credit for the background: Google.

The MWA is composed of 128 tiles spread in a non-redundant formation. The core of the MWA is dense, and tapers off after about 50 λ, or approximately 100 m. The longest baseline is about 3 km. This layout helps to generate images that are extremely sensitive to bright, large EoR modes due to the dense core, yet also have angular resolution of approximately an arcminute from the long baselines [4]. Figure 2.2 shows a birds-eye view of the configuration. Upgrades of the array in 2016, which changed the layout, are discussed in §2.2.4.

The beginning of the hardware signal chain, illustrated in Figure 2.3, starts with the tiles. Each **tile** is comprised of a set of 16 dual-polarization **dipoles** in a square. These dipoles are placed on top of a square metal mesh, which allows a firm clip-in for the dipoles

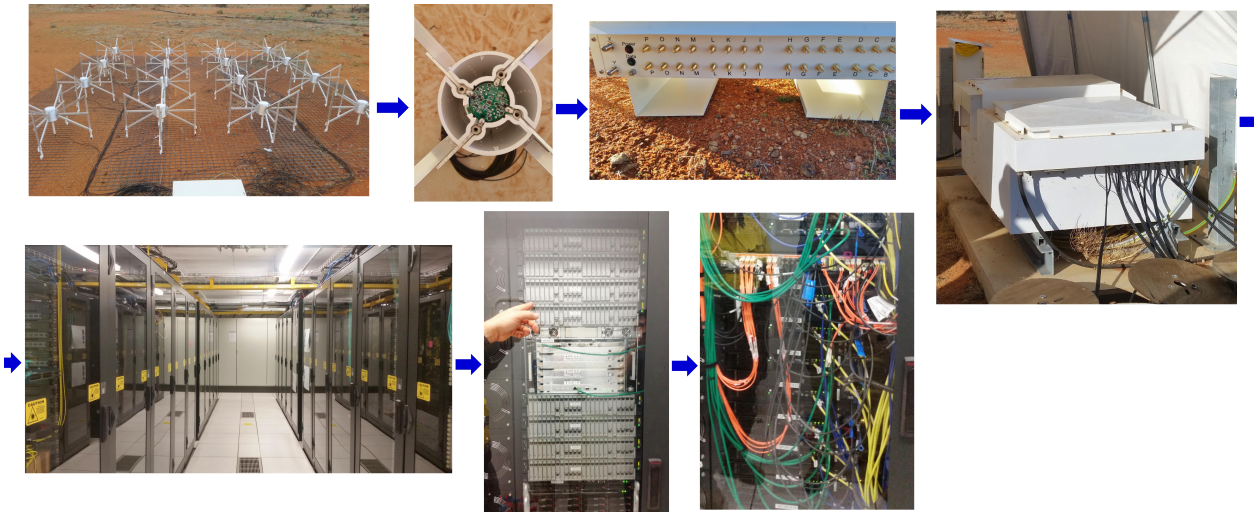


Figure 2.3: The hardware signal chain of the MWA. From left to right: a tile of dipoles, a LNA, a beamformer box, a receiver, the correlator building, a set of PFBs (and storage disks), and the correlator GPUs.

and provides extra signal from electric field reflection. Each dipole has aluminum arms that extend outward from a small, central hub that contains low-noise amplifiers (LNA). The arms are perpendicular to each other, and thus are sensitive to perpendicular linear polarizations for a source at zenith. We call these two polarizations X and Y for the E–W and N–S dipole, respectively. Two cables, one each for X and Y polarization, connect each LNA to the beamformer box.

The analog beamformer box adds together the signals from the 16 dipoles to create a unified element with roughly one square meter collecting area at 150 MHz [105]. The beamformer can also add extra physical path lengths to the dipoles, which effectively point the tile in a different direction. By adding path length, the sky directly above will decorrelate, but the sky pointed off-center will add coherently. We can create a series of coarse **pointing** adjustments in the beamformer throughout the night to roughly track one spot on the sky (see §2.2.2).

Cables run from the beamformers to the receivers (see §2.2.1 for detailed cable informa-

tion). Each receiver can manage 8 tiles, and thus for MWA's 128 tiles, there are 16 receivers. The receivers perform a series of operations, starting with a large bandpass filter with a range of 80 MHz to 300 MHz. The analog signal is then converted to a digital signal using an 8-bit A/D converter. Then the signal is digitally filtered, or split, into 1.24 MHz portions called **coarse channels** [88]. Only 24 coarse channels are selected for a total of 30.72 MHz bandwidth in analysis. All of these operations must be done quickly to keep up with the data stream. In turn, this causes a lot of heat dissipation, so each receiver is equipped with an air conditioning system inside a double-Faraday-cage box.

Optical fiber transports the digital signal from the receivers to the correlator building shared with ASKAP. Inside the correlator building is the correlator infrastructure, including racks of water-cooled, dedicated hardware for each task. More filters take the digital signal and create even smaller frequency channels of 10 kHz in each coarse band (see §2.2.3). Specialized GPU nodes then perform the correlation by multiplying and averaging.

The data then begins its final journey along optical cable to Geraldton, about 300 km away. There, it joins up with the AARNet network from Geraldton to Perth, which is another 350 km. It finally reaches the Pawsey High Performance Computing Centre for SKA Science, where the data is stored and further analyzed. Further MWA instrumental details can be found in [105].

2.2.1 Cable specifics

While not usually thought of as a prominent part of the instrument, cable specifics are critically important to our analysis. There are many different types and lengths of cables used at the MWA. They fall into three categories depending on their function: LNA-to-beamformer, beamformer-to-receiver, and receiver-to-correlator.

The LNA-to-beamformer cables carry one polarization per dipole, so there are 32 cables of this type per tile. They are all 7 m LMR-100 coaxial cables. Uniformity is crucial in these cables because the pointing of the tile depends on the timing delay.

The beamformer-to-receiver cables carry both polarizations in the same shielding per

beamformer, and thus there are 256 cables in 128 sheaths. However, these are made in a variety of lengths. There are 16 receivers in the field, located at various points along the trenching shown in Figure 2.2 (lighter color lines radiating from the center). Each beamformer has a different length requirement needed to reach a receiver, and receivers are placed in such a way that minimize the total cable used. There are five cable lengths among the tiles: 90 m, 150 m, 230 m, 320 m, 400 m, and 524 m. The 90 m, 150 m, and 230 m cables are RG-6 coaxial cables, and the 320 m, 400 m, and 524 m cables are LMR-400-75 coaxial cables.

In addition, all of the beamformer-to-receiver cables longer than 90 m are fitted with whitening filters at the receiver. Higher frequencies attenuate faster in coaxial cables than lower frequencies, causing a reddening effect in the data. A whitening filter boosts signals with higher frequencies to counteract this.

The receiver-to-correlator cables are fiber optic. Each receiver has three fibers which transmit one third of the total bandwidth for both polarizations. The digital signal is 5+5 bit complex, and the total amount transferred across the system exceeds $80\,\mathrm{Gb\,s^{-1}}$ [105].

2.2.2 Pointing specifics

Observations are naturally divided up into pointing groups, and thus analysis techniques have formed around their variation. Understanding how these pointings differ is crucial in calibration techniques.

A pointing is created by the physical addition of delay lines for each dipole in the beamformer, as shown in Figure 2.4. This allows the instrument to be steered in hardware, rather than with physical moving parts as is typical in dish telescopes. The pointings are quantized; there are five different additional lengths that are commonly used in EoR measurements. This allows for rough tracking of the sky as it moves across the night.

The beam remains fixed—the radiation pattern of the dipoles do not move. Therefore, as the tiles are pointed off-center, the beam pattern will change *from that perspective.* As a result, pointings experience different gains and modeling errors that are unique to the

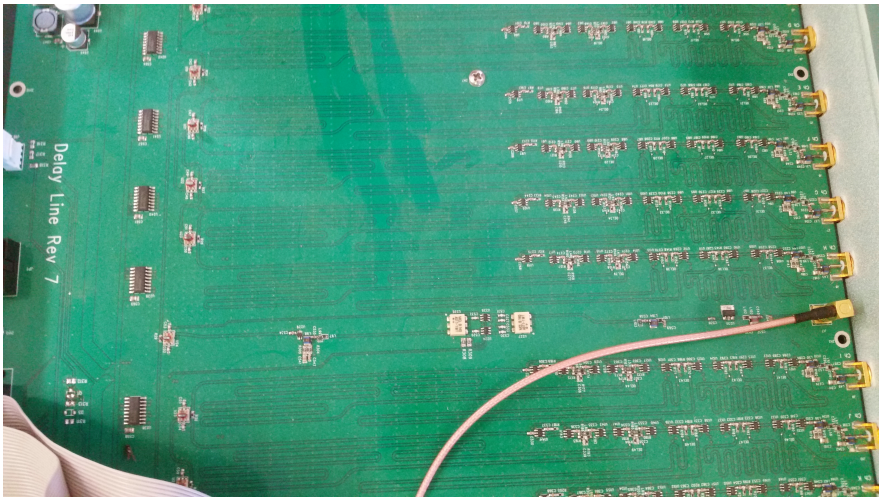


Figure 2.4: The delay line board in the beamformer. Extra physical path length is added to certain dipoles in the tile to point to a different spot on the sky.

pointing.

EoR observations are taken with many different pointings. However, our beam modeling techniques are less accurate further away from zenith. For that reason, the limits presented in §9 are only calculated from a subset of pointings near zenith.

2.2.3 PFB specifics

The polyphase filter banks, or **PFBs**, create coarse and fine resolution frequency channels. The first stage of PFBs is in the receiver hub, and they create the coarse channelization every 1.28 MHz. Not only does this take time-series data and create frequency-series data, it suppresses out-of-band **RFI** (radio frequency interference, e.g. FM, AM, and TV) which could potentially obscure real data in our band. However, this introduces a characteristic band shape to our data every 1.28 MHz and aliasing effects in the edge channels which must be accounted for in the analysis.

Data is then transported through three optical fibers from each receiver, and re-sorted into the full bandwidth at the second PFB channelization stage in the correlator building. The

data itself is in the form of Fourier transform coefficients from the first-stage channelization at the receiver.

Fine-resolution channelization is performed to generate 10 kHz samples through the use of another Fourier transform. Data is then reduced from 5-bit to 4-bit at the end of the PFBs to counteract data growth in the Fourier transform steps. This process is usually called a cascaded Fourier transform, and is similar in concept to the FFT algorithm.

Fine channels are sensitive to signal in channels on either side, but to a very small degree. Issues arise, however, when the aliasing occurs in the channels that are close to the coarse channel edge. These channels are very sensitive to signal exactly 1.28 MHz away due to the first-stage PFB channelization. This is a fundamental ringing of exactly one coarse channel, and is inherent to the design [70]. While it can be expected that aliasing occurs in all channels, it is most influential on the edges.

The final result of this aliasing is the necessary flagging of the fine frequency channels on the edge of the coarse bands. Unfortunately, this causes a regular ringing contamination in the power spectrum. In addition, the channels on the edge which are unflagged experience a characteristic bandpass shape to them, which if not modeled correctly, will cause further contamination in the power spectrum due to their regularity.

2.2.4 Phase II configuration

The MWA is going through a few upgrade and redesign stages, namely Phase II and Phase III. Phase II was successfully deployed during 2016–2017, with Phase III set for the future at the time of this writing. Commissioning efforts for Phase II are currently ongoing, and will advise the upgrades for Phase III.

For Phase II, we built a new set of 128 tiles. A subset of 72 tiles were built in two compact arrays, and a subset of 56 tiles were built in an extended array. These two designs address two different problems: EoR sensitivity/calibration, and resolution for other scientific goals. The MWA is an observatory; while its main observing program has been for EoR science, a significant portion of its program is for heliospheric science, ionospheric science, cluster

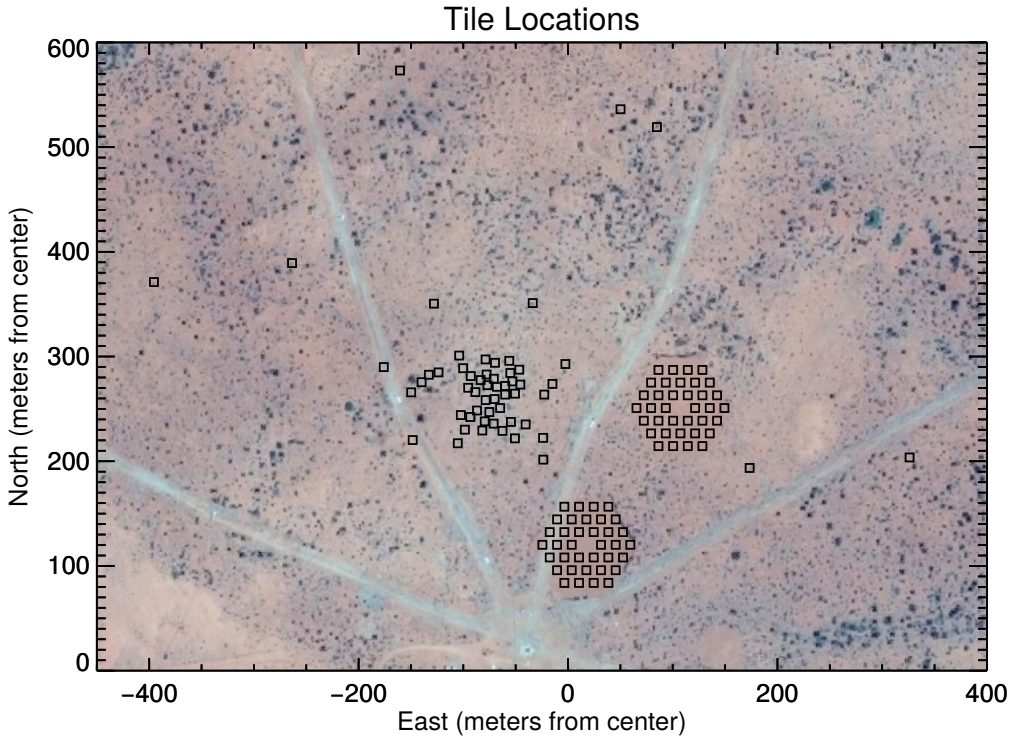


Figure 2.5: The locations of tiles for the compact configuration of the MWA, Phase II. Short, redundant baselines will increase sensitivity on EoR modes and will allow for redundant calibration. Credit for the background: Google.

statistics, and transients [9]. These goals would benefit from increased resolution on the sky.

For the purposes of measuring the EoR, however, the two compact arrays are of the most interest. These arrays are configured into hexagons of closely packed tiles, as shown in Figure 2.5. Many of the baselines within and between the two hexagons are repeated, and thus will measure the same mode on the sky many times. There are two main benefits from this configuration: 1) there will be increased sensitivity on the repeated, short baselines which are best for EoR science and 2) the ability to test redundant calibration, which needs repeated baselines in order to create a solvable system of equations on instrumental parameters [114].

This dissertation will focus heavily on sky calibration in later chapters and the resulting systematics which could make the EoR undetectable. Calibration has been a topic of great

interest in recent years due to the realization that the precision of the calibration must be unprecedented in the realm of astronomy. Redundant calibration may provide the necessary precision, but has thus far been untested on a system that can perform *both* sky and redundant calibration. Phase II commissioning will pursue this.

Phase II is a simple upgrade; we have only increased the number of tiles on the ground. The correlator can still only correlate 128 tiles at a time. As such, only 128 tiles are used during an observation at any one time, even though there are a total of 256 tiles. Phase III will explore options for correlating all tiles at one time.

2.3 Visibility space

Radio interferometers generate visibilities, the cross-correlation between a pair of voltage signals as a function of frequency and time, as their data product. However, visibilities are neither in image space nor in Fourier space, but rather an intermediate mix between the two. Therefore, it is important to discuss the space of the measurement and the various spaces of the analysis.

The visibility can be mathematically defined as $\langle V_A(f,t)\ V_B^*(f,t)\rangle$, as mentioned in Figure 1.3. For each AB antenna pair, the resulting time and frequency voltage responses $V(f,t)$ are correlated and averaged in time. In addition to this definition, we can also mathematically describe the visibility as

$$\mathbf{v}_{AB}(f) = \int_S C(\boldsymbol{\theta}, f) I(\boldsymbol{\theta}, f) e^{-2\pi i f \mathbf{b}\cdot\boldsymbol{\theta}/c} d\boldsymbol{\theta}^2, \tag{2.1}$$

where $C(\boldsymbol{\theta}, f)$ is the normalized beam response of the two antennas, $I(\boldsymbol{\theta}, f)$ is the intensity of the source on the sky, S is the full sky to be integrated over, $\boldsymbol{b}$ is the baseline vector between antenna A and antenna B, f is frequency, and $\boldsymbol{\theta}$ are the spatial coordinates on the sky [102]. In this case, we have expanded the visibility in terms of its contributions: the instrument, the sky intensity, and the complex nature of the incoming waves. If we can ignore the contribution of the beam, and substitute c/f for λ, then the measurement

equation becomes

$$\mathbf{v}_{AB}(f) = \int_S I(\boldsymbol{\theta}, f) e^{-2\pi i \frac{\boldsymbol{b}}{\lambda} \cdot \boldsymbol{\theta}} d\boldsymbol{\theta}^2. \tag{2.2}$$

This is a Fourier transform between $\frac{\boldsymbol{b}}{\lambda}$ and $\boldsymbol{\theta}$—the distance between antennas measured in wavelengths and the sky coordinates. While many assumptions were used in writing Equation 2.1 and Equation 2.2, this brief exercise has demonstrated the space in which visibilities live. They are in the Fourier-dual space of the spatial coordinates of the sky *as a function of frequency.*

A commonly used notation is to describe the wavelength-dependent baseline vector, $\{\frac{b_x}{\lambda}, \frac{b_y}{\lambda}, \frac{b_z}{\lambda}\}$, as $\boldsymbol{u} = \{u, v, w\}$. This is usually referred to as $\boldsymbol{uv}$**-space**. The visibility measurements are therefore in $\{\boldsymbol{u}, f\}$ space. This poses a slight problem; they are neither in the proper space for imaging ($\{\boldsymbol{\theta}, f\}$) nor for power spectrum (coordinate transformations of $\{\boldsymbol{u}, \eta\}$ where η is the Fourier dual of f [58]). We must use Fourier transforms or similar periodograms to utilize either analysis space.

Figure 2.6 illustrates the various transforms needed to go from one space to another. For example, a frequency Fourier transform is required to go from visibility space to full Fourier space. A coordinate transformation can then be used to generate wavenumber space, the space of our power spectrum measurement. In contrast, we can take a Fourier transform of the spatial dimensions in visibility space to go to image space. A further coordinate transform can generate a comoving distance space from this image space. In the upcoming chapters, both analysis spaces are utilized for different purposes. However, this requires large computational overhead due to the various Fourier transforms. This affects our choices of analysis variables, window functions, and resolution.

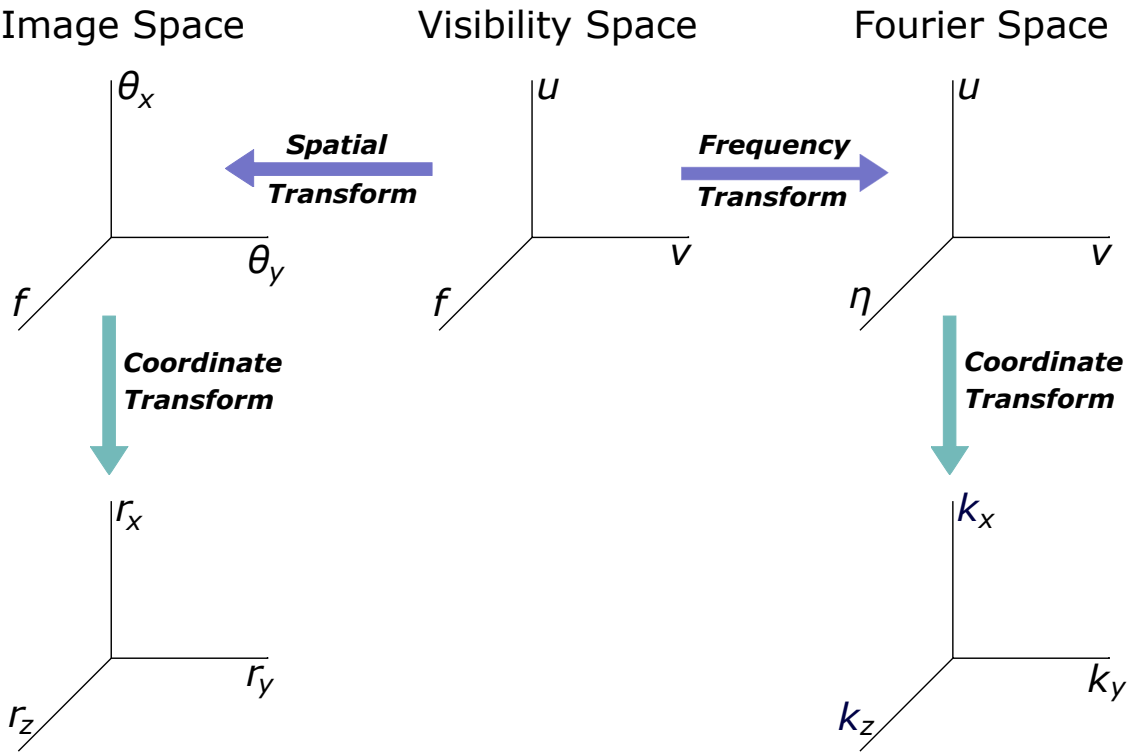


Figure 2.6: The various transformations needed to switch between image, visibility, and Fourier space. The full Fourier space $\{u, v, \eta\}$ and sky/frequency coordinates $\{\theta_x, \theta_y, f\}$ are related by Fourier transforms. The comoving distance space $\{r_x, r_y, r_x\}$ and wavenumber space $\{k_x, k_y, k_z\}$ are related to image space and Fourier space by coordinates transformations, respectively.

Chapter 3

DATA ANALYSIS PIPELINE

The MWA measures various modes on the sky given interference between pairs of antennas. A visibility is generated for every antenna pair *and* frequency channel *and* integration time *and* polarization, which amounts to over 347 million data points per typical observation in a space which is neither an image nor a power spectrum. As such, radio interferometers are often referred to as *software arrays* due to the sheer amount of data to be processed and Fourier-transformed even for the most basic of analyses.

Data analysis pipelines therefore must be very sophisticated, especially for a precision measurement like the search for the EoR. Improving this critical element is the main focus of the research performed at UW, and is core to this thesis. This chapter outlines the main pipeline components to build the infrastructure needed for future context, focusing on particular subjects like calibration.

In addition, we detail our final power spectrum products extensively. We perform many diagnostics with various types of power spectra, which provide justifications and confirmations for our improvements to the pipeline.

3.1 Fast Holographic Deconvolution

Fast Holographic Deconvolution, or FHD[1], is an interferometric radio analysis package that produces calibrated sky maps from measured visibilities [97]. Built by Ian Sullivan and the UW team in the early development of the MWA, it serves as a quick testing sandbox for EoR science. Its twin pipeline, the Real Time System (RTS), is optimized differently and less flexible [56]. Therefore, FHD has been the ongoing prototype for the past 6 years, fine-tuning

[1] `https://github.com/EoRImaging/FHD`

analysis techniques for implementation in other pipelines.

FHD is a learning tool in how to analyze radio interferometric data. It has a series of main functions at its core: 1) creating model visibilities from sky catalogs for calibration and subtraction, 2) gridding calibrated data and transforming into sky space, and 3) making images for analysis and integration. In order to understand each of these stages, we will describe the signal path and the measurement equation in detail. Using that framework, we will then describe the steps in building a gridding kernel, forming model visibilities, calibrating data visibilities, gridding all data products, and making images.

3.1.1 The signal path framework

The sky signal is modified during the journey from when it was emitted to when it was recorded. There are three main types of signal modification: those that occur before, during, and after interaction with the antenna elements, as shown in Figure 3.1. We adhere to notation from [33] whenever possible in our brief catalog of interactions.

1. **I**: Before antenna
    - Faraday rotation from interaction with the ionosphere, **F**.
    - Source position offsets **O** due to variation in ionospheric thickness.
    - Time correlations caused by unchecked RFI, **U**.
2. **S**: At antenna
    - Parallactic rotation **P** between the rotating basis of the sky and the basis of the antenna elements.
    - Antenna correlations from cross-talk between antennas, **X**.
    - Nominal configuration of the antenna elements and the interferometer as a whole, **C**, usually referred to as the beam.

- Errors in the expected nominal configuration, $\mathbf{D}$[2].

3. $\mathbf{E}$: After antenna

   - Electronic gain amplitude and phase $\mathbf{R}$ from a typical response of each antenna.
   - Gain amplitude changes from temperature effects $\mathbf{T}$ on amplifiers.
   - Gain amplitude and phase oscillations $\mathbf{K}$ due to cable reflections, both at the end of the cables and at locations where the cable is kinked.
   - Frequency correlations $\mathbf{A}$ caused by aliasing in polyphase filter banks.

Each contribution can be modeled as a matrix which depends on $[f, t, PP', AB]$: frequency, time, antenna polarization products, and antenna cross-correlations. The expected contributions along the signal path are thus $\mathbf{I} = \mathbf{UOF}$, $\mathbf{S} = \mathbf{DCXP}$, and $\mathbf{E} = \mathbf{AKTR}$. The measurement equation takes the form

$$\mathbf{v}^{\text{meas}} \sim \mathbf{E}\,\mathbf{S}\,\mathbf{I}\,\mathbf{v}^{\text{true}} + \text{noise}, \tag{3.1}$$

where $\mathbf{v}^{\text{meas}}$ are the measured visibilities, $\mathbf{v}^{\text{true}}$ are the emitted visibilities, $\mathbf{I}$ are contributions that occur between emission and the ground, $\mathbf{S}$ are contributions from the antenna configuration, and $\mathbf{E}$ are contributions from the electronic response. The visibilities can be condensed into vectors since the measurements are naturally independent over the $[f, t, PP', AB]$ dimensions. We have included all known contributions and modifications to the signal. However, there are most likely contributions that we did not include, hence we describe Equation 3.1 as an approximation.

One of FHD's goals is to reconstruct the true sky visibilities, $\mathbf{v}^{\text{true}}$, given the measured sky visibilities, $\mathbf{v}^{\text{meas}}$. We will use our generalized framework described in this subsection to detail the assumptions and methodology for the FHD pipeline.

---

[2]We deviate somewhat from [33] here. Any polarization correlation from [33] was counted as an error in the nominal configuration, whereas we expect and model polarization correlation in the Jones matrices.

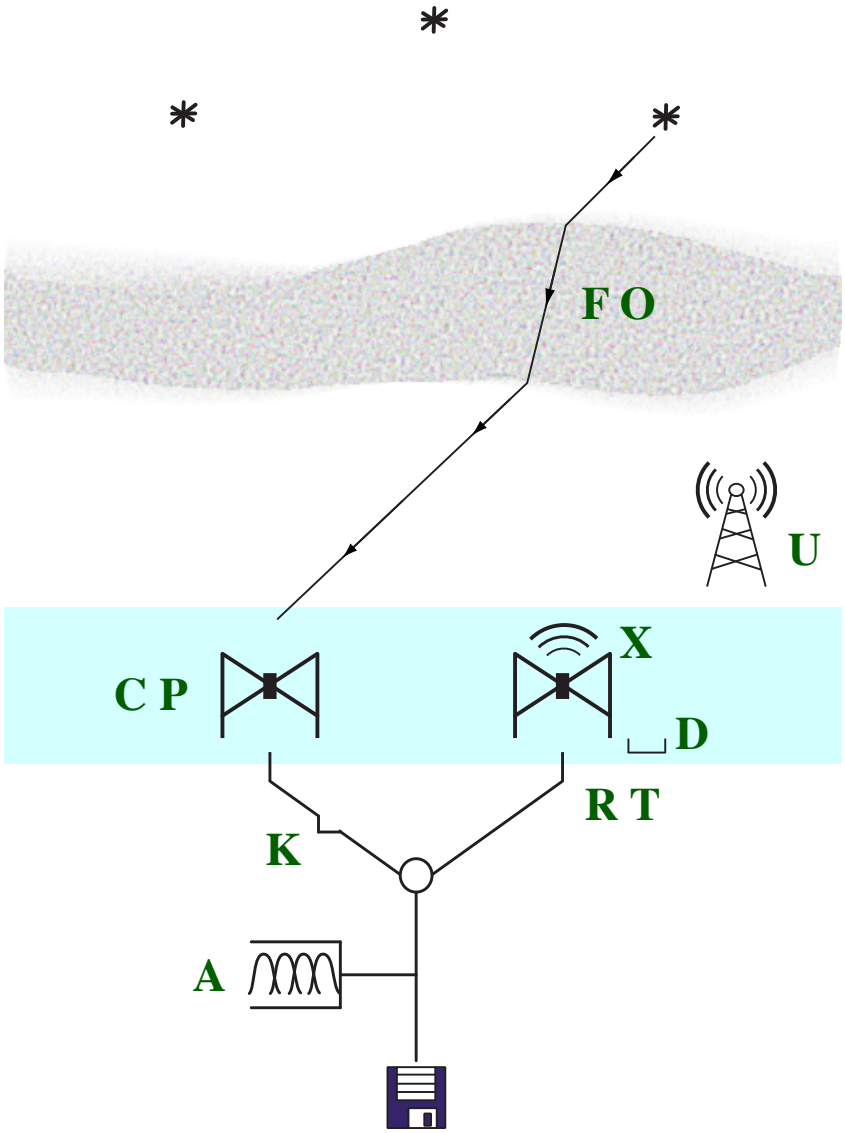


Figure 3.1: The signal path through the instrument. There are three categories of signal modification: before antenna, at antenna (colored light blue), and after antenna. Each modification matrix (green) is detailed in the text and Table 3.1.

Type	Variables	Definition
I	**F**	Faraday rotation
	O	Source position offsets
	U	RFI time correlations
S	**P**	Parallactic rotation
	X	Cross-talk antenna correlations
	C	Nominal configuration
	D	Errors in the nominal configuration
E	**R**	Electronic gain amplitude and phase
	T	Temperature amplitude changes
	K	Cable reflections
	A	Frequency correlations
v	$\mathbf{v}^{\text{true}}$	Emitted visibilities
	$\mathbf{v}^{\text{meas}}$	Measured visibilities
	noise	Independent thermal noise

Table 3.1: Brief definitions of the variables used within the signal path framework, organized by type. There are four types, **I**) interactions that occur before the antenna elements, **S**) interactions that occur at the antenna elements, **E**) interactions that occur after the antenna elements, and **v**) visibility-related variables.

3.1.2 Pre-pipeline flagging

Before any analysis can begin, the data must be RFI-flagged, averaged, and formatted. Radio frequency interference (RFI) in our data is an issue due to the popularity of FM radio and digital TV. However, RFI has characteristic signatures in time and frequency which allow it to be systematically removed by trained packages.

We use the package AOFLAGGER[3] to RFI-flag the data [68]. This removes bright line-like emission, but has difficulty removing faint, broad emission like TV. Luckily, the radio

[3] `https://sourceforge.net/p/aoflagger/wiki/Home/`

environment at the Murchison Widefield Array is extremely quiet [68], especially in comparison to other low-frequency arrays [66]. We completely remove any observations that have signatures of TV, which does not occur often. Therefore, we remove contributions from $\mathbf{U}$.

3.1.3 Generating the beam

The measurement collecting area of the instrument, $\mathbf{C}(\theta, \phi, f)$, is commonly referred to as the **beam**. A deep knowledge of the beam is critical for precision measurements of faint astrophysical signals, especially with widefield interferometers. Since our visibility measurements are correlations of the voltage response between tiles, we must understand the footprint of each tile's voltage response to reconstruct images [59].

A lot of research has gone into the generation of image beam models for the MWA, including both experimentally determined beam shapes [61] and theoretical modeling [100, 99]. We input a theoretical dipole beam image as a function of RA/DEC and frequency, and smoothly interpolate to a fine spatial resolution. We then build a tile beam with this fine dipole beam. This happens in the instrument's **coherency domain**, or instrumental polarization. Each dipole inherently has an X or Y response due to its design (see §2.2). Therefore, each correlation will have either have an XX, XY, YX, or YY response in the coherency domain [33]. We build these by calculating the dot product between the two tiles' response patterns,

$$\mathbf{C}_{XX}(i,j,f) = \mathbf{C}_{X_1X_2}(i,j,f) = \mathbf{J}_{X_1\to i}\mathbf{J}^*_{X_2\to i} + \mathbf{J}_{X_1\to j}\mathbf{J}^*_{X_2\to j}, \tag{3.2}$$

where $\mathbf{C}_{XX}(i,j,f)$ is the beam in the XX instrumental polarization as a function of the coordinate system (i,j) and frequency, $\mathbf{J}$ is the vector field for a tile of dipoles (also known as the **Jones matrix**), and the subscript $X_1 \to i$ is the contribution for the first tile of X dipoles in the coordinate i (and likewise for the other subscripts)[4] [99]. The Jones matrices

[4]We have described $\mathbf{C}_{XX}$ as a function of generalized coordinates (i,j). At this point, we have discretized square pixels which we describe in (x,y), but it is confusing to have X and x in the same equation where X refers to a particular object and x refers to the coordinate.

describe the transformation needed to account for $\mathbf{P}$, or parallactic rotation, and known inter-dipole mutual coupling. We generate $\mathbf{C}_{XX}(i,j,f)$, $\mathbf{C}_{XY}(i,j,f)$, $\mathbf{C}_{YX}(i,j,f)$ and $\mathbf{C}_{YY}(i,j,f)$ for a pair of tiles, which can be applied to all other tile pairs.

FHD has the flexibility to generate personalized beams for each tile. At this point, more complex beams can be made by modifying the overall intensity of dipole beams within the tile to match known metadata and on-site testing. This corrects for one form of $\mathbf{D}$, or tile errors due to dead or ineffective dipoles. New beams are generated for each new tile pair possibility, which greatly increases memory usage. We have found that this has not significantly changed results thus far, but the infrastructure is in place for future testing. We have also found that cross-talk $\mathbf{X}$, or potential inter-tile reflections, has not been an issue. Therefore, only one beam template is needed for all tile pairs at this time.

We then take the various beams and Fourier transform them along the spatial dimensions to get beams in uv-space, or $\mathbf{C}_{XX}(u,v,f)$, $\mathbf{C}_{XY}(u,v,f)$, $\mathbf{C}_{YX}(u,v,f)$ and $\mathbf{C}_{YY}(u,v,f)$. Even though a fine resolution dipole beam with a wide field-of-view was used in forming the image, we would like to have even finer uv-beams. Since we plan on using the beam as a gridding kernel later on, it is very important that the beam varies as smoothly as possible. This is achievable by hyperresolving even more, down to typically $\frac{1}{7000}\lambda$.

The beams we have made will be used during our gridding process to take visibilities and make an image, and vice versa. Our hyperresolved approach is critical to maintaining smooth contributions to minimize spectral contamination from performing Fourier transforms.

3.1.4 Creating model visibilities

We must calibrate our input visibilities. Due to our wide field-of-view and accuracy requirements, this must be done in ways that deviate from traditional radio astronomy methods [102]. FHD achieves this by simulating all reliable sources that the instrument can see, and comparing the data visibilities to these **model visibilities**. This type of analysis is referred to as sky calibration, and from this, we can estimate the instrument's contribution to the data visibilities. In addition, model visibilities can be used in subtracting known sky

contributions.

Generating accurate model visibilities is an important step in calibration. Therefore, the estimate of the sky, including source positions, morphology, and brightness, must be as complete as possible. Cataloging these features from all reliable sources in the field EoR0 has been the work of [11, 12], and is the main catalog (KGS) used in all analyses in this thesis. KGS is very reliable in the primary beam, but we must resort to other catalogs in the sidelobes. Therefore, we use GLEAM [39], a catalog with large coverage and high completeness, in the sidelobes to model more sources reliably.

We take the RA/DEC floating-point locations and Stokes I brightness of unresolved point sources and perform a discrete Fourier transform to the *uv*-plane. A discrete Fourier transform must be used due to the nonregularity of the locations of the sources. This significantly impacts efficiency of the code because of the sheer number of reliable sources; over 11,000 known sources exist in the field EoR0 with our catalogs! However, we must include as many sources as possible to create an accurate model[5].

Not all sources are only bright in Stokes I, or in other words, are unpolarized. There are cases where a source can be bright in Stokes Q (horizontally or vertically polarized), Stokes U (diagonally polarized), or Stokes V (circularly polarized). Only five out of the thousands of sources seen in the EoR0 field are known to be reliably polarized [44]. Therefore, we assume there are no polarized sources and only make model visibilities from Stokes I sources. With this assumption, we avoid complications from $\mathbf{F}$, or Faraday rotation in the polarization components due to the ionosphere.

Not all sources are unresolved. We also can optionally include extended sources by modeling their contribution as a series of unresolved point sources [11]. This can also be done for creating models of diffuse synchrotron emission [2]. However, the diffuse emission is significantly polarized [44] and difficult to include.

Once a model *uv*-plane is created with all source contributions, we simulate what the

[5]Why this level of accuracy is needed is discussed in §6.

instrument actually measured. The hyperresolved uv-beam from §3.1.3 is the sensitivity of the cross-correlation of two tiles. We calculate the uv-locations of each cross-correlated visibility, and multiply the model uv-plane with the uv-beam sampling function. The sum of the sensitivity multiplied by the model at the sampled points is the estimated measured value for that cross-correlated visibility.

These visibilities represent our best estimate for what the instrument should have measured, disregarding any source position offsets $\mathbf{O}$ from varying thicknesses in the ionosphere. Any deviations from these model visibilities (whether instrumental or not) will be captured during the comparison between the data and model during calibration.

3.1.5 Calibration

We are left with one type of modification to the signal that has not been accounted for by the model visibilities or beam: $\mathbf{E}$, the electronic response. This is what we classify as our calibration, and here we go into more mathematical detail for this process.

At this point in the analysis, we are left with a measurement equation that looks like $\mathbf{v}^{\text{meas}} \sim \mathbf{E}\,\mathbf{m} + \text{noise}$, where $\mathbf{m}$ are the estimated model visibilities, which capture effects that occur before the antenna and at the antenna. The electronic response $\mathbf{E}$ varies slowly with time, and thus does not change significantly over a two-minute observation. Due to our model-based assumptions, we do not have any time correlations, antenna correlations, or unknown polarization correlations. The electronic response is thus simply a time-independent gain $\mathbf{G}$ per observation which is independent per antenna and polarization.

We begin by rewriting Equation 3.1 using these assumptions. The measured cross-correlated visibilities are a function of frequency, time, and polarization. The sets of antennas $A = \{a_0, a_1, a_2, \ldots, a_{127}\}$ and $B = \{a_0, a_1, a_2, \ldots, a_{127}\}$ are iterated through independently to calculate these cross-correlations. A visibility is measured for each linear polarization $P = \{X, Y\}$ and $P' = \{X, Y\}$, creating the polarization product, $PP' = \{XX, XY, YX, YY\}$.

The resulting relation between the measured visibilities and the model visibilities is

$$\mathbf{v}_{AB,PP'}([f_o,t]) \sim \mathbf{G}_{A,P}(f_o,f_i)\mathbf{G}^*_{B,P'}(f_o,f_i)\mathbf{m}_{AB,PP'}([f_i,t]) + \mathbf{n}_{AB,PP'}([f_o,t]), \tag{3.3}$$

where $\mathbf{v}_{AB,PP'}([f_o,t])$ are the measured visibilities and $\mathbf{m}_{AB,PP'}([f_i,t])$ are the model visibilities. Both are frequency and time vectors $[f,t]$ of the visibilities over all A and B antenna pairs and over all P and P' polarization products. $\mathbf{G}_{A,P}(f_o,f_i)$ is a frequency matrix of gains given input frequencies f_i which affect multiple output frequencies f_o for instrumental polarization P for antenna A (and likewise for B). Thermal noise $\mathbf{n}$ is independent for each visibility.

Our notation has been specifically chosen. Naturally discrete variables, antenna pairs and polarization products, are described in the subscripts. Naturally continuous variables, frequency and time, are function arguments. We group frequency and time into a set $[f,t]$ for the visibilities to create vectors. Since frequency and time are independent, this notation is more compact. In contrast, the gain matrices $\mathbf{G}$ are not independent in frequency. A full matrix must be used to accurately capture correlation due to aliasing in the polyphase filter banks (see §2.2.3 for more information on PFBs and their effects).

However, we make the assumption that the frequencies are independent (contribution $\mathbf{A} \to \mathbb{I}$) to reduce Equation 3.3 significantly,

$$\mathbf{G}_{A,P}(f_o,f_i) \to \mathrm{diag}(\mathbf{g}_{A,P}(f)). \tag{3.4}$$

The instrumental gains $\mathbf{g}$ are now an independent vector of frequencies for antenna A (and likewise for B) per instrumental polarization P. We flag frequency channels which are most affected by aliasing to make this assumption viable. However, this does introduce contamination in confined areas in the power spectrum due to the regularity of the flagging.

We can fully vectorize the variables in Equation 3.3:

$$\mathbf{v}_{AB,PP'}([f,t]) \sim \mathrm{diag}(\mathbf{g}_{A,P}(f))\,\mathrm{diag}(\mathbf{g}^*_{B,P'}(f))\mathbf{m}_{AB,PP'}([f,t]) + \mathbf{n}_{AB,PP'}([f,t]). \tag{3.5}$$

The gains in Equation 3.3 and Equation 3.5 will encode systematics between the model visibilities and the true visibilities due to the imperfection of the model. This is not instru-

Type	Variables	Definition
Observation Parameters	f	Measured frequencies of the observation.
	P	Instrumental polarizations $\{X, Y\}$. Also a set.
	a	Antennas in the array.
	t	Time steps within an observation.
Sets	$A = \{a_0, \ldots, a_{127}\}$	All antennas of the array. A and B can be iterated separately to form cross-correlated antenna pairs.
	$B = \{a_0, \ldots, a_{127}\}$	
	$C = \{c_2, c_1, c_0, \phi_1, \phi_0\}$	Coefficients of a 2^{nd} order amplitude polynomial and a linear phase fit across the frequency band.
	$[f, t]$	A combined set of all frequencies and times.
Groups, Matrices, & Vectors	$\boldsymbol{\alpha}_L$	An antenna grouping, where parameters are per antenna group rather than per antenna.
	$\mathbf{G}_{A,P}(f_o, f_i)$	The full gain matrix for each antenna in group A per P. Input and output frequencies are correlated.
	$\mathbf{g}_{A,P}(f)$	A vectorized approximation of the gains $\mathbf{G}$ for each antenna in the group A per P over frequency.
	$\mathbf{m}_{AB,PP'}([f, t])$	Vector of simulated model visibilities for each antenna pair AB and polarization product PP' over $[f, t]$.
	$\mathbf{n}_{AB,PP'}([f, t])$	Thermal noise vector for each antenna pair AB and polarization product PP' over $[f, t]$.
	$\mathbf{v}_{AB,PP'}([f, t])$	Uncalibrated visibilities vector for each antenna pair AB and polarization product PP' over $[f, t]$.

Table 3.2: Definitions of all variables used to describe our calibration procedure in this subsection.

mental, thus the calibration solutions will be contaminated. In later chapters, we describe our attempts to remove some of this effect, based on work done in [1]. Our model currently does not include large-scale diffuse emission, so we constrain our calibration to be calculated only from visibilities from baselines longer than $50\,\lambda$ to reduce overfitting effects [76].

Equation 3.5 can be used to solve for the instrumental gains for all frequencies and polarizations independently. This allows the use of Alternating Direction Implicit (ADI) methods for fast and efficient solving of $\mathcal{O}(N^2)$ [90]. Due to this independence, parallelization can also be applied. Noise is ignored during the ADI for simplicity; true thermal noise is Gaussian and will average out during the ADI. However, if any noise has a non-Gaussian distribution, it will contribute to the instrumental gains.

We begin solving Equation 3.5 by estimating an initial solution for $\mathbf{g}^*_{B,P'}(f)$. This choice must force the gains into a region with a local minimum for the ADI method. A good choice is the average gain expected across all antennas, and in practice a set of 1's has been sufficient for the MWA. For a more complicated method, scaled auto-correlations have also been used.

With an input for $\mathbf{g}^*_{B,P'}(f)$, Equation 3.5 can then become a linear least-squares problem.

$$\boldsymbol{\chi}^2_{A,P}([f,t]) = \sum_B \left| \mathbf{v}_{AB,PP'}([f,t]) - \mathrm{diag}(\mathbf{g}_{A,P}(f))\ \mathrm{diag}(\mathbf{g}^*_{B,P'}(f))\ \mathbf{m}_{AB,PP'}([f,t]) \right|^2, \quad (3.6)$$

where $\mathbf{g}_{A,P}(f)$ is found given a minimization of $\boldsymbol{\chi}^2_{A,P}([f,t])$ for each antenna A and instrumental polarization P. All time steps are used to find the temporally constant gains over the two-minute observation. We have also assumed $PP' \rightarrow PP$ (resulting in $\{XX, YY\}$ only) for computation efficiency since these contributions are most significant. A full polarization treatment is being developed.

The current estimation of $\mathbf{g}^*_{B,P}(f)$ is then updated with knowledge from $\mathbf{g}_{A,P}(f)$ by adding together the current and new estimation and dividing by 2. By updating in partial steps, a smooth convergence is ensured. The linear least-squares process is then repeated with an updated $\mathbf{g}^*_{B,P}(f)$ until convergence is reached[6].

[6]We have found that allowing the first 10 iterations to only update the phase of $\mathbf{g}^*_{B,P}(f)$ helps to converge faster.

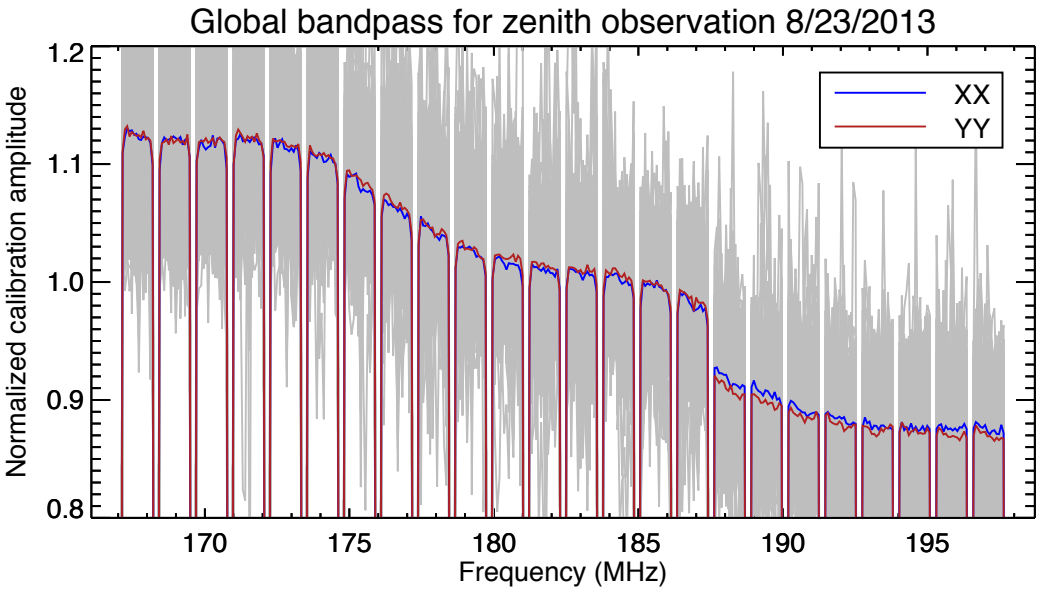


Figure 3.2: The global bandpass for the zenith observation of August 23, 2013 for instrumental XX (blue) and YY (red). All the per-frequency solutions that went into the global bandpass average are shown in the background (grey). This historical approach greatly decreased expected noise on the solutions.

The resulting gains from the least-squares iteration process are completely independent in frequency, antenna, time, and polarization. This is not entirely correct. We also did not account for noise contributions during the per-frequency ADI fit; this adds spurious deviations from the gain's true value with mean of zero. Historically, we accounted for these effects by creating a **global bandpass,**

$$|\mathbf{g}_P(f;\boldsymbol{\alpha})| = \left\langle \left|\mathbf{g}_{A,P}(f)\right| \right\rangle_{\boldsymbol{\alpha}}, \tag{3.7}$$

where the normalized amplitude average is taken over all antennas $\boldsymbol{\alpha}$ as a function of frequency to create a global bandpass $|\mathbf{g}_P(f;\boldsymbol{\alpha})|$ independent of antennas. Figure 3.2 shows an example of the global bandpass alongside the noisy per-frequency inputs. We did not account for cable reflection effects $\mathbf{K}$ with this methodology. For updated methods, see §4.

An overall amplitude still must be accounted for within the gains. These vary slowly per antenna and differ from day to day, therefore they cannot be included in the average bandpass. We have found that temperature is the main cause in the amplitude variations due to the temperature dependence of the amplifiers, which we have described as $\mathbf{T}$ in our signal path framework. This dependence varies slightly as a function of frequency, and is easily characterized with a low-order polynomial.

In addition to fitting polynomials to the amplitude as a function of frequency, we also fit the phase. The phase is extremely linear for the MWA. We have found that using a per-antenna linear fit as the calibration phase solution has been a good estimate.

For the amplitude, we fit

$$c_2 f^2 + c_1 f + c_0 = \mathcal{P}\left(\frac{|\mathbf{g}_{A,P}(f)|}{|\mathbf{g}_P(f;\boldsymbol{\alpha})|} \right), \tag{3.8}$$

where the global bandpass contribution, $|\mathbf{g}_P(f;\boldsymbol{\alpha})|$, is removed before the fit and c_2, c_1, and c_0 are the resulting coefficients. For the phase, we fit

$$\phi_1 f + \phi_0 = \mathcal{P}\left(\arg \frac{\mathbf{g}_{A,P}(f)}{|\mathbf{g}_P(f;\boldsymbol{\alpha})|} \right), \tag{3.9}$$

where the polynomial fit is done over the phase of the residual and ϕ_1 and ϕ_0 are the resulting coefficients. Due to phase jumps between $-\pi$ and π, special care is taken to ensure the function is continuous across the π boundary[7]. We can create a set of these coefficients, $C = \{c_2, c_1, c_0, \phi_1, \phi_0\}$, for easy reference.

Our final calibration solution is

$$\mathbf{g}_{A,P}(f;\boldsymbol{\alpha},[C]) = \underbrace{|\mathbf{g}_P(f;\boldsymbol{\alpha})|}_{\textbf{global bandpass}} \underbrace{(c_2 f^2 + c_1 f + c_0)\ e^{i(\phi_1 f + \phi_0)}}_{\textbf{per antenna gain and phase}}. \tag{3.10}$$

Improvements have been made on this historical solution, which will be described in §4 & §7.

[7]We "unwrap" the phase to account for this, where we take the Riemann sheets and create a new continuous plane. We then solve, and "rewrap." There can be ambiguity in which Riemann sheet to place the phase if the phase varies quickly, but this is not an issue with the MWA.

3.1.6 Imaging

The final goal of FHD is to take calibrated data visibilities and transform them into a space where they can be combined across observations. We have already described the reverse algorithm we need in §3.1.4. Instead of taking a sky and making visibilities, we would like to take visibilities and make a sky. We do this through a process called **gridding**.

We convolve the visibility value by the beam sensitivity calculated in §3.1.3. Essentially, we take an integrated value and de-integrate it given our knowledge of the beam in uv-space. This will make an estimate of the instrument voltage response for that location in uv-space. We continue to **grid** the uv-plane with every voltage response estimate from every visibility for a set of regular gridding points.

We perform gridding to the uv-plane for calibrated raw visibilities, model visibilities, and the residual visibilities generated from their difference. By gridding each separate data product, we can make diagnostic images of all of them. Images generated from residual visibilities help to ascertain the quality of foreground removal in image space, and are thus an important aspect in quality assurance.

This final uv-plane generated from our visibilities is not completely filled; there are modes we did not measure because there wasn't an antenna pair with that spacing. Even more, some modes are measured many times and some modes are measured once or twice. Therefore, we need to weight our generated uv-plane.

There are choices on particular weighting schemes: some have high resolution but have high noise, and others have low noise but have low resolution. We choose high-resolution, high-noise weightings (i.e. uniform weighting) if images are the goal, and we choose low-resolution, low-noise weightings (i.e. natural weighting) if EoR science is the goal.

In addition to gridding the data visibilities, we grid visibilities of value 1 with the beam gridding kernel to generate natural uv-space weights. This effectively gives us a **sampling map** which describes how much of a measurement went into each pixel. We also create a map where we grid with the beam-squared kernel. This is called a **variance map** and will

be used in error propagation in later packages.

We then have three types of *uv*-plane products: the sampling map, the variance map, and the data-generated *uv*-planes. In addition, we create weighted-data *uv*-planes using the sampling map and data-generated *uv*-planes. All these products are transformed via 2D FFTs (Fast Fourier Transforms) to image space in slant orthographic projection, which is a flat projection of the sky that is slanted to be parallel with the measurement plane (see §3.2.1 for more discussion on what this entails).

At this point, the various images are made for two different purposes. The sampling map, variance map, and data-generated image planes are for future packages. The weighted-data image plane is for diagnostic images per two-minute observation. We divide by the average beam in image space and combine instrumental polarizations to generate the true sky in Stokes parameters for the diagnostics. The result of calibrated data and residual snapshot images is shown in Figure 3.3.

However, our slant orthographic images are in a basis that changes with LST[8]. Therefore, we interpolate to HEALPix pixel centers, or the Hierarchical Equal Area isoLatitude Pixelization of a sphere [31], which is the same for all LSTs. We interpolate all data products—calibrated data, model, sampling map, and variance map—to HEALPix pixel centers *separately* for use in future packages. If we want to combine multiple observations, we will need to do a weighted average. Therefore, we keep the numerator (data) separate from the denominator (sampling map) for this purpose. FHD outputs all of these data products, which we can then analyze with other packages.

3.2 Integration

In order to reduce noise, we must now integrate. Since our signal is very small, we must integrate for long periods of time. The MWA can theoretically detect the EoR with a signal-to-noise of 7.1 on the amplitude with an integration of 450 hours [3], or about half of an

[8]Local Sidereal Time: a basis that fixes Earth rotation time to celestial objects.

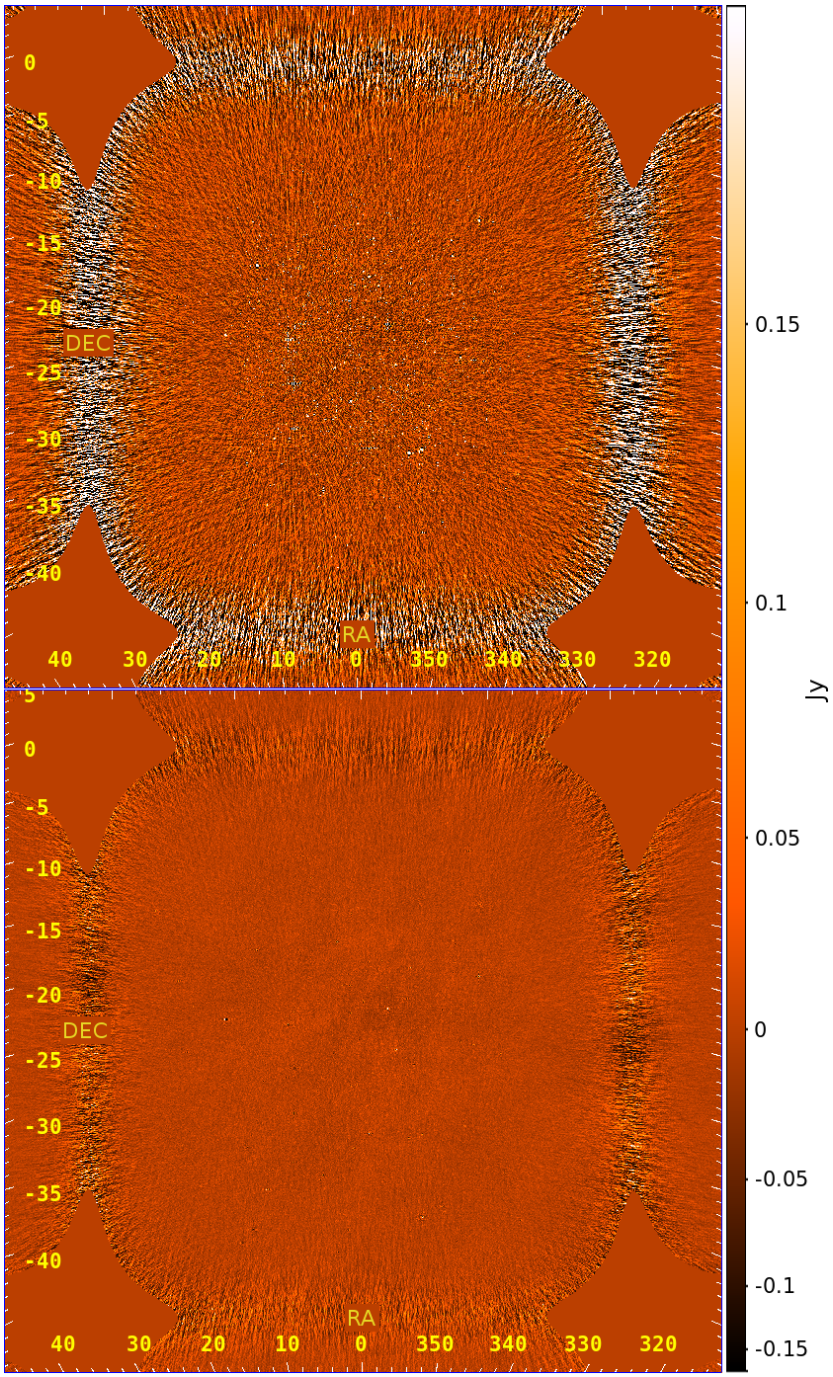


Figure 3.3: The calibrated data (top) and residual (bottom) images for the zenith observation of August 23, 2013 in apparent brightness, in uniform weighting, and in zero-mean. There is significant reduction of sources and point spread functions.

observing season. This requires integration of thousands of observations.

FHD outputs a variety of cubes that we must now integrate together before we input the cubes into our next package, εppsilon (described in §3.3). These cubes are numerators and denominators which create meaningful maps in *uv*-space; they have no well-defined meaning in image space. The various cubes are:

Calibrated data cubes: HEALPix images for each frequency of the unweighted calibrated data in units of $[\frac{\mathrm{Jy}^2}{\mathrm{str}}]$.

Model data cubes: HEALPix images for each frequency of the unweighted generated model in units of $[\frac{\mathrm{Jy}^2}{\mathrm{str}}]$.

Sampling map cubes: HEALPix images for each frequency of the sampling map, or the estimate of how well-sampled each mode is by the instrument, in units of $[\frac{\mathrm{Jy}}{\mathrm{str}}]$.

Variance map cubes: HEALPix images for each frequency of the variance map, or the estimate of the squared sampling of each mode, in units of $[\frac{\mathrm{Jy}^4}{\mathrm{str}}]$.

These cube types are split by instrumental polarization and by an interleaved time sampling set of even and odd time indices. Each of the cubes are added together, pixel by pixel, using the HEALPix coordinate system. The choices of separating even and odd time sampling sets, keeping the data and the weighting maps separate, and performing this integration in image space rather than *uv*-space are all very deliberate.

3.2.1 Averaging in image space

Averaging *in image space*, rather than in *uv*-space, is a crucial aspect of the analysis. Either approach can theoretically be used, but the computational requirements vary greatly.

The measurement *uv*-plane appears non-parallel to the tangent plane of the observed sky for measurements that are not at zenith. For example, a plane wave from near the horizon will be measured at one end of the array sooner than the opposing side. The instrument's measurement plane appears tilted compared to the wave's propagation direction; this tilt

causes a measurement delay. Depending on the tilt angle, a source's plane wave will be measured at different phases in its wave propagation across the array. This uv-plane cannot be added to a uv-plane pointing in a different direction since each measurement appears tilted in a different way. This is a classic decoherence problem, except that it is in Fourier space.

There are two methodologies to fix this geometry problem. The first is to project the measurement uv-plane to be parallel to the tangent plane of the observed sky. This is called w-projection [15] since it projects $\{u, v, w\}$ space to $\{u, v, w = 0\}$ space. Unfortunately, the projection requires the propagation of a Fresnel pattern to reconstruct precisely what the wave looked like on the $w = 0$ plane. This must be done for every observation and visibility! This requires intense computational overhead, but has been successfully implemented by other packages [110].

The second methodology is to fix the geometry problem in image space. Instead of adding uv-planes, we add image planes. We create an image for each observation by performing a 2D spatial FFT of the uv-plane, which results in a slant orthographic projection of the sky [10]. This projection is parallel to the uv-plane due to the fact that no w-terms were used. This inherently assumes the array is **coplanar**; the integration time must be small enough such that the measurement plane does not vary much over the course of the observation and altitude deviations must be accounted for with accurate calibration phases.

We can then easily interpolate from the slant orthographic projection to the HEALPix projection [71], which is a constant basis. We do not need to propagate waves in image space, thus it is computationally efficient in comparison. The HEALPix maps of each observation are now projected the same and are in the same space, allowing for an integration method that is manageable.

3.2.2 Interleaved cubes

The integration cubes are split up by interleaved time steps. We group these into time steps that have an even index and time steps that have an odd index for the observation's two-

second cadence over 112 seconds. The even–odd distinction is arbitrary; what really matters is that they are interleaved. While this doubles the number of Fourier transforms to perform, it allows for crucial error analysis.

The sky does not vary in any substantial way over two seconds. Any significant variation can be attributed to either RFI (which has been accounted for in §3.1.2) or thermal noise on the observation. By subtracting the even–odd groupings and enforcing consistent flagging, we should be left with thermal noise contribution to the observation.

We carry even–odd interleaved cubes throughout the power spectrum analysis and check at various stages against the noise calculation. It is a robust way to ensure that error propagation and normalization has occurred correctly. In addition, a calculation of the thermal noise allows us to generate a cross power spectrum.

3.3 Error Propagated Power Spectrum with Interleaved Observed Noise

Error Propagated Power Spectrum with Interleaved Observed Noise, or εppsilon[9], is a power spectrum analysis package designed to take integrated images and create various types of diagnostic and limit power spectra. It was created by Bryna Hazelton as a way to propagate errors into power spectrum space, rather than have *estimated* error bars. Other image-based power spectrum analyses exist for radio interferometric data [22, 79, 110], but εppsilon is unique in this respect.

The main functions of εppsilon are to 1) transform integrated images into $\{u, v, f\}$ space, 2) calculate observed noise using even–odd interleaving, 3) transform frequency to k-space, and 4) average k-space voxels together for diagnostic and limit power spectra. We will focus on these main points in the following subsections, which will help to build the groundwork needed to describe 2D power spectra, 1D power spectra, and their respective error bars.

---

[9] `https://github.com/EoRImaging/eppsilon`

3.3.1 From integrated images to uv-space

The input products of εppsilon are image cubes that have been integrated over many observations, and are a function of RA, DEC, and frequency. While this space was useful and well-defined for integration, it is not useful nor well-defined in creating error-propagated power spectra. Our measurements were inherently taken in uv-space, and that is where our error bars on the measurement are easy to understand. The error bars in image space, however, are a mix of *all* mode errors, which is difficult to propagate. Therefore uv-space is the natural space in which to propagate errors.

In going back to uv-space, we are faced with a complication: not all image pixels on the sky have contributions from every observation. The observations within a pointing measure slightly different portions of the sky. We keep a square portion of the sky where all observations contribute and with an extent that provides the necessary resolution in uv-space[10]. See §8.3 for an update to this procedure.

We then Fourier transform the integrated image back into uv-space using a direct Fourier transform between the curved sky and the flat, regularly spaced uv-plane. This differs from the methodology described in §3.1.6. Instead of an interpolation to a regular grid in image space and performing a FFT, we DFT a non-regular grid directly to the uv-plane. Since this happens once per observation integration *set,* rather than once per observation, we can perform the slower process of a DFT without too much overhead.

After transforming into uv-space, we calculate the resulting $\{k_x, k_y\}$ values for each pixel. The uv-space is related to the wavenumber space by the simple transforms [58]

$$k_x = \frac{u2\pi}{D_M(z)} \qquad k_y = \frac{v2\pi}{D_M(z)}, \tag{3.11}$$

where $D_M(z)$ is the transverse comoving distance dependent on redshift [37], calculated from the cosmological parameters of the universe.

In addition, we also perform the 3D pixel-by-pixel subtraction of the integrated model from the integrated data to create residual cubes. We could perform this at the end of the

[10]We actually apply this stipulation in FHD during the image-making process to reduce computation.

analysis after we have transformed $f \rightarrow k_z$ to save on computation time, but it is helpful to make diagnostic plots of the residual cube in $\{u, v, f\}$ space.

3.3.2 Mean and noise calculation

We have twenty various *uv*-products: calibrated, model, residual, sampling map, and variance map data, for each polarization product XX or YY, and for each interleaved even–odd time sample set. Each of these cubes are a function of two spatial dimensions in units of wavelengths and one frequency dimension in units of Hz. The actual measurements are not in meaningful units; they are numerators and denominators which form meaningful units when combined.

We have kept the numerators and denominators (data and weights) separate up until this point so that we can perform variance-weighted sums and differences. First, we weight the calibrated data, model, and residual by the sampling map, thereby upweighting strongly measured modes and downweighting weakly measured modes. Second, we weight the variance map by the square of the sampling map to scale our error bars with our choice of weighting scheme. We then perform sums and differences using our sampling-map-weighted error bars as the weights,

$$\hat{\mu} = \frac{\sum\limits_{n=e,o} \frac{x_n}{\sigma_n^2}}{\sum\limits_{n=e,o} \frac{1}{\sigma_n^2}} \qquad \hat{n} = \frac{\sum\limits_{n=e,o} (-1)^n \ \frac{x_n}{\sigma_n^2}}{\sum\limits_{n=e,o} \frac{1}{\sigma_n^2}}, \tag{3.12}$$

where $\hat{\mu}$ is the mean, $\hat{n}$ is the noise, $\{e, o\}$ are the interleaved even–odd sets, x_n is the sampling-map-weighted data (calibrated data, model, or residual) for a given even–odd set, and σ_n^2 is the sampling-map-weighted variance map for a given even–odd set. The calculated mean and noise are the **maximum likelihood estimates** for a weighted Gaussian probability distribution, hence using the variables $\hat{\mu}$ and $\hat{n}$ instead of μ and n.

We also perform the same weighting and maximum likelihood estimation for simulated cubes. We randomly generate a real distribution and an imaginary distribution of numbers given the sampling-map-weighted standard deviation and zero mean for each of the even–odd sets. Sum and difference cubes are then made the same way as in Equation 3.12; inverse-

weighted sums and differences to make maximum likelihood estimates of the signal and noise. In all of our subsequent data analysis steps, we will perform the same transformations, averages, and estimations to the simulations as we do to the data. By modifying pure simulation in the same way as the data and comparing at each step, we can be sure that we are not removing signal accidentally to lead to artificially optimistic limits. This will provide important justification for claiming no **signal loss** on our limits.

Finally, we also calculate our error bars using the maximum likelihood estimation:

$$\hat{\epsilon}^2 = \frac{1}{\sum\limits_{n=e,o} \frac{1}{\sigma_n^2}}, \tag{3.13}$$

where $\{e, o\}$ are the interleaved even–odd sets and σ_n^2 is the sampling-map-weighted variance map for a given even–odd set. Due to the propagation of errors, this maximum likelihood error is the same for all of our previously generated estimates regardless of addition or subtraction.

3.3.3 Transforming frequency to k-space

We now have a variety of cubes in $\{u, v, f\}$[11] space. In order to go to power spectrum space, we must perform a spectral analysis in the frequency direction to go from f to k_z. Due to the nature of the data, this includes several steps.

Fourier-transforming a relatively small, finite set of data will cause **leakage**. Finite data can be thought of as a top-hat function of the length of the set multiplied by the data—and a Fourier transform of two multiplied functions is the convolution of the Fourier transform of each function. The Fourier transform of a top-hat function is a *sinc* function. When convolved with the spectrum of the data, leakage occurs in other modes. We mitigate this by multiplying the data by a Blackman-Harris **window function** [34]. This will decrease the leakage from about –15 dB to about –90 dB at the first sidelobe, but at the cost of about half the effective bandwidth. To reduce leakage even more, we artificially force the data to be

[11] Related to $\{k_x, k_y, f\}$ through simple conversions, see §3.3.1 and §2.3.

close to zero mean by removing the mean of the data before applying the window function. We save this value to be added back in later.

The data is also not regularly spaced in frequency. We have deliberately flagged channels due to polyphase filter bank aliasing in §3.1.5, and we also have flagged RFI in §3.1.2. The array also does not regularly sample as a function of frequency due to baseline length evolving with frequency. As a result, the *sin* and *cos* basis functions of the Fourier transform are not orthogonal to the noise distribution on our sampling; the noise is not independent in the *sin* and *cos* basis. However, we can find a basis which is orthogonal. We use the Lomb-Scargle periodogram to find this basis [49, 91],

$$P_{X_j}(\eta) \propto \frac{\left(\sum_j X_j \cos(\eta(f_j - \tau))\right)^2}{\sum_j \cos^2(\eta(f_j - \tau))} + \frac{\left(\sum_j X_j \sin(\eta(f_j - \tau))\right)^2}{\sum_j \sin^2(\eta(f_j - \tau))}, \tag{3.14}$$

where X_j is the data in the j^{th} channel, η is the periodogram dual of frequency f, $P_{X_j}(\eta)$ is the power at that channel as a function of η, and τ is the basis rotation phase that creates orthogonality, which can be found with

$$\tan(2\eta\tau) = \frac{\sum_j \sin(2\eta f_j)}{\sum_j \cos(2\eta f_j)}. \tag{3.15}$$

$P_{X_j}(\eta)$ effectively loses the phase of the data, which is not an issue in our situation because our final goal is to create a power spectrum (a naturally phase-less product). We then add the mean value of the data removed during application of the window function to the zeroth mode of the *cos*-like eigenfunction. To calculate 3D power cubes, we add each squared eigenfunction together, weighting by the variance of their respective Rayleigh distributions. The noise on the k-space voxels follows an Erlang distribution [118], which has a well-defined variance depending on the summed distributions.

The periodogram dual, η, is related to wavenumber space through [37]

$$k_z \approx \frac{2\pi H_0 f_{21} E(z)}{c(1+z)^2}\eta, \tag{3.16}$$

where c is the speed of light, z is redshift, H_0 is the Hubble constant in the present epoch, f_{21} is the frequency of the 21 cm emission line, $E(z)$ describes how H_0 evolves as a function of redshift ($E(z) = \sqrt{\Omega_M(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda}$ [80]), and k_z is the wavenumber along redshift.

3.3.4 Various power spectrum products

We now have the maximum likelihood estimate for the mean, the noise, simulations of the mean, simulations of the noise, and the error bars as a function of $\{u, v, \eta\}$, or through simple conversion, as a function of $\{k_x, k_y, k_z\}$. These estimates are for the power, since we squared each pixel during the Lomb-Scargle transformation. Specifically, this is the power of the combination of the even and odd data sets. We actually prefer the cross power, or the power generated from the even *multiplied* by the odd data set, which will have even less correlated thermal noise. This can be constructed by subtracting the power of the even–odd difference from the power of the even–odd summation, and dividing by 4. Since we have been carrying around the even–odd difference, or the noise, throughout all of our analysis, it is simple to construct the cross power.

We would like to perform averages over these cubes to generate the best possible limits and to generate diagnostics. To do so, we must assume that the EoR is spatially homogeneous and isotropic [58]. There is no spatial preference for the EoR signal in any particular direction, and all of the EoR signal within a small enough redshift measurement is from the same statistical ensemble. Averaging in spherical shells in $\{k_x, k_y, k_z\}$ space or averaging in cylindrical shells in $\{k_x, k_y\}$ space is valid and will average down the noise.

Much like the creation of even–odd sum and difference cubes, we perform the weighted average of pixels given their error bars. However, the value of the pixels are not complex Gaussian-distributed anymore since we took the square, so we are not directly performing a maximum likelihood estimate. However, we perform averages with large sample sets. The central limit theorem states that averaging independent values that are distributed with their own well-defined set of parameters will tend towards a Gaussian distribution as the sample

set becomes large [16]. Therefore, we can continue to use the maximum likelihood estimate for the mean in Equation 3.12 to average together many pixels and the maximum likelihood estimate for the error in Equation 3.13 to estimate the error bars. The only difference is that the variances σ^2 are no longer Gaussian, but rather Erlang variances which can be propagated from the original variances.

We can now create various power spectrum products in 2D and 1D with our 3D power spectrum cube. Error bars, measured noise contributions, and expectation values that we have generated in εppsilon will also be used in our diagnostics.

3.4 Power spectrum diagnostics

We create a variety of power spectrum plots with various averaging regions to make diagnostics along the way to publishing a limit. These are essential to understanding contributions to the power spectrum, which will be vital to following chapters.

3.4.1 2D

The most used diagnostic we have is the 2D power spectrum. We average our $\{k_x, k_y, k_z\}$ power measurements along only the angular wavenumbers $\{k_x, k_y\}$ in cylindrical shells. This creates the power spectrum as a function of modes perpendicular to the line-of-sight ($k_\perp$) and modes parallel to the line-of-sight ($k_{||}$) shown in Figure 3.4. Axes are displayed in units of Hubble constant (h) times inverse megaparsec (Mpc^{-1}). The $\{k_{||}, k_\perp\}$ axes have been converted into $\{\tau, \lambda\}$ on the right and top axes—delay in nanoseconds and baseline length in wavelengths, respectively.

Wavenumber space is crucial for statistical measurements due to the spectral characteristics of the foregrounds. Diffuse synchrotron emission and bright radio sources, while distributed across the sky, vary smoothly in frequency (e.g. [51, 81]). Only small $k_{||}$ values are theoretically contaminated by bright, spectrally smooth astrophysical foregrounds. Since the foreground power is restricted to only a few low $k_{||}$ modes, larger $k_{||}$ values tend to be free of these **intrinsic foregrounds** in wavenumber space.

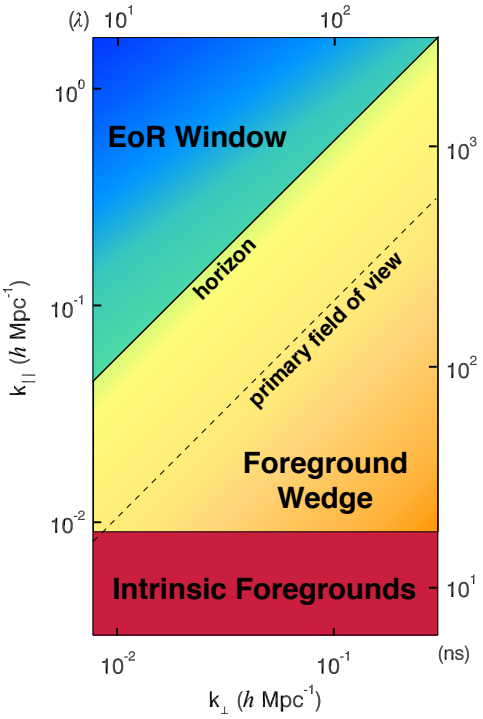


Figure 3.4: A schematic representation of a 2D power spectrum. Intrinsic foregrounds dominate low $k_{||}$ (modes along the line-of-sight) for all $k_{\perp}$ (modes perpendicular to the line-of-sight) due to their relatively smooth spectral structure. Chromaticity of the instrument mixes foreground modes up into the foreground wedge. The primary-field-of-view line and the horizon line are contamination limits dependent on how far off-axis sources are on the sky. Foreground-free measurement modes are expected to be in the EoR window.

However, interferometers are naturally chromatic. This chromaticity distributes foreground power into a distinctive **foreground wedge** due to the mode-mixing of power from small $k_{||}$ values into larger $k_{||}$ values as demonstrated in Figure 3.4 [18, 57, 113, 75, 112, 35, 104, 87, 46]. The primary field-of-view line and the horizon line are the expected contamination limits caused by measured sources in the primary field of view and the sidelobes, respectively. The remaining region, called the **EoR window**, is expected to be contaminant-free. Because the power of the EoR signal decreases with increasing $|k|$, the most sensitive

measurements are expected to be in the lower, left-hand corner of the EoR window.

The intrinsic foregrounds, the foreground wedge, the primary field-of-view line, the horizon line, and the resulting EoR window all have characteristic shapes in 2D power spectrum space. Thus, it is a very useful diagnostic space for identifying contamination in real data. We generate a 2D power spectrum as a function of $k_\perp$ and $k_{||}$ for each of the calibrated data, model, and residual data sets in instrumental XX and YY. Shown in Figure 3.5 is an example 2D power spectrum panel of one night's worth of integrated data from August 23, 2013.

We use this 2D power spectrum panel to help determine if expected contamination occurred in expected regions, and if there is any deviation from the norm. There are some additional shapes and features which are expected, but were not shown in Figure 3.4. First, there is contamination at high $k_\perp$ due to a lack of complete uv-coverage by the instrument. The MWA measures all small modes less than $100\,\lambda$ very well, but becomes sparse for longer wavelength modes. Second, there is harmonic $k_{||}$ contamination in the EoR window which is constant in $k_\perp$. This is caused from regular flagging of aliased frequency channels due to the polyphase filter banks, which creates harmonics in k-space. More information about these instrumental effects are detailed in §2.2 and §3.1.5.

In order to provide reliable results, we also report the 2D noise power spectra. We regularly generate expected noise, observed noise, error, and the noise ratio 2D power spectra, shown in Figure 3.6. The expected noise and the error are related; for a Gaussian distribution, these two quantities are the same. However, these quantities diverge when we square and sum our measurements to create power spectra. We can propagate variances to calculate how these two power spectra depend on the original Gaussian variances,

$$\mathrm{Var}[N] = \frac{1}{\sum_{i=0}^{n} \frac{1}{4\sigma_i^4}} \qquad \mathrm{E}[N] = \frac{\sum_{i=0}^{n} \frac{1}{2\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{4\sigma_i^4}}, \tag{3.17}$$

where $\mathrm{Var}[N]$ is the variance on the noise, $\mathrm{E}[N]$ is the expected noise, n is the number of pixels in the average, and σ^2 is the original Gaussian variance. We also calculate the

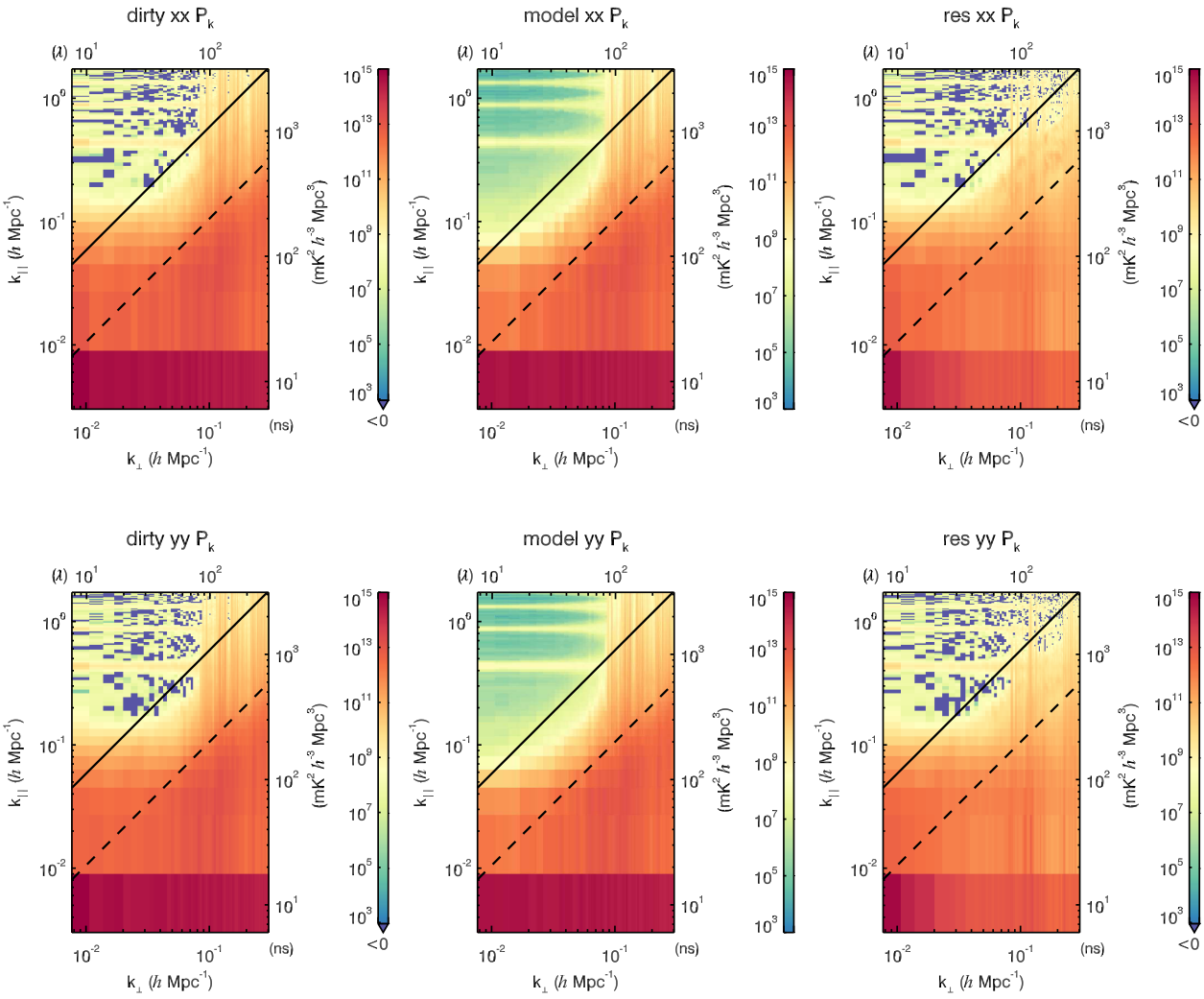


Figure 3.5: The 2D power spectra for the calibrated data, model, and residual (each for instrumental XX and YY) of an integration of 64 observations from August 23, 2013. The characteristic locations of contamination are very similar to Figure 3.4, with the addition of contamination at high $k_\perp$ due to the lack of uv-coverage at longer baselines and $k_{||}$ harmonics due to flagging aliased frequency channels. Regions consistent with noise are dark purple-blue.

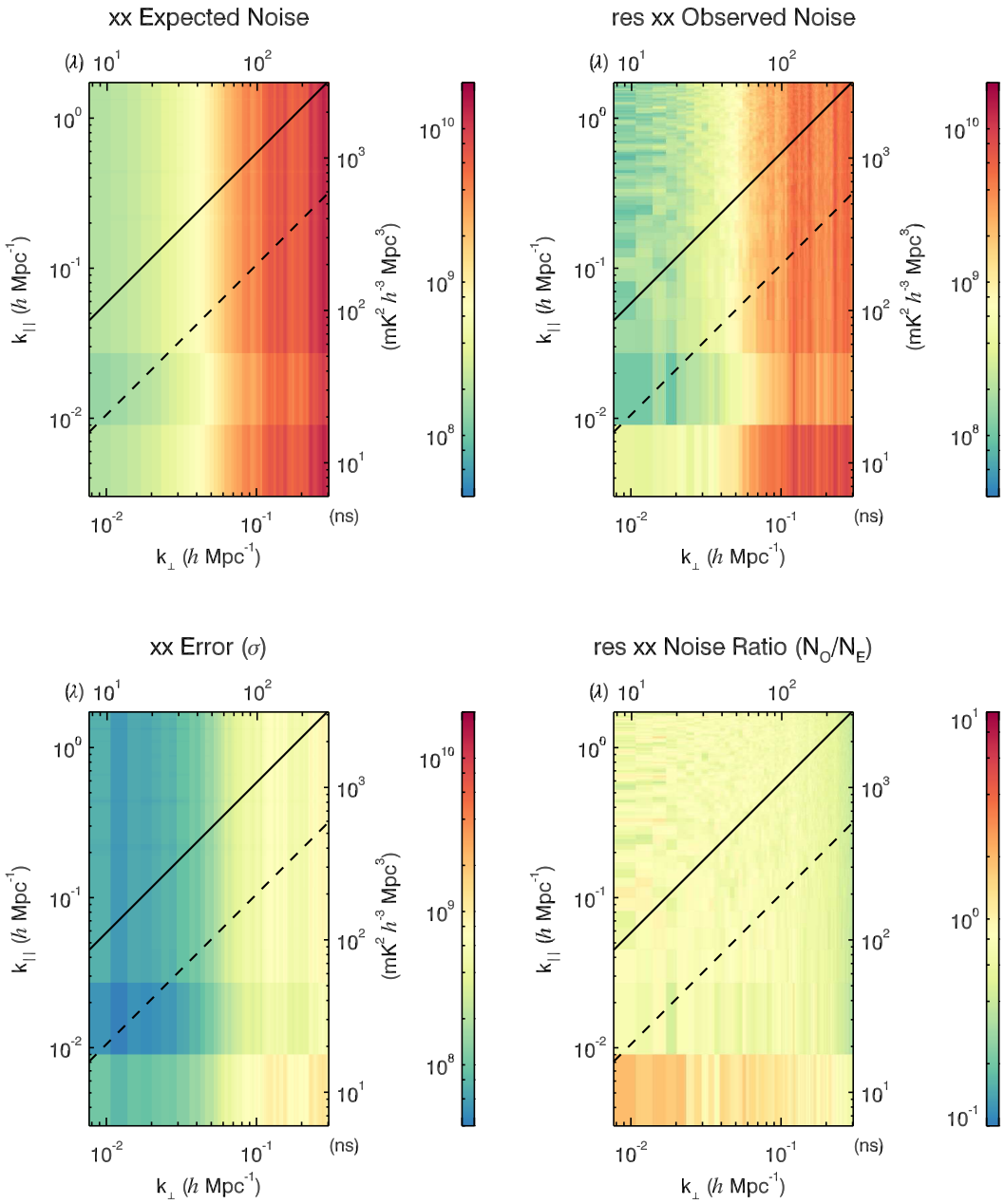


Figure 3.6: The 2D power spectra for expected noise, observed noise, error bars, and noise ratio of an integration of 64 observations from August 23, 2013 in instrumental XX. The expected noise is the calculated noise propagation, and the observed noise is the actual observed noise. Their ratio is near 1, indicating proper treatment. The error bars are related to the observed noise via Equation 3.17.

observed noise, which is the propagation of the maximum likelihood noise in Equation 3.12 throughout the power spectrum analysis. To ensure that our estimations are correct, we compare the observed noise to the expected noise in a ratio plot. There will be fluctuations given different noise realizations, but the ratio should fluctuate around one.

3.4.2 1D

Averaging to 1D harnesses as much as possible of the data, making the most robust limits. However, the 1D power spectrum space also has the ability to be a secondary diagnostic after the 2D power spectrum. While characteristic locations of contamination are easier to distinguish on 2D plots, the 1D diagnostics are more able to distinguish subtleties.

The most basic 1D secondary diagnostic is a 1D power spectrum, averaged in spherical shells, of the entire 3D power spectrum cube. This includes all areas of contamination explored in §3.4.1, which will block out nearly all nuanced signal and contamination areas. Therefore, a typical secondary diagnostic is a 1D power spectrum generated only from pixels which fall within 10 and 50 wavelengths in $k_\perp$, shown in Figure 3.7. This will include intrinsic foregrounds and the foreground wedge, but will avoid contamination from lack of coverage.

A few characteristic contamination shapes are present in this secondary diagnostic plot, much like those in the 2D power spectrum. Intrinsic foregrounds contaminate the lowest k bin, and the foreground wedge contaminates up through about $0.2\,h\,\mathrm{Mpc}^{-1}$. The contamination of the foreground wedge decreases in k because fewer and fewer pixels include the foreground wedge in the spherical averaging. In a true limit, the entire foreground wedge is completely excluded from the averaging. There is also a series of harmonic contaminates at about $0.43\,h\,\mathrm{Mpc}^{-1}$, $0.83\,h\,\mathrm{Mpc}^{-1}$, and $1.3\,h\,\mathrm{Mpc}^{-1}$ due to the aliased channels that were flagged, which can also easily be seen in the 2D power spectra in Figure 3.5.

All those characteristics are expected given the 2D power spectra. However, there is now an obvious contamination seen in Figure 3.7 that was relatively hidden before. An extra contamination feature occurs between the first and second coarse band harmonics at about $0.7\,h\,\mathrm{Mpc}^{-1}$. This is contamination from a cable reflection in the tiles with 150 m

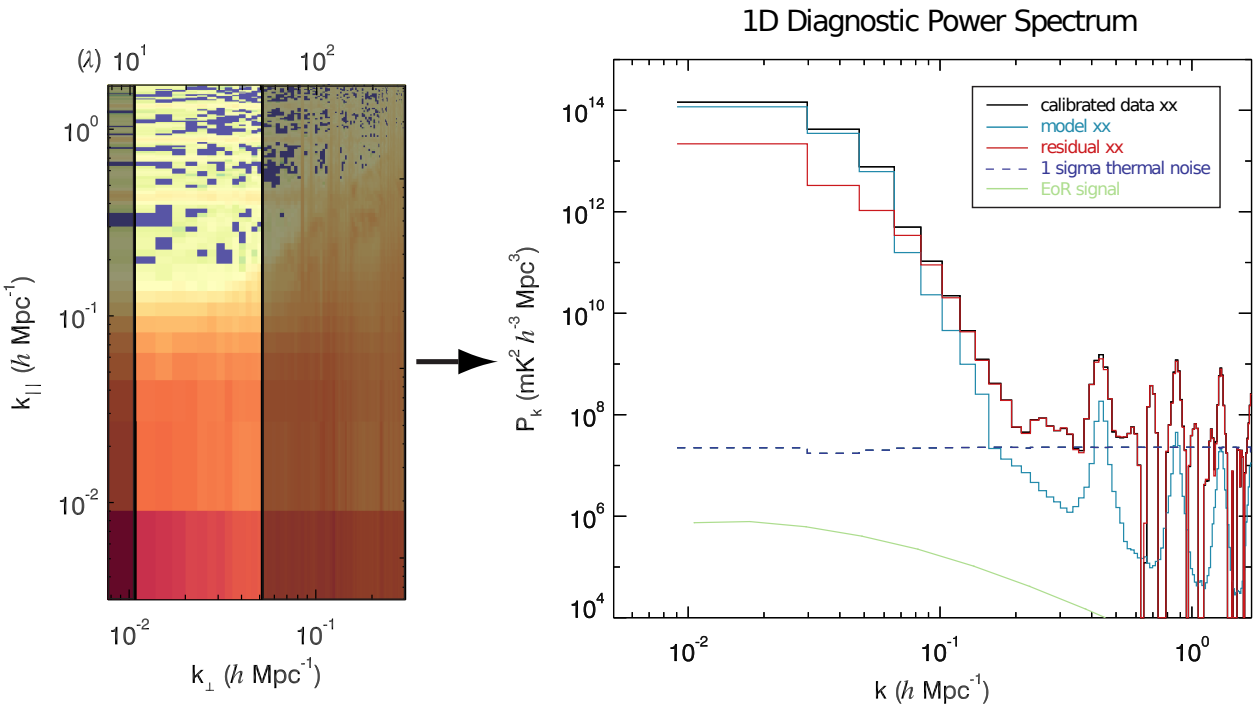


Figure 3.7: 1D power spectra as a function of k for calibrated data, model, residual, theoretical EoR, and the accompanying thermal noise contribution of an integration of 64 observations from August 23, 2013 for instrumental XX. The 2D power spectrum highlights the bins that went into the 1D averaging, which we can modify to exclude the foreground wedge when making limits. A cable-reflection contamination feature at $0.7\,h\,\mathrm{Mpc}^{-1}$ is more obvious in this 1D power spectrum, which highlights the importance of using 1D space as a secondary diagnostic.

cables between the beamformer and low-noise amplifier. The reflection coefficients of a cable reflection are a function of frequency, and thus constructive and destructive interference modulates as a function of frequency. Much like with the aliased channel flagging, spectrally repetitive signals will appear as a bright contamination along $k_\perp$, translating to a near-constant 1D contribution if high in $k_{||}$.

This highlights the potential of utilizing multiple types of power spectra. The 2D power spectrum is a powerful diagnostic due to characteristic locations of contamination. However, it is useful to also use 1D diagnostic power spectra to more quantitatively measure contamination, which helps in discerning smaller contributions.

The cable reflections and flagged aliased channels lead to bright contamination due to their modulation as a function of frequency. This is an important revelation; we must minimize all forms of spectrally repetitive signals during power spectrum analysis. If a spectral mode is introduced in the instrument or in the processing, that mode cannot be used to detect the EoR. This topic will be revisited many times in the following chapters, with diagnostics and confirmations occurring in the 2D and 1D power spectra.

3.4.3 2D difference

Often, we would like to compare a new data analysis technique to a standard to assess potential improvements. We can do this with a side-by-side comparison of 2D power spectra, but are limited by the large dynamic range of the color bar and our eyes' ability to distinguish color. Alternatively, we can compare the 1D power spectra side by side, but lose information about characteristic locations of contamination. We mitigate these issues by creating 2D difference power spectra.

We take a bin-by-bin difference between two power spectra to generate a 2D difference power spectrum. The reference or standard is subtracted *from* the new run, and the 2D difference power spectrum varies positive and negative depending on power levels. We choose a red-blue color bar to indicate sign, where red indicates an increase in power relative to the reference and blue indicates a decrease in power.

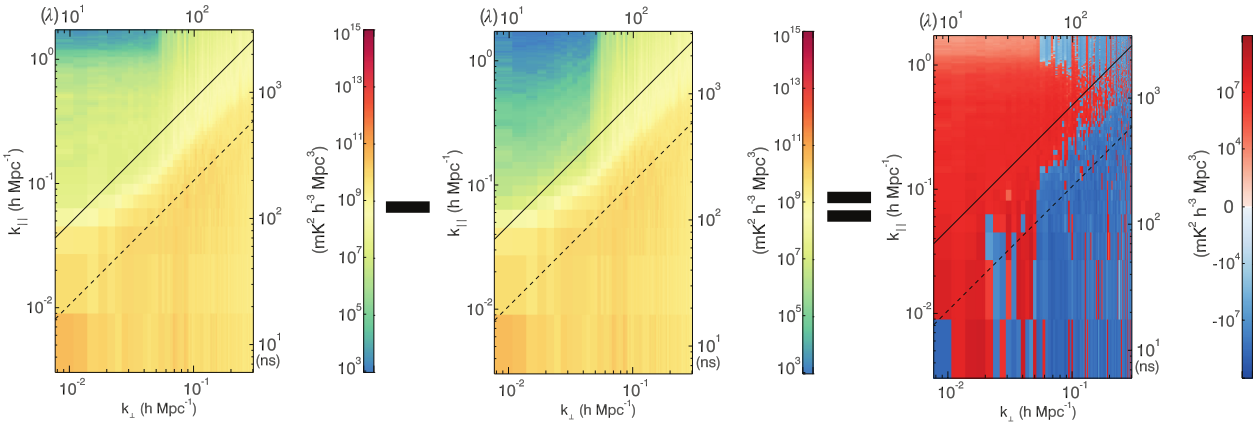


Figure 3.8: The subtraction of a residual 2D power spectrum (left) and a reference residual 2D power spectrum (middle) to create a difference 2D power spectrum (right). Red indicates a relative excess of power, and blue indicates a relative depression of power.

Figure 3.8 shows an example of a 2D difference power spectra. The new data analysis run is the left panel, the reference is the middle panel, and the 2D difference is the right panel. For this example, we have chosen a new data analysis run that had excess power in the window, and a decrease in power in the foreground wedge, compared to the reference. In general, we make 2D difference power spectra for dirty, model, and residual, for both XX and YY polarizations to match the six-panel plot in Figure 3.5.

Whenever we make changes to the pipeline, we make 2D difference power spectra compared to a reference. This ensures that all changes we implement are actually beneficial. This will be heavily used in much of the following chapters to prove validity of our changes to the pipeline.

3.5 Overview

We have built an analysis framework that takes large quantities of measured visibilities and generates images, power spectra, and other diagnostics. Four modularized components exist: pre-analysis flagging and averaging, calibration and imaging, integration, and error-

propagated power spectrum calculations. The accuracy of each component is crucial due to the level of precision needed in an EoR experiment. In particular, the analysis handled by FHD and εppsilon is complicated and multifaceted, necessitating constant refinement and development to ensure accuracy, precision, and reproducibility.

The previous subsections have covered the computational tasks of FHD and εppsilon in detail, which can be summarized in the following list.

FHD: Fast Holographic Deconvolution

1. Calculate the image space beam between tile pairs and transform into $\{u, v, f\}$ space.
2. Generate the uv-plane of all known and reliable sources using a direct Fourier transform from a sky catalog's floating-point locations.
3. Sample the model uv-plane with the uv-beam to generate model visibilities, or estimates of the measurement.
4. Calibrate by performing an iterative least-squares analysis between raw and model visibilities, accounting for temperature and bandpass effects in the amplitude and ramps in the phase.
5. Grid calibrated visibilities and model visibilities onto the uv-plane using the beam sensitivity as the gridding kernel.
6. Grid visibilities of value 1 with the beam to make a sampling map, and with the square of the beam to make a variance map.
7. Weight the gridded uv-data with the sampling map and Fourier transform the spatial dimensions to generate diagnostic images in slant orthographic projection.
8. Fourier transform the spatial dimensions of the uv-sampling map, variance map, and unweighted data products to slant orthographic projection and interpolate to HEALPix for even–odd time samples and instrumental polarizations.

εppsilon: Error Propagated Power Spectrum with Interleaved Observed Noise

1. Input the interpolated HEALPix data products from FHD, or integrate several HEALPix data products from multiple observations.
2. Cut the image plane to an appropriate extent that is fully covered by all observations, and direct Fourier transform the spatial dimensions to uv-space.
3. Weight the data products by the sampling map and weight the variance map by the square of the sampling map.
4. Perform weighted sums and differences between even–odd interleaved cubes, using calculated error bars as the weights to generate maximum likelihood estimates.
5. Apply a window function and use a Lomb-Scargle periodogram to transform frequency to k-space.
6. Calculate cross power from even–odd sums and differences.
7. Calculate weighted averages to generate 1D power spectra, 2D power spectra, and their respective error bars.

Many data products are produced with our analysis pipeline. We make instrumental polarization images for each two-minute observation for calibrated data, model data, and residual data. The resulting 2D and 1D power spectra from the integrated cubes are made for the same set with various k-space cuts. Noise power spectra, both simulated and propagated, are generated along with the power spectrum error bars. All of these diagnostic images and plots are essential to creating our final figure, the EoR limit in 1D power spectrum space.

FHD and εppsilon form the core of our analysis, and have been used to make EoR power spectrum limits in the past [2]. In the following chapters, we will introduce new analysis techniques which utilize this framework. However, lessons learned from improving our data analysis pipeline are not exclusive; the underlying importance of the calibration precision is vital to power spectrum analyses in general. Nevertheless, FHD and εppsilon were essential to their discovery, both due to their flexibility and efficiency.

Chapter 4

BANDPASS CALIBRATION

The bandpass is a significant source of contamination in our measurements of the EoR through power spectra, and has the potential to completely obscure any detection [1]. However, it is necessary; there are many instrumental effects which must be captured in calibration that are not easily describable without many degrees of freedom. This chapter will focus on experimentally determined best methods for generating a bandpass calibration solution, and the variety of instrumental parameters we would like to represent.

The bandpass calibration solution is the amplitude of the instrumental response per fine frequency—the result of the linear least-squares solution between the data and the model visibilities as a function of frequency described in §3.1.5. It captures all fine detail and amplitude changes, whether inherent to the instrument or not. For example, any thermal noise that is captured in the linear least-squares method will propagate into the bandpass calibration solution. This is applied to the data, and the frequency modulations present in the bandpass show up in the power spectra. Therefore, reducing noise contributions and false features due to our linear least-squares assumptions is crucial.

In this chapter, the subtleties and consequences of creating bandpasses to reduce potential contamination effects will be explored. Motivated by power spectrum differences, we have developed various averaging techniques in the search for low-amplitude instrumental effects. All of these methods are experimentally determined—theoretical reasoning is described in §6.

4.1 Typical features

We must accurately portray instrumental effects in the bandpass, and there are several. The main contributors can be seen by calculating our historical calibration solution, the global bandpass. We average together all tiles' separate bandpass solutions to reduce noise, generating $|\mathbf{g}_P(f;\boldsymbol{\alpha})|$ in Equation 3.7, which we illustrate in Figure 4.1 for one of the polarizations. The major instrumental characteristics of our bandpass calculation are described in the following list.

Coarse bandpass shape: The first stage PFB introduces a bandpass shape per coarse band, highlighted in green in Figure 4.1. In this global bandpass, we have already divided out a theoretically calculated coarse band shape—any residual shape deviates from expected values. This shape appears to change over the full band; the first coarse bandpass shape is very different from the last coarse bandpass shape.

Large-scale ripple and attenuation: There is an overall shape to the bandpass, shown in purple in Figure 4.1. It decreases over the full band, and has a large-scale ripple feature. If this feature were a cable reflection, it would correspond to a cable roughly 7 m in length. For this reason, we suspect this feature originates from the LNA-to-beamformer cable, which is about 7 m long. There is also an overall attenuation over the band.

Digital gain jump: At higher frequencies, the gain decreases enough for there to be under-saturation. Our bits are selected such that they are most accurate within a certain amplitude range. We maintain this range by applying **digital gain jumps**—multiplying the data by integer values at certain frequency ranges, and then correcting in post-processing once the data has past any potential bit selection effects [88]. A digital gain jump effect can be seen in orange in Figure 4.1.

Regular flagging: There is a systematic aliasing that occurs every 1.28 MHz due to the first

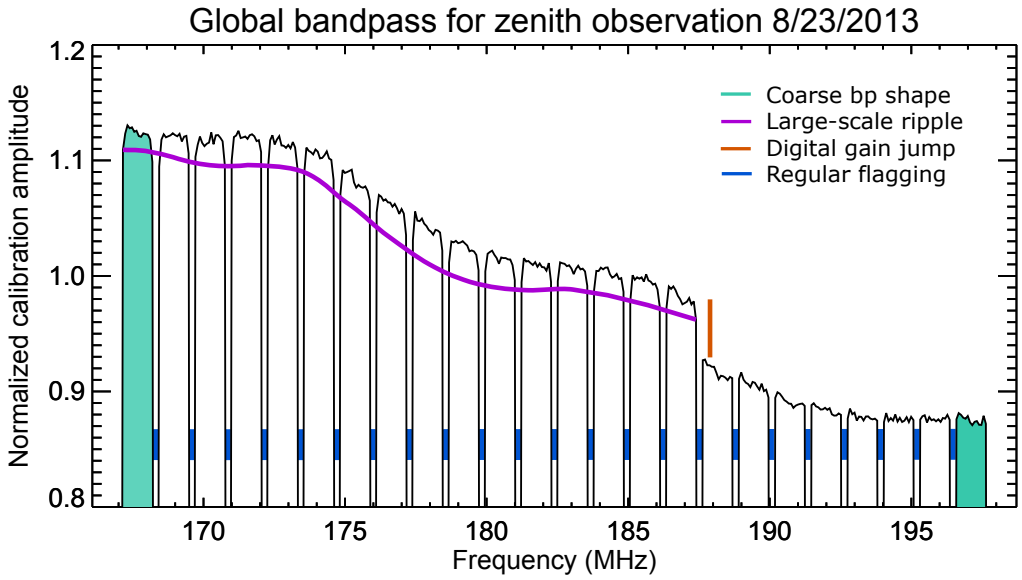


Figure 4.1: Schematic of instrumental features in the bandpass. Green highlights the change in polyphase filter bank shape over the bandwidth. Purple emphasizes an overall decrease in amplitude and a large-scale ripple over the bandwidth. The digital gain jump effect, a prominent feature in 2013 data, is highlighted in orange. Blue emphasizes the regular flagging to avoid polyphase filter bank aliasing.

stage in the PFBs (see §2.2.3 for more details). We flag the fine frequency channels that are most affected in order to avoid this aliasing in the final power spectrum. However, this regular flagging, highlighted in blue in Figure 4.1, will propagate to the final power spectrum.

Cable reflections: Not shown in Figure 4.1 are the ripples caused by cable reflections from the beamformer to the receiver. This figure is the average over all the bandpass calibration solutions from all the tiles, which effectively erases any cable reflection features. The reflection's amplitude and phase are dependent on the cable material and length, as well as environmental factors like temperature, and thus averaging solutions can artificially remove them.

Cable Length	Cable Type	Tile Amount
90 m	RG-6	19
150 m	RG-6	31
230 m	RG-6	23
320 m	LMR-400-75	8
400 m	LMR-400-75	17
524 m	LMR-400-75	30

Table 4.1: Summary of the cable groups, dependent on length and type.

Some of the typical features are characterized well with a global average, while others are not. By taking the average over all tiles to calculate the bandpass, we have artificially removed any contributions that happen on a per-tile or per-cable level. These instrumental dependencies must be accounted for in the calibration, or they will contribute directly to power spectrum contamination. We have experimentally explored ways to include these effects and reduce contamination, detailed in the following sections.

4.2 Cable reflections

Back in 2014, the very first hints of cable reflection contamination were seen in the power spectra [24, 5]. It is a faint and variable effect; cable reflections can vary between being nonexistent and being about 15% of the gain amplitude depending on cable type and tile, even though the cables meet engineering specifications. The various types of cable materials and lengths allow for various modes of cable reflections, and are quantized into five different groups. The groups are summarized in §2.2.1 and Table 4.1.

One way to see the cable reflections easily in the calibration solutions is to look at the auto-correlation solutions. Auto-correlations are the visibilities of a tile correlated with itself, literally $\mathbf{v}_{AA,PP}([f,t])$ in our framework (see §3.1.5). The baseline is of zero length, and thus should not measure any structure from the sky. Therefore, any structure seen in the auto-correlation was a contribution by the instrument. The auto-correlations are better able to

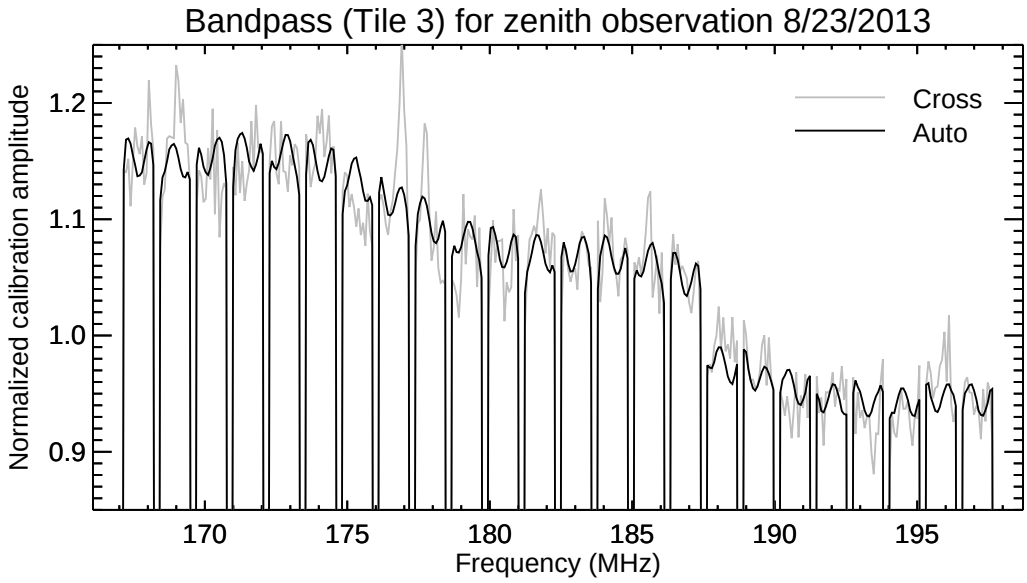


Figure 4.2: The auto-correlation bandpass (black) for a specific tile and observation compared to the cross-correlation bandpass solution (grey). It is much easier to see the 150 m cable reflection ripple in the auto-correlation.

pick up amplitude instrumental systematics, apparent in Figure 4.2[1].

The cable reflection creates an amplitude modulation and a phase modulation as a function of frequency, due to a mismatched impedance between cables (contribution **K** in the signal path framework). Auto-correlations have been used in the past to model the particular 150 m cable reflection, but correlated noise prevented any significant improvement. The standard method, then, is to implement a hyperresolved cable-reflection fit on the cross-correlations.

Averaging the gains across antennas and/or time effectively removes cable reflections from the solutions, thus we must specifically incorporate them. We only fit the cable reflection for the l_{150}[2] antennas because of its prime location within the EoR window in power spectrum

[1]See §4.5 for more information on using the auto-correlations in the bandpass.

[2]We can group antennas into sets based on their beamformer-to-receiver cable lengths: $L = \{l_{90}, l_{150}, l_{230}, l_{320}, l_{400}, l_{524}\}$.

space. Some cable lengths in the window are difficult to model due to other contamination in the mode (l_{90} is degenerate with the PFB width) or due to low signal-to-noise for the reflection amplitude (l_{320}).

We find the theoretical location of the potential mode using the nominal cable length and the specified light travel time of the cable. We then perform a hyperfine DFT of the gain *for each observation* around the theoretical mode

$$\mathbf{g}_{A,P}(\tau_\kappa) = \sum_{n=0}^{N-1} \left(\frac{\mathbf{g}_{A,P}(f_n)}{|\mathbf{g}_P(f;\boldsymbol{\alpha})|} - (c_2 f_n^2 + c_1 f_n + c_0)\, e^{i(\phi_1 f_n + \phi_0)} \right) e^{-2\pi i \kappa \frac{n}{N}}, \tag{4.1}$$

where τ_κ is the delay, $[n, N] \in \mathbb{Z}$, and κ is the hyperfine index component. Typically, we set the range of κ to be $[k_{\tau_o} - \frac{1}{20}k, k_{\tau_o} + \frac{1}{20}k]$, where k_{τ_o} is the index of the theoretical mode and k is the normal DFT index in the range of $[0, N-1]$ [2]. The cable reflection fit depends on how well other sources of contamination are removed, especially from the mode in question. Therefore, Equation 4.1 is updated with bandpass and polynomial fit calculation improvements whenever they are made.

The maximum $|\mathbf{g}_{A,P}(\tau_\kappa)|$ around k_{τ_o} is chosen as the experimental cable reflection. The associated amplitude c, phase ϕ, and mode τ are then calculated to generate the experimental cable reflection contribution $ce^{-2\pi i \tau f + i\phi}$ to the gain for each observation. We can create a set of these coefficients, $D = \{c, \tau, \phi\}$, for easy reference.

The updated calibration solution for l_{150} tiles is then

$$\mathbf{g}_{A,P}(f;\boldsymbol{\alpha},[C],[D]) = \underbrace{|\mathbf{g}_P(f;\boldsymbol{\alpha})|}_{\textbf{global bandpass}} \Big(\underbrace{(c_2 f^2 + c_1 f + c_0)\, e^{i(\phi_1 f + \phi_0)}}_{\textbf{per antenna gain and phase}} + \overbrace{ce^{-2\pi i \tau f + i\phi}}^{\textbf{per antenna cable reflection}} \Big). \tag{4.2}$$

We experimentally tested this new addition to the calibration solutions using our standard set of data—64 of the least-contaminated observations from the night of August 23, 2013. We can compare these new power spectra with cable-fitting procedures (Equation 4.2) to the power spectra with no cable-fitting procedures (Equation 3.10) using a 2D power spectrum difference, shown in Figure 4.3.

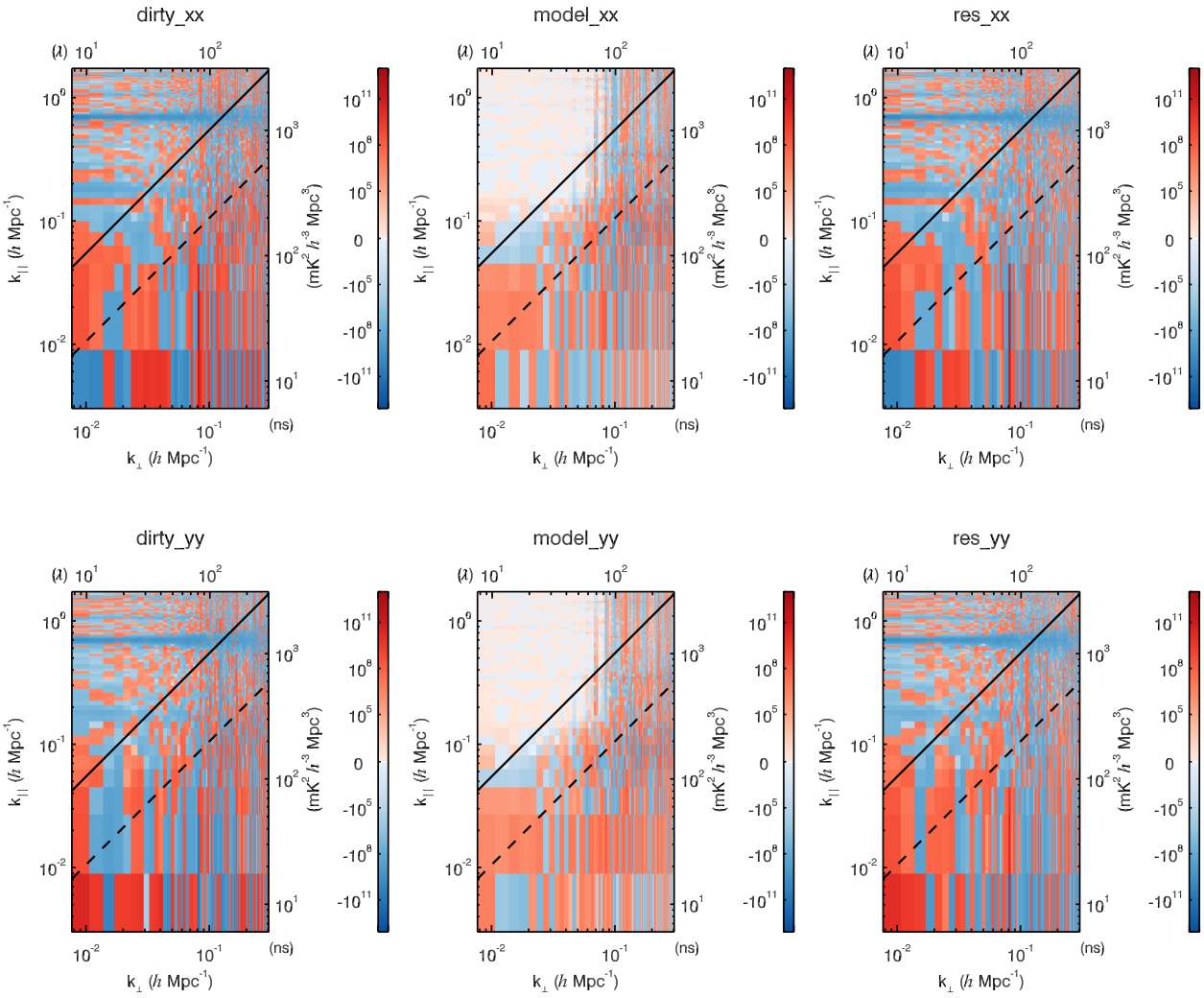


Figure 4.3: 2D power spectrum differences (see §3.4.3) comparing an integration of 64 observations from August 23, 2013, where one test has cable fitting in the calibration for l_{150} tiles and one does not. Blue indicates a decrease in power for the cable-fitting test compared to the reference, and red indicates an increase in power. There is a significant decrease in power at the mode associated with the 150 m cable reflection at constant $k_\perp$.

Since the only difference between the two tests was the calibration, there is a noticeable change in the calibrated data and residual, with only noise fluctuations in the model. The largest difference is a reduction in power in the cable-fitted power spectrum (blue) at constant $k_\perp$ for a delay of about 1235 ns, which is the light travel time of the 150 m cable reflection. We have successfully removed contamination at the mode in question.

4.3 Temperature dependence

The bandpass only describes fine frequency structure; it does not describe overall amplitude fluctuations. We use polynomial fitting to capture this critical component in calibration. As described in Equation 3.8, a low-order polynomial is fit to the full band once fine frequency structure from the bandpass is removed.

When we began to look at data quantities that spanned many months, it became apparent that the overall amplitude was changing with time. This tile amplitude variation is correlated with the temperature at the beamformer, which is very close to ambient temperature. As the season went from Winter to Summer, the amplitude decreased, consistent with amplifiers being more efficient at lower temperatures.

This could lead to estimable quantities, and a predetermined polynomial amplitude in the calibration solutions. However, there are complications. Every LNA-receiver chain has a slightly different dependence on temperature. In addition, there can be shifts in total amplitude, as shown in Figure 4.4. Two separate tiles have different temperature dependencies, with Tile 2 (right panel) experiencing a drastic shift. This particular instance was traced back to an air conditioning service trip on-site on October 10, 2013, which highlights the difficulty in using the temperature dependence in the amplitude estimation. We continue to use polynomial fitting per observation to capture overall amplitude changes with temperature.

The temperature can affect more than just the overall amplitude as well. The digital gain jump and its corresponding bit selection is affected by temperature. This means that the amplitude of the digital gain jump appears to change, usually in a non-linear way due to efficiency of the bit range (see §7.1.2 for details). We have therefore implemented a change

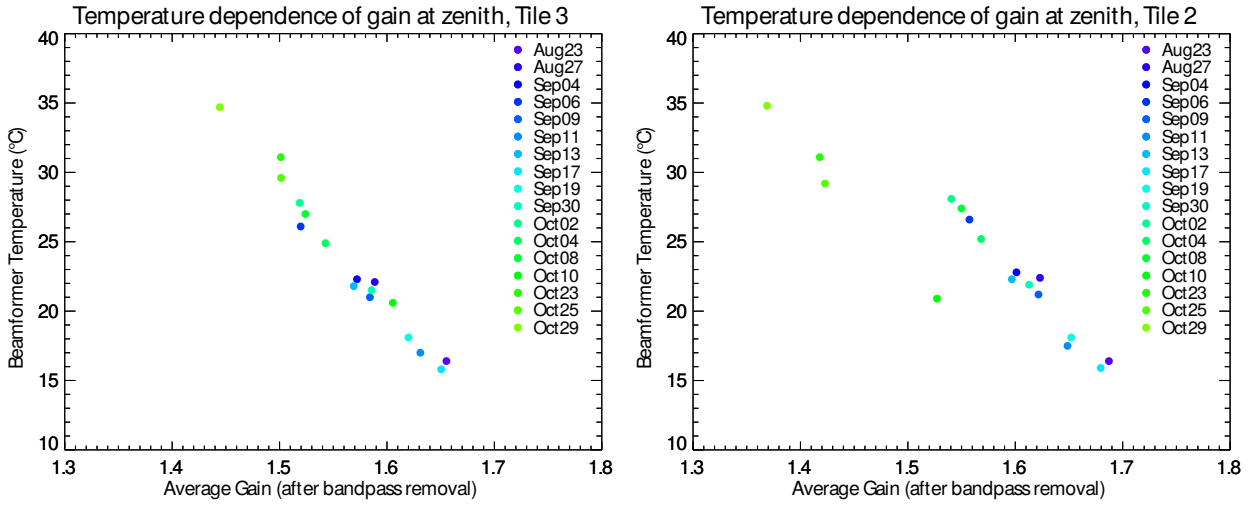


Figure 4.4: The average overall gain at zenith over days in 2013 for Tile 3 (left) and Tile 2 (right) as a function of beamformer temperature. There is a strong negative correlation. Any change to the system (e.g. an air conditioning service trip on October 10 for Tile 2) can dramatically change the parameters.

to the amplitude polynomial fitting ($c_2 f^2 + c_1 f + c_0$, Equation 3.8) based on this non-linear temperature dependence:

$$\{c_2, c_1, c_0\} = \begin{cases} \{c_{2'}, c_{1'}, c_{0'}\}, & \text{if } f \leq f_D \\ \{c_{2''}, c_{1''}, c_{0''}\}, & \text{if } f > f_D \end{cases} \tag{4.3}$$

A separate set of coefficients are calculated for the frequency ranges on either side of f_D, the frequency of the digital gain jump. While this increases the number of degrees of freedom in the calibration solutions, it does attempt to model a true instrumental effect which should be included in the bandpass.

Again, we have experimentally tested this new calibration with the least-contaminated observations of the night of August 23, 2013. We compare the addition of Equation 4.3 to a calibration standard (Equation 4.2); Figure 4.5 shows the 2D power spectrum difference. In the lower portion of the EoR window and throughout most of the foreground wedge, there

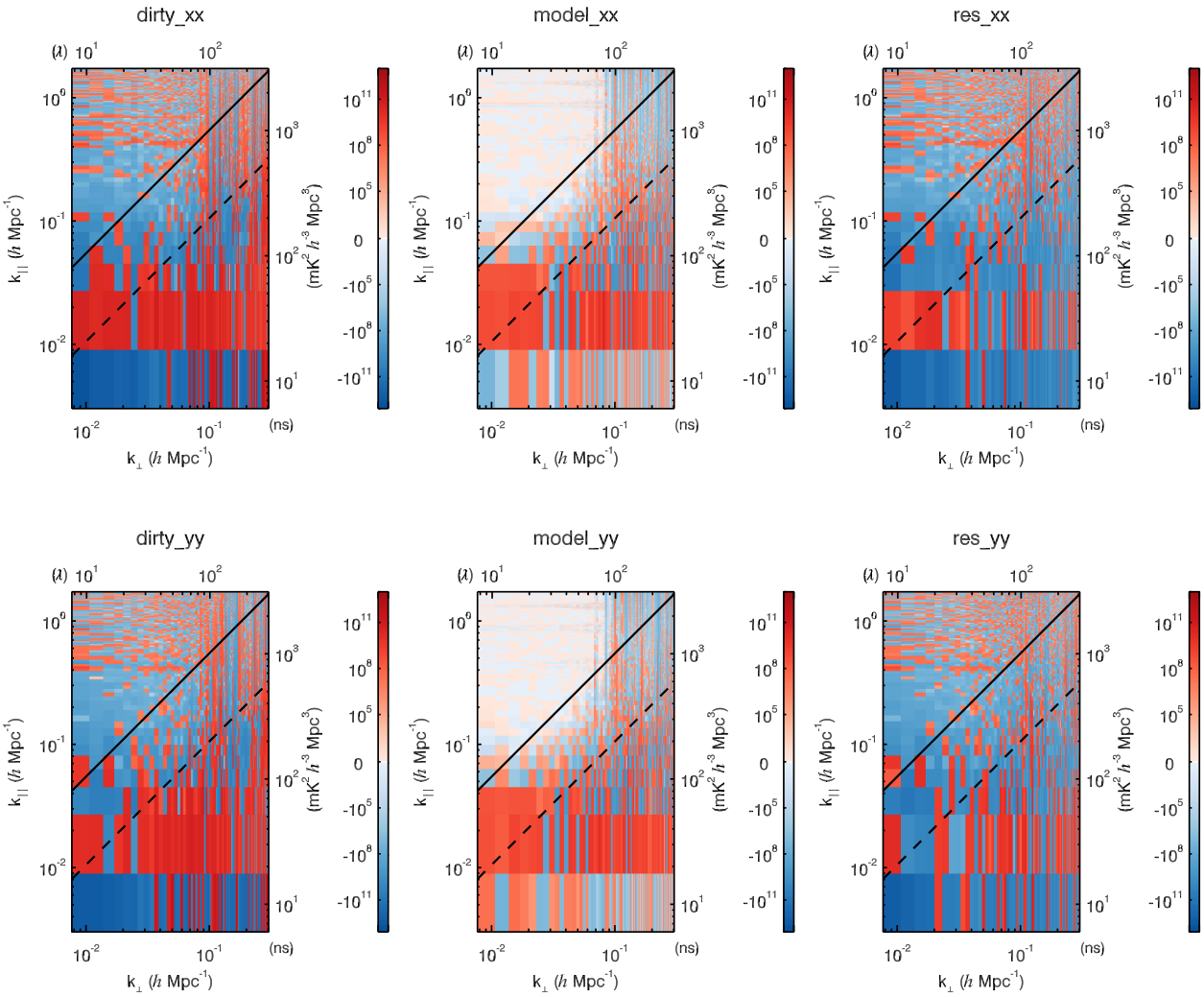


Figure 4.5: 2D power spectrum differences comparing an integration of 64 observations from August 23, 2013, where one test has an independent fit to the digital gain jump in the calibration and one does not. Blue indicates a decrease in power relative to the reference, and red indicates an increase in power. There is a significant decrease in power near the lower portion of the EoR window and throughout most of the foreground wedge, which are the highest signal-to-noise regions.

is a significant reduction of power (blue) for the power spectrum when we fit the digital gain jump (Equation 4.3). A miscalibrated digital gain jump creates a step function, and a Fourier transform over a step function causes leakage on all modes. By accounting for the digital gain jump directly, we have reduced the amount of contamination from this effect, especially in higher signal-to-noise regions of the power spectrum.

4.4 Averaging

When we averaged all of the tiles' responses to generate a global bandpass, we were motivated to reduce noise on the calibration solutions (see §3.1.5). However, we lose something inherent to the bandpass in this average: a cable- or tile-dependent instrumental term. For the MWA, sets of antennas experience the same attenuation as a function of frequency due to cable types, cable lengths, and whitening filters. Individual tiles may also have particular characteristics, either due to environment or hardware variation. Instead of averaging over all tiles and potentially removing these effects, we explore other averaging sets.

4.4.1 By cable

We group antennas into the set $L = \{l_{90}, l_{150}, l_{230}, l_{320}, l_{400}, l_{524}\}$ where the subscript denotes the length of the cable type. Noise and spectral structure from unmodeled sources and their point spread functions will differ from one antenna to another, and thus the average will decrease these unwanted effects, while capturing cable-dependent terms. We use

$$\left|\mathbf{g}_{L,P}(f;\boldsymbol{\alpha}_L)\right| = \mathcal{R}\left\langle \left|\mathbf{g}_{A\in L,P}(f)\right|, 2\sigma \right\rangle, \tag{4.4}$$

where $\mathcal{R}$ is the resistant mean function that calculates the distribution of the amplitudes of a similar antenna set $\boldsymbol{\alpha}_L$ for each frequency and polarization, excludes Gaussian 2σ outliers, and computes the resulting mean. We choose the resistant mean because outlier contributions are more reliably reduced than in median calculations. The variable change $\boldsymbol{\alpha} \rightarrow \boldsymbol{\alpha}_L$ indicates one parameter per cable grouping of antennas.

There is a compromise in generating bandpasses per cable type, as opposed to one global bandpass. While modeling of cable-dependent parameters is correctly being taken into account, there are less contributions to the mean. The resulting bandpass is noisier, highlighted in Figure 4.6. More cable-specific features have been modeled (e.g. slopes due to differing attenuations, a low-order fluctuation in the 400 m cable tiles, various digital gain jump effects that are tied to attenuations, etc.), but the increase in noise is unmistakable.

We investigated this new calibration scheme in the space of the measurement by comparing 2D difference power spectra from the least-contaminated observations of August 23, 2013. We generate new power spectra using Equation 4.4 to create the bandpass, and then compare with multiple other standards. Figure 4.7 shows three different comparisons for the calibrated data in instrumental XX and YY using the cable-averaged calibration.

The first column of Figure 4.7 is referenced against the global bandpass and cable-reflection fit calibration (Equation 4.2). There is power reduction in the EoR window with the cable-averaged calibration. However, this is mostly due to a better handling of the digital gain jump. Tiles of a specific cable group have similar attenuation, and thus the amplitude change of the digital gain jump is similar. This is apparent in the second column, where the reference calibration scheme is comprised of the global bandpass and fits to the cable reflection *and* the independent digital gain jump (Equation 4.2 with Equation 4.3). Most of the beneficial reduction in EoR window power is gone, and the difference is noise-like. Thus, the benefits of the cable-averaged calibration and the independent fit to the digital gain jump are degenerate.

We can investigate even further by comparing only the low part of the frequency band before the digital gain jump. Power spectra can be generated for subsets of the frequency band, allowing us to highlight certain regions of interest. The last column of Figure 4.7 is the same comparison as the middle column, but only using data before the digital gain jump to remove any of its contribution. We can see that modes within the EoR window are contaminated—this is due to the significant increase in noise on the calibration solutions compared to the reference.

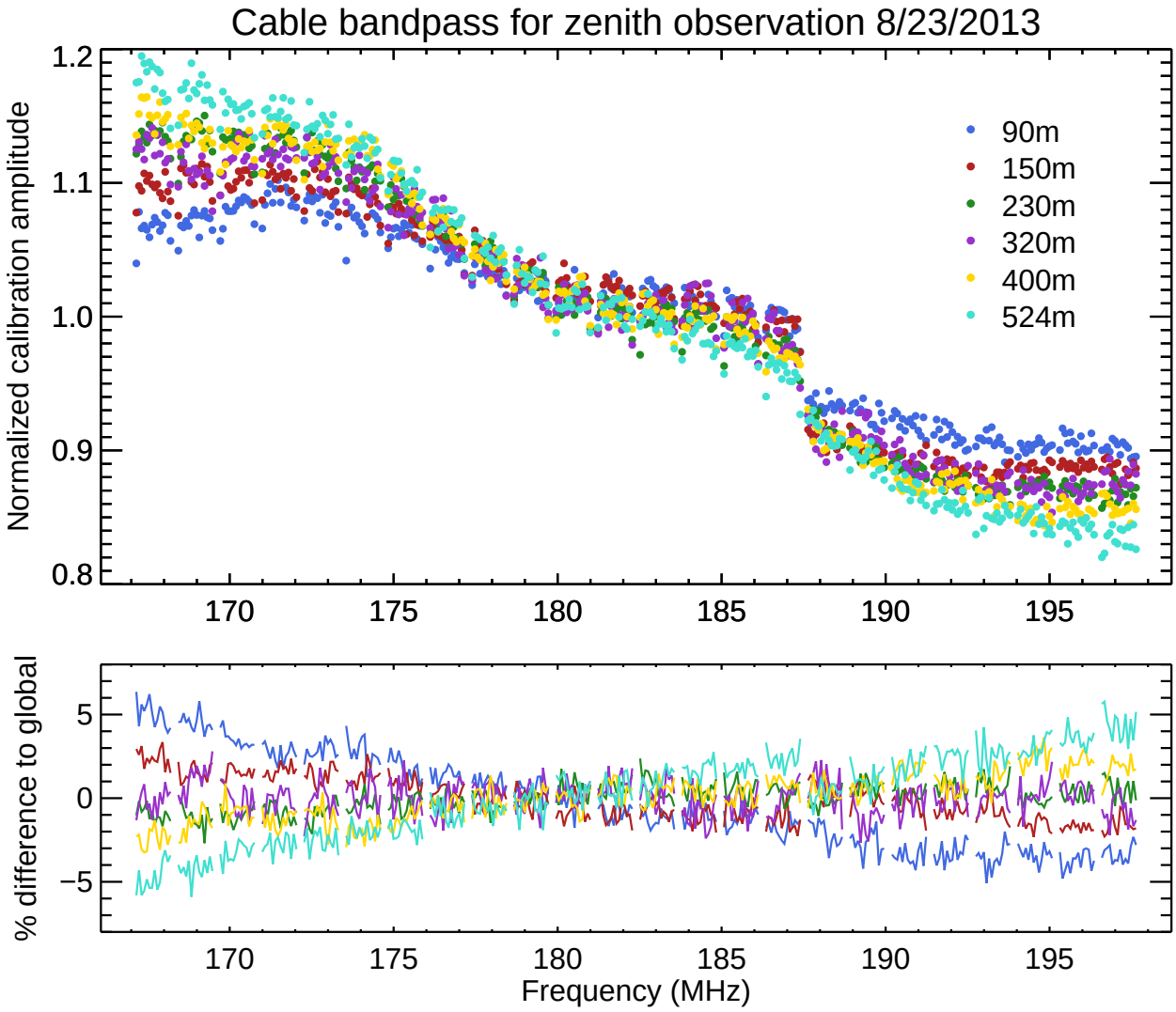


Figure 4.6: The bandpass-per-cable grouping for the zenith observation of August 23, 2013. The top plot is the resulting amplitude for each of the five LNA-to-beamformer cable lengths, and the bottom plot is the percent difference of the cable bandpasses compared to the global bandpass.

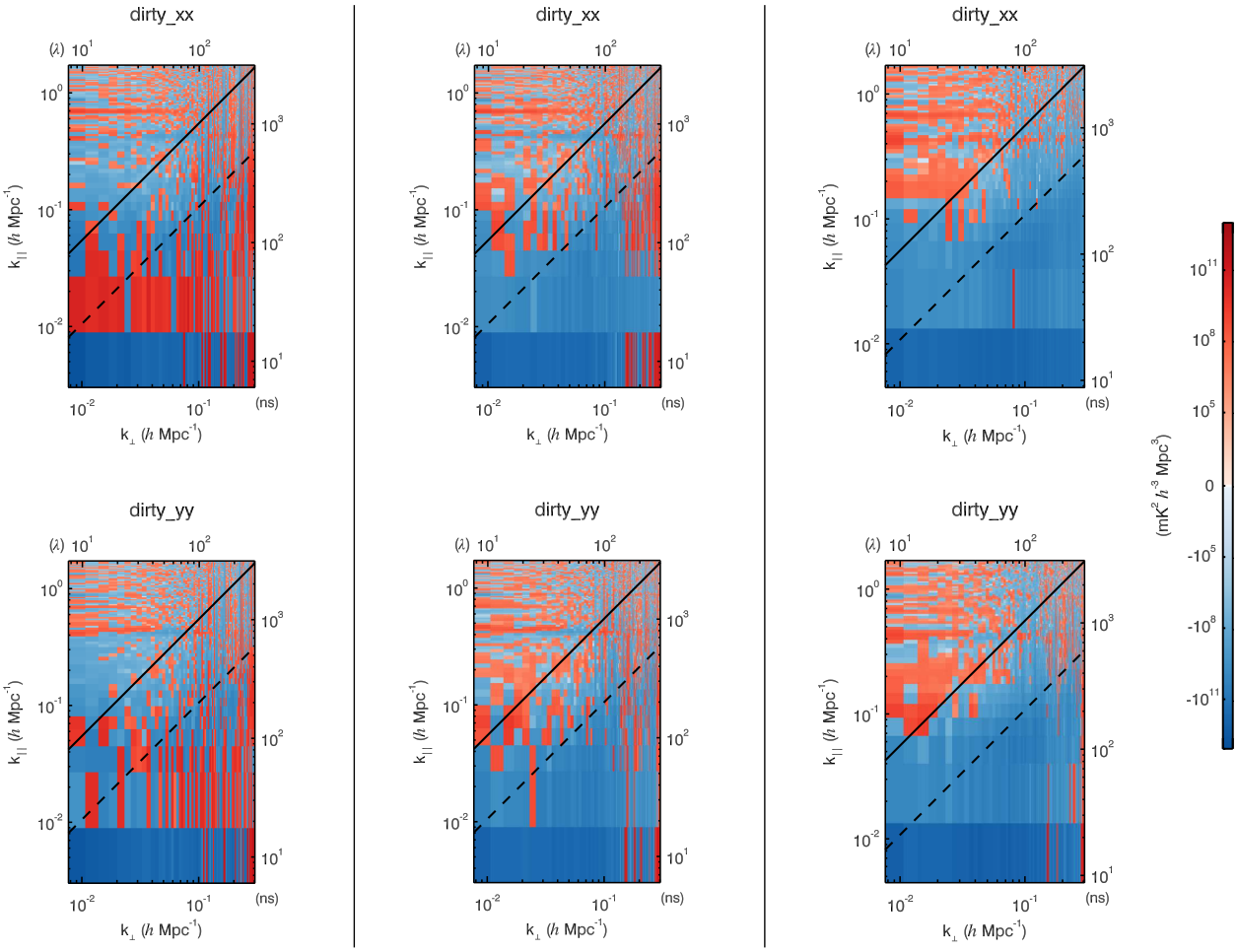


Figure 4.7: 2D power spectrum differences comparing three different references of 64 observations from August 23, 2013, each with a different calibration scheme. The cable-averaged calibration scheme is referenced against: (left column) the global bandpass and fitted cable reflection; (middle column) the global bandpass, fitted cable reflection, and independent digital gain jump; and (right column) the global bandpass, fitted cable reflection, and independent digital gain jump, but only for a subset of the frequency band before the digital gain jump. Blue indicates a decrease in power for the cable-averaged calibration compared to the reference, and red indicates an increase in power. The cable-averaged calibration is degenerate with fitting for an independent digital gain jump, and is noisier than the global bandpass.

4.4.2 By cable and time

If more observations are available, we can follow a similar averaging process over time to reduce the noise. Disregarding temperature-dependent effects, the MWA is very stable in time. A normalized bandpass per antenna should be nearly identical from one moment to the next, excluding potential bit and noise contributions. The thermal noise has a mean of zero, and thus will average away.

Different pointings have different beams and resulting beam errors, and thus make a natural instrumental grouping. We create a time set of $T = \{\rho_{-2}, \rho_{-1}, \rho_0, \rho_{+1}, \rho_{+2}\}$ where times are grouped by how many pointings they are away from zenith. Whenever possible, we use

$$\left|\mathbf{g}_{L,P}(f; \boldsymbol{\alpha}_L, \theta_T)\right| = \mathcal{R}\left\langle \left|\mathbf{g}_{A\in L,P}(f)\right|, 2\sigma, \theta_T \right\rangle, \tag{4.5}$$

where θ_T includes observations within a single pointing and as many days as possible. The variable change $t \to \theta$ indicates one parameter per timing group.

Figure 4.8 shows the result of averaging over many reliable days in **Season 1** for the zenith pointing by cable type. There is a significant decrease in noise, while still capturing cable-dependent effects. Clearer coarse band PFB shapes emerge after the digital gain jump, as seen in the percent-difference panel. Small wavelength ripples are also apparent: some due to specific PFB shapes, some due to cable reflections.

We experimentally tested this calibration scheme using 2D difference power spectra and various references, shown in Figure 4.9. The first column is referenced against a calibration with the global bandpass and a fitted cable reflection (Equation 4.2), where again we see a reduction in power in the EoR window consistent with better handling of the digital gain jump. The middle column is referenced against a calibration with the global bandpass, fitted cable reflection, and independent fit of the digital gain jump (Equation 4.2 with Equation 4.3), and the EoR window is now relatively noise-like. The last column is the same as the middle column, except that it only compares frequencies below the digital gain jump. There is less power contamination in the EoR window compared to Figure 4.7, and it is relatively noise-

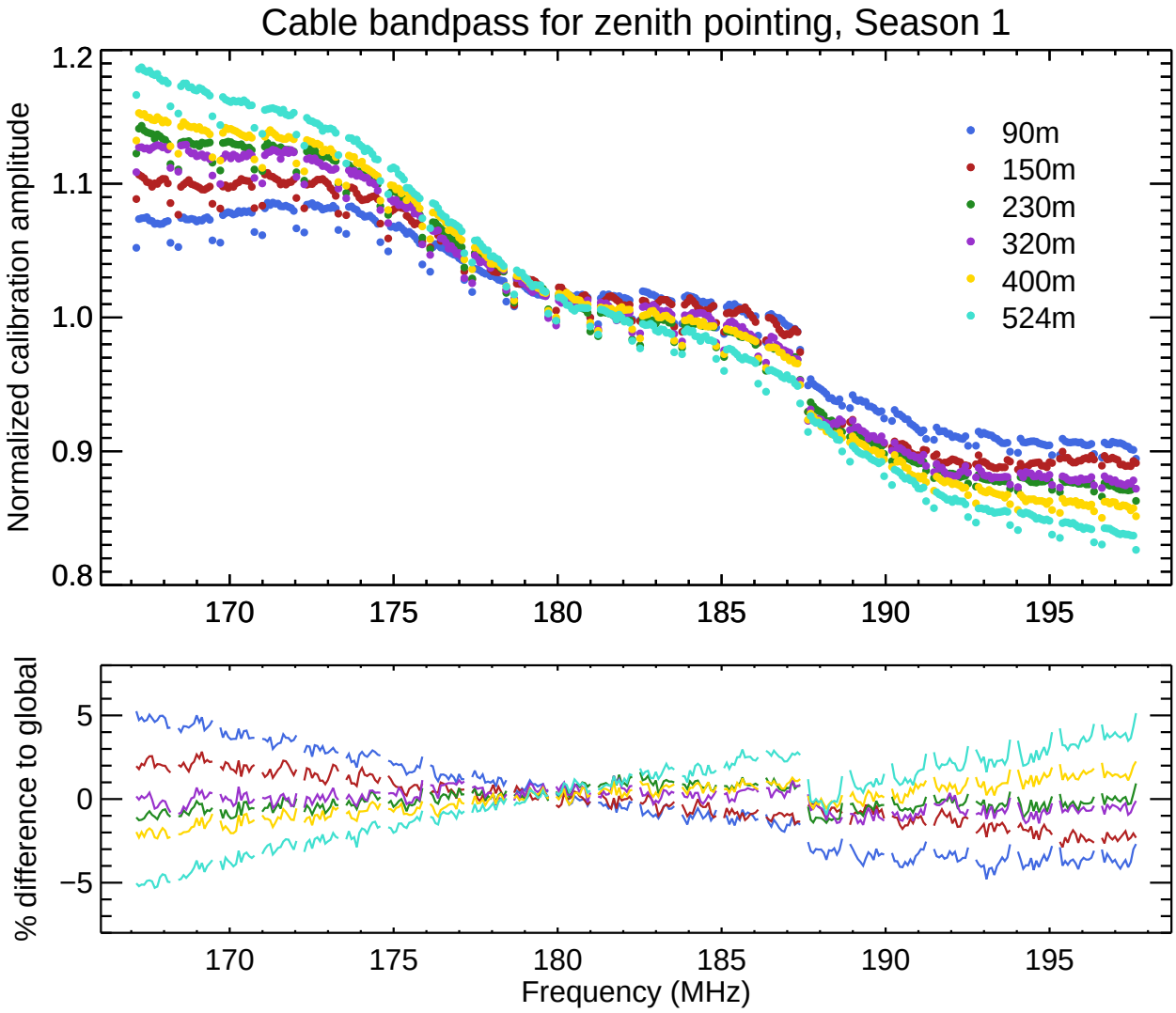


Figure 4.8: The bandpass-per-cable grouping for the zenith pointing of Season 1. The top plot is the resulting amplitude for each of the five LNA-to-beamformer cable lengths, and the bottom plot is the percent difference of the time-averaged cable bandpasses compared to the global bandpass.

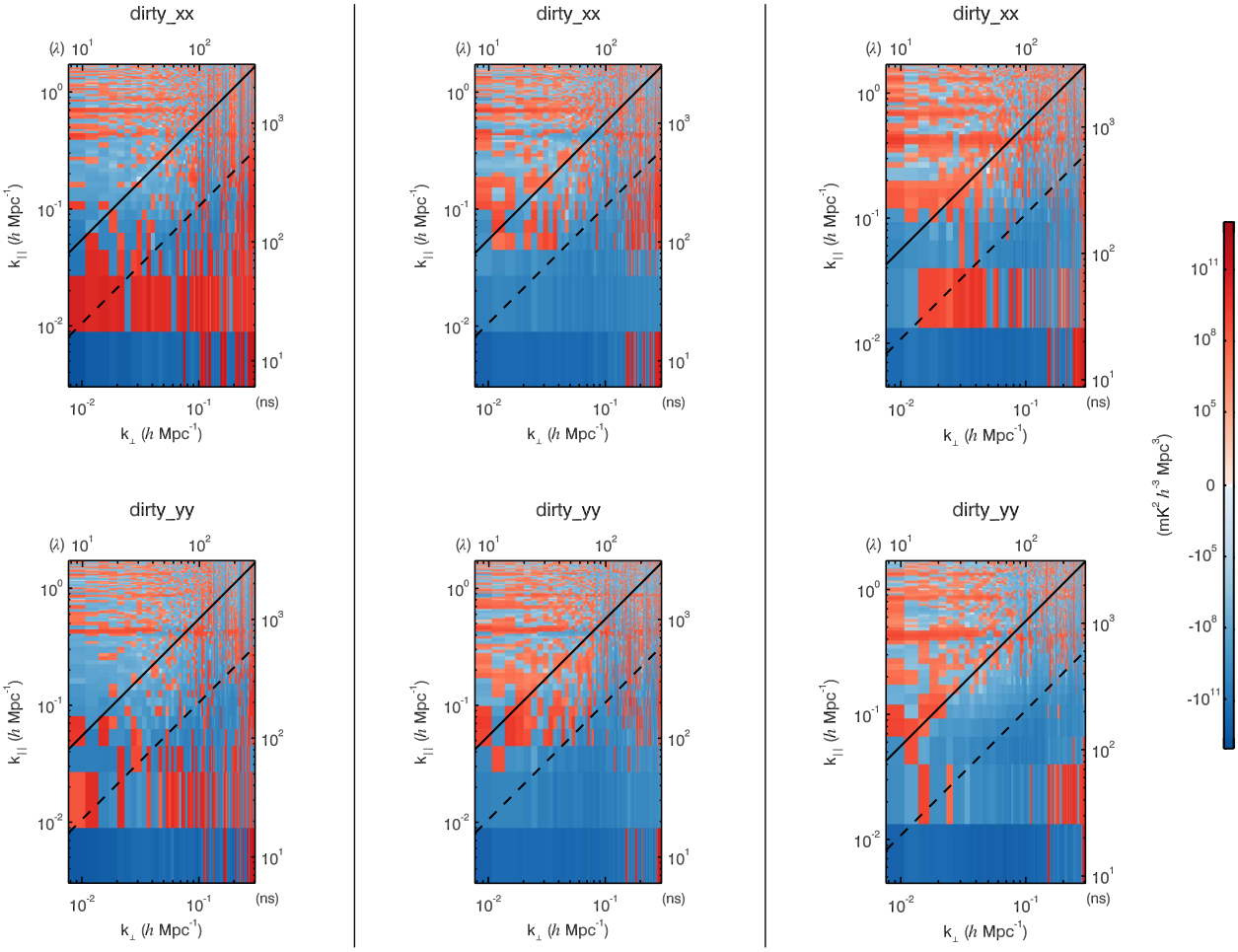


Figure 4.9: 2D power spectrum differences comparing three different references of 64 observations from August 23, 2013, each with a different calibration scheme. The cable- and time-averaged calibration scheme is referenced against: (left column) the global bandpass and fitted cable reflection; (middle column) the global bandpass, fitted cable reflection, and independent digital gain jump; and (right column) the global bandpass, fitted cable reflection, and independent digital gain jump, but only for a subset of the frequency band before the digital gain jump. Blue indicates a decrease in power for the cable- and time-averaged calibration compared to the reference, and red indicates an increase in power. The cable- and time-averaged calibration is degenerate with the fit to the digital gain jump, and it is unclear whether this calibration improves the regions of interest when the degeneracy is removed.

like for YY with perhaps some $k_\perp$-banded structure in XX. Additional averaging in time by pointing has improved the cable-averaging calibration, but has not significantly improved the power spectra compared to previous methods. However, longer integrations would be necessary in order to make definitive conclusions.

4.4.3 By tile and time

When we average over many tiles for the bandpass solution, we are making the assumption that there are no tile-dependent effects in the fine frequency structure. However, this is not a valid assumption. Individual instrumental or hardware variations on each signal chain will influence the bandpass in a unique way. Bandpasses calculated from individual tiles are noisy, hence the justification for the averaged historical bandpass.

This can be mitigated. Disregarding temperature-dependent effects, the MWA is extremely stable in time, and the per-tile instrumental effects will be consistent from one moment to the next. We can average in a similar fashion to §4.4.2 by grouping the observations into a time set $T = \{\rho_{-2}, \rho_{-1}, \rho_0, \rho_{+1}, \rho_{+2}\}$, where times are grouped by number of pointings away from zenith. We then use the equation

$$\left|\mathbf{g}_{A,P}(f;\theta_T)\right| = \mathcal{R}\left\langle\ \left|\mathbf{g}_{A,P}(f)\right|, 2\sigma, \theta_T\ \right\rangle, \tag{4.6}$$

where each tile has a separate bandpass calculation over as much time as possible within a pointing group, excluding outliers. This increases the degrees of freedom compared to Equation 4.5, but attempts to capture per-tile discrepancies. With this new formalization, we no longer need to fit cable reflections.

Figure 4.10 illustrates the result of Equation 4.6 for six representative tiles with separate LNA-to-beamformer cable lengths. The individuality of the tiles is apparent. Cable reflection features are being captured without any additional procedures, even for l_{90} and l_{320} cable types. There are particular deviations near the beginning of the band and near the digital gain jump, especially for the 150 m and 524 m tiles chosen. The coarse band shapes are also very different across frequency. In addition, it should be noted that the noise level is

approximately constant across the tiles, compared to the cable- and time-averaged bandpass, which is particularly noisy for cable types with few contributing tiles.

As with all the other previous bandpass formalisms, we can test the validity of the new calibration in power spectrum space. We integrate over the least-contaminated observations of the night of August 23, 2013 and compare to a series of previous calibrations. The first column of Figure 4.11 is referenced against a calibration with the global bandpass and a fitted 150 m cable reflection (Equation 4.2). The middle column is referenced against a calibration with the global bandpass, fitted cable reflection, and independent fit of the digital gain jump (Equation 4.2 with Equation 4.3). The last column is the same as the middle column, except that it only compares frequencies below the digital gain jump.

Analyzing the 2D power spectrum difference is more difficult in this particular case. The immediately apparent characteristic of Figure 4.11 is a scale change that causes a reduction of power in YY and an increase of power in XX. All modes that are coupled to the foreground wedge will reflect the scale change, which can obscure subtleties. Nevertheless, a similar story emerges.

The EoR window experiences a reduction of power compared to the reference (blue) in the left column of Figure 4.11, indicating a benefit compared to the historical bandpass. However, the EoR window increases in power compared to the reference (red) in the middle column, revealing that the majority of the benefit was from a better handling of the digital gain jump. The final column attempts to ascertain if the new calibration performs better for frequencies below the digital gain jump. With the tile- and time-averaged scheme, noisier bandpasses are yet again the dominant feature; the EoR window experiences a slight increase in power compared to a bandpass with fewer degrees of freedom per observation.

4.5 Auto-correlations

In our search for an experimentally determined bandpass, we try to mitigate thermal, spectral, and bit noise while still maintaining instrumental parameters. However, some noise contributions are more important than others. Using the auto-correlations, or the visibili-

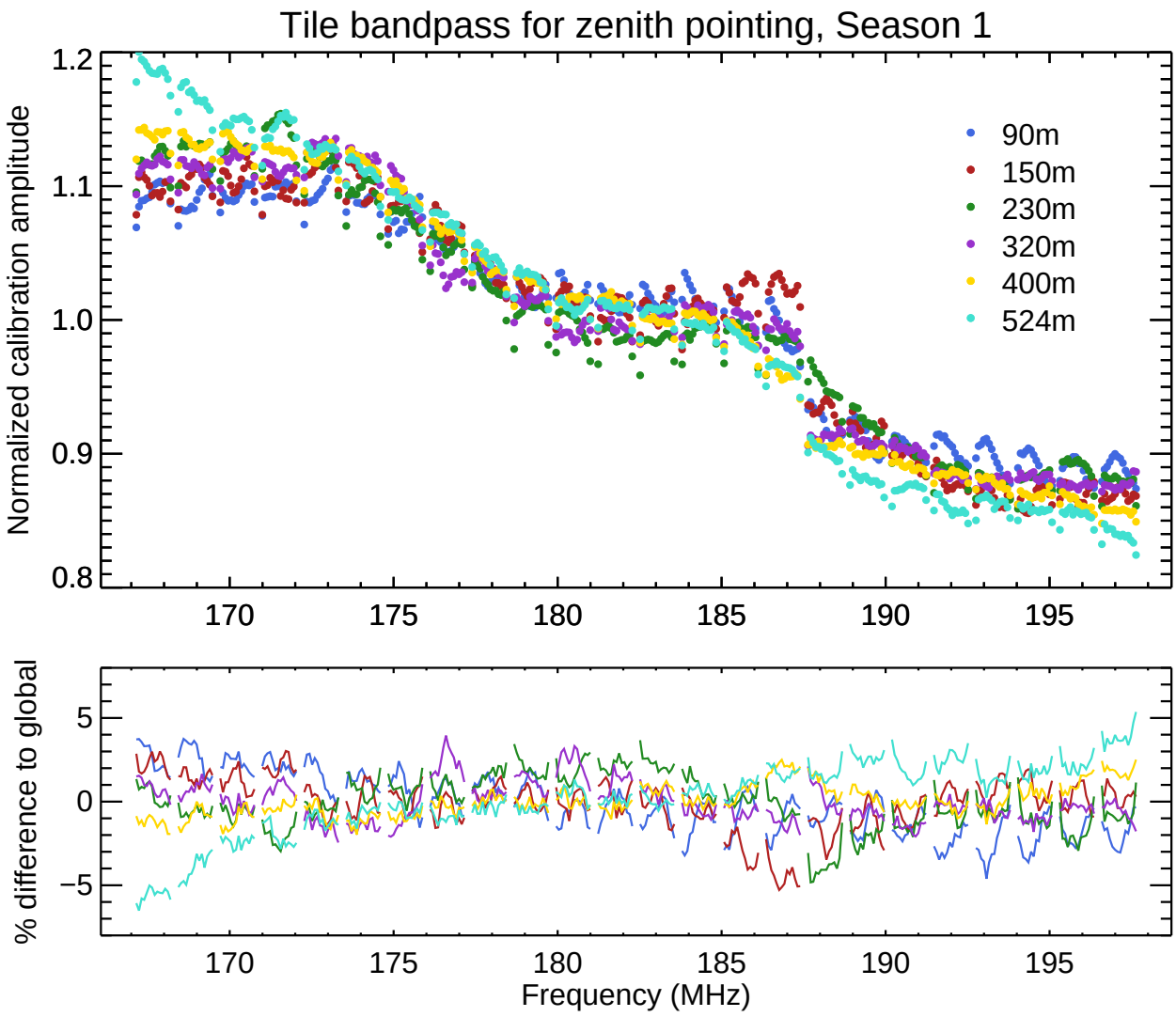


Figure 4.10: The time-averaged bandpass for a variety of tiles representing each cable type (Tile 0 (90 m), Tile 2 (150 m), Tile 37 (230 m), Tile 35 (320 m), Tile 33 (400 m), and Tile 36 (524 m)) for the zenith pointing of Season 1. The top plot is the resulting amplitude, and the bottom plot is the percent difference of the time-averaged tile bandpasses compared to the global bandpass.

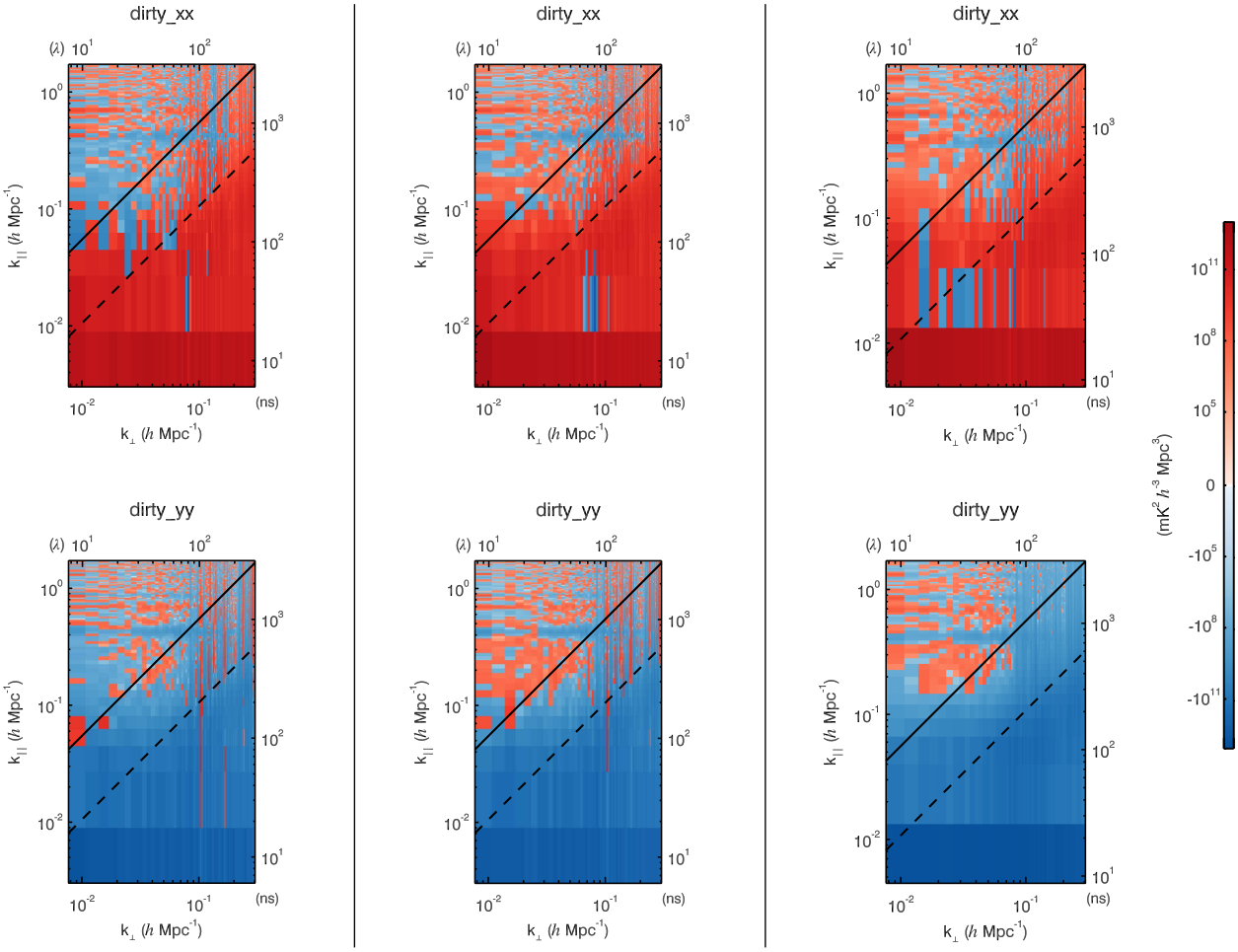


Figure 4.11: 2D power spectrum differences comparing three different references of 64 observations from August 23, 2013, each with a different calibration scheme. The tile- and time-averaged calibration scheme is referenced against: (left column) the global bandpass and fitted cable reflection; (middle column) the global bandpass, fitted cable reflection, and independent digital gain jump; and (right column) the global bandpass, fitted cable reflection, and independent digital gain jump, but only for a subset of the frequency band before the digital gain jump. Blue indicates a decrease in power for the tile- and time-averaged calibration compared to the reference, and red indicates an increase in power. The tile- and time-averaged calibration captures per-tile variation, but is still too noisy to be of benefit in the EoR window.

ties of a tile correlated with itself, as the bandpass solution investigates this balance. The auto-correlations do well to reduce some noise inputs, but enhance others. Whether or not the auto-correlation bandpass has advantages over the averaging techniques depends on the relative importance of the noise contributions.

The auto-correlation bandpass is determined from the same least-squares fitting between raw and model visibilities used for the cross-correlations in §3.1.5. There are fewer equations used to calculate the χ^2 since there is only one baseline pair that creates the auto-correlation. However, all of the time steps are included and thus there is some reduction in the thermal noise, which varies for each time step.

The resulting bandpass is thus $\left|\mathbf{g}_{AA,P}(f)\right|$. No extra averaging in time, tile, or cable grouping is done. Figure 4.12 illustrates the auto-correlation bandpass for the same six representative tiles as in §4.4.3 with separate LNA-to-beamformer cable lengths. The bandpasses exhibit similar features as those averaged by tile and time in Figure 4.10: cable reflection features are being captured for many of the various lengths, individualized deviations are present near the beginning of the band and near the digital gain jump, and per-coarse-band features are apparent.

However, the bandpass is much smoother. The noise due to the incorrectly modeled foregrounds is essentially absent in the auto-correlation bandpass; the auto-correlations are not affected by structure on the sky. While the thermal noise is expected to be higher in the auto-correlation bandpass, it is 1) reduced by including the time steps in the χ^2 calculation and 2) must not be an important contributor to the smoothness of the bandpass.

Nevertheless, there is one major setback. The bit noise is correlated in the auto-correlation bandpass. The edges of each coarse band are subject to an increased contribution from bit noise (see §7.1.1), and the digital gain jump is also subject to the same effect. This will contaminate certain modes in the power spectrum.

We analyzed the auto-correlation bandpass in Figure 4.13 by comparing 2D difference power spectra with the same calibration references as in §4.4. First, we reference against a power spectrum with a global bandpass and cable-reflection fit (Equation 4.2) in the

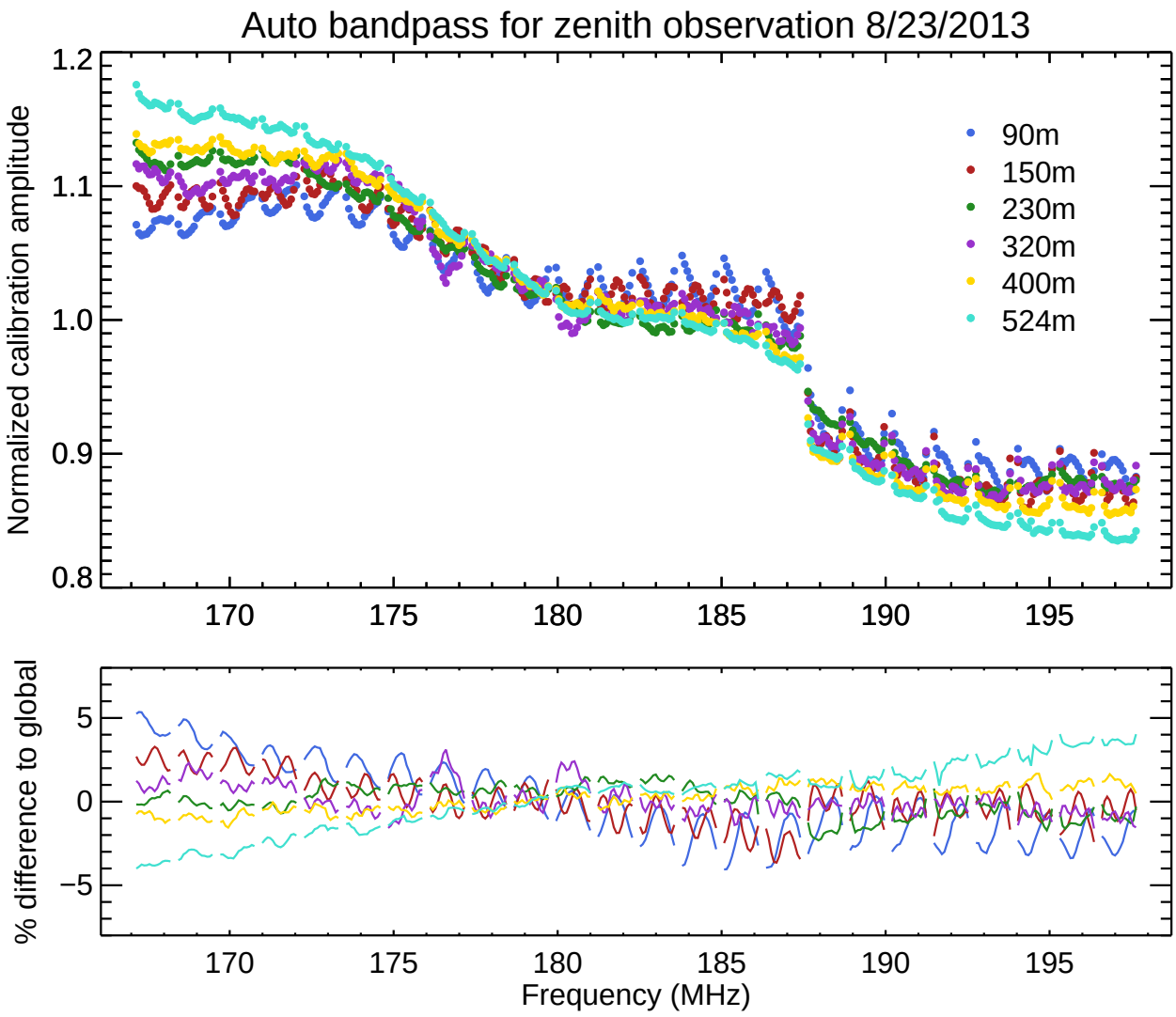


Figure 4.12: The auto-correlation bandpass for a variety of tiles representing each cable type (Tile 0 (90 m), Tile 2 (150 m), Tile 37 (230 m), Tile 35 (320 m), Tile 33 (400 m), and Tile 36 (524 m)) for the zenith observation of August 23, 2013. The top plot is the resulting amplitude, and the bottom plot is the percent difference of the auto-correlation tile bandpasses compared to the global bandpass.

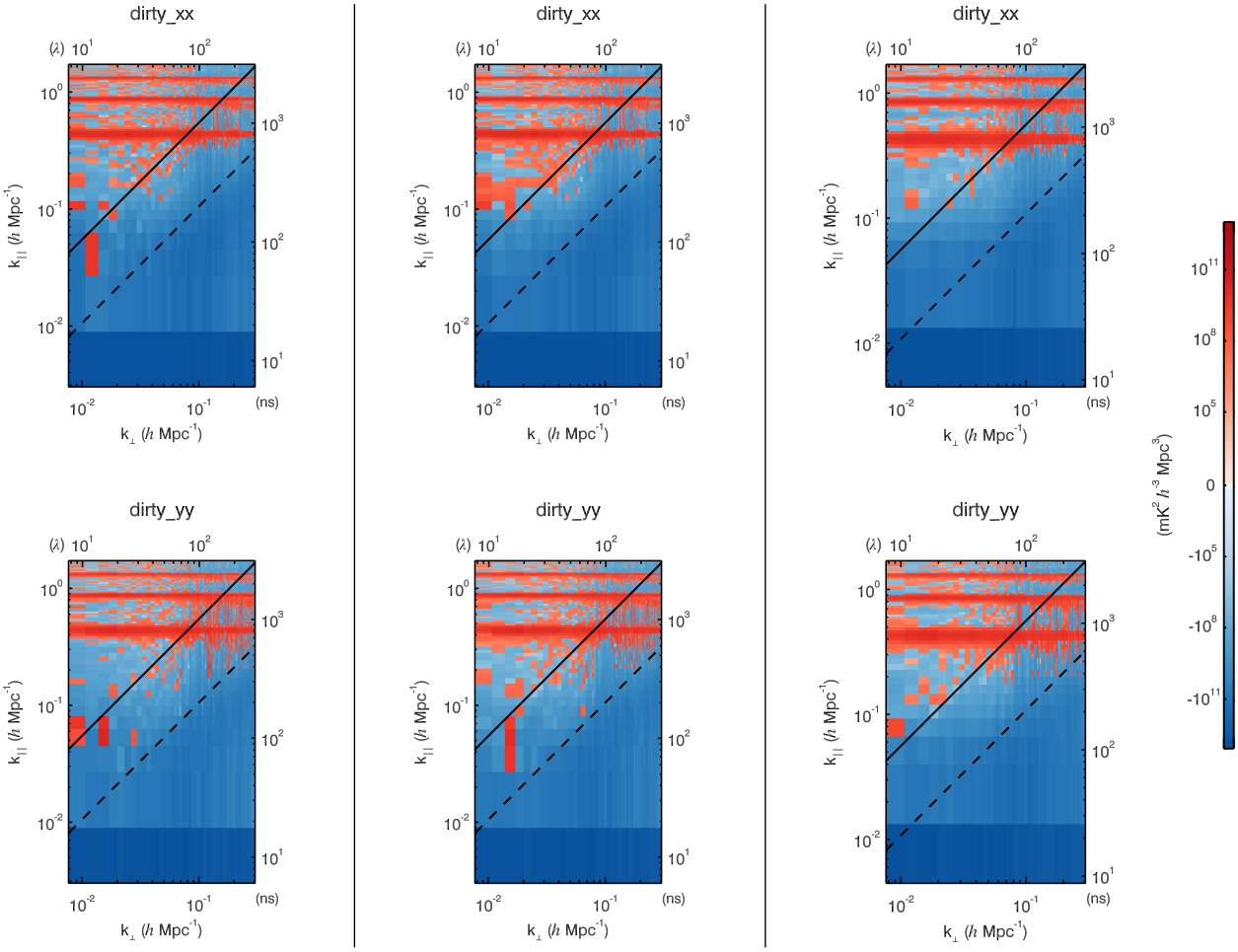


Figure 4.13: 2D power spectrum difference comparing three different references of 64 observations from August 23, 2013, each with a different calibration scheme. The auto-correlation calibration scheme is referenced against: (left column) the global bandpass and fitted cable reflection; (middle column) the global bandpass, fitted cable reflection, and independent digital gain jump; and (right column) the global bandpass, fitted cable reflection, and independent digital gain jump, but only for a subset of the frequency band before the digital gain jump. Blue indicates a decrease in power for the auto-correlation calibration compared to the reference, and red indicates an increase in power. The auto-correlation calibration captures per-tile variation and is less noisy overall, but suffers on the coarse band edges due to bit noise correlation.

left column. There is a reduction of power everywhere except the coarse band harmonics, consistent with an improved bandpass apart from the coarse band edges. Second, we compare to a power spectrum that also includes an independent digital gain jump in the middle column. This result contrasts with all of §4.4; the benefits of the auto-correlation bandpass are *not* fully degenerate with an independent fit to the digital gain jump. Finally, in the right column, we reference with the same calibration scheme as the middle column, but only for frequencies below the digital gain jump. There is still a large benefit in the regions of interest.

4.6 Overview

The bandpass calibration solution captures instrumental parameters which are too complicated to model, and thus it requires many degrees of freedom. As a result, it can be a large source of contamination in our final measurements; it is difficult to constrain and easy to get wrong.

The coarse bandpass shapes, large-scale ripples, band attenuations, digital gain jump, flagged frequency channels, cable reflections, and bit statistics all contribute to features in the bandpass. Most of these effects are variable across tiles, times, and temperatures. We can model these instrumental effects as global contributions, but there will be some error associated with the assumption of stability. In contrast, we can model these contributions along each axis that they vary, but there will be increased thermal and spectral noise associated with the increase in degrees of freedom and fewer averaging components.

We've experimentally tested a variety of different averaging schemes on one night's worth of data. In general, we've found that the largest improvement on the historical, global bandpass comes from improved modeling of the digital gain jump. Thus, all averaging schemes share some degeneracy with a separate, independent fit to the digital gain jump.

Incorrect modeling of the foregrounds causes significant noise on the solutions, and thus there is less contamination in the power spectrum as more is averaged. Once the degeneracy regarding the digital gain jump is removed, it is hard to make improvements on the power

spectrum. Accounting for more instrumental parameters usually results in reducing the number of averaged components.

We also experimentally tested a calibration scheme using the auto-correlations as the bandpass. There are benefits to this method: the auto-correlations are not affected by incorrectly modeling structure on the sky, correlated thermal noise is not a concern due to many time steps, and cable reflections are accounted for without extra fitting. However, the bit noise is correlated and is an issue in the coarse band edges and digital gain jump (see §7.1).

Given all of these experimental results, what is the best bandpass calibration method? All of our investigations demonstrate benefits as well as drawbacks. The choice purely depends on the regions of interest in a particular EoR power spectrum study. Since we desire to make a measurement as low as possible in the EoR window to maximize signal-to-noise, we will use the auto-correlation method for longer integrations in §9.

Chapter 5

FRAMEWORK OF THE IN SITU CALIBRATION SIMULATION

Calibration is inherently an instrumental aspect, and thus the solutions chosen are found experimentally. However, we are still left asking a vital question with our calibration experiments: what is the price of getting the calibration wrong? One potentially incriminating result is signal loss through calibration errors. Does the calibration incorporate the EoR signal itself, model it as an instrumental effect, and essentially remove it from the data? We needed a full pipeline simulation in order to address these concerns.

This motivated a theoretical calibration simulation, a sandbox from which to explore calibration effects in a hypothetical space. In this chapter, we will explore the calibration simulation framework in order to test the consequences of miscalibration. The setup of the simulation framework is vital; the usefulness and flexibility of the pipeline depends on the inclusion of realistic errors.

While made for the testing of calibration schemes and signal loss, our simulation framework has since grown to be a sensitive test for all new instrumental and precision analysis techniques. Therefore, we highlight the methodologies which allowed for its crucial contributions to the analysis pipeline. This chapter is drawn from our published work in [1].

5.1 Simulation setup and rationale

The calibration simulation retools the FHD package, using its existing infrastructure to add functionality. It is an in situ simulation, testing the code's ability to process real data. There are four main steps to the in situ calibration simulation: 1) input simulated data, 2) generate a model sky based on a subset of the input data, 3) calibrate, and 4) investigate the result

in the space of the measurement.

The process to generate step 1 is easy: the sky model made for calibration and subtraction is already a full pipeline simulation. There is no need to bring in outside instrument simulators as FHD is already an instrument simulator. From brightness and sky locations in a catalog, FHD grids with the hyperresolved beam as the kernel to build model visibilities for the MWA layout. These model visibilities are a simulation of reliable point sources on the sky as seen by the MWA. Regularly, we simulate over 11,000 point sources spread throughout the primary lobe, the first sidelobes, and the second sidelobes.

However, there are a few key aspects missing from our typical sky model visibilities. A reliable diffuse emission catalog is not available at this time, and thus the visibilities are inaccurate at the largest modes. Ionospheric effects have not been included. The EoR signal is also missing from the model visibilities; we do not know its exact distribution.

We then input these model visibilities as the input data visibilities into FHD, and run the pipeline again. Thus, we can use FHD end-to-end for the in situ simulation, thereby testing the validity of the pipeline itself. We can optionally add (in visibility space) simulated EoR and/or thermal noise to the input to make the simulated data more realistic.

If the calibration and subtraction model visibilities are the same as the input visibilities, there is not much story to tell. There will be perfect calibration and perfect subtraction, given a converging nature to the pipeline and no signal loss. One of the most important aspects to the in situ calibration simulation is an *imperfect* calibration and subtraction model. There will always be unmodeled faint sources in real data, either due to confusion limits or the inability to resolve morphology, and catalogs will always include small errors in flux, position, or compactness of sources. These imperfections in the sky model will affect calibration, and subsequently the power spectrum. A true test of a calibration scheme must include realistic errors in the model.

Once a calibration scheme has been tested, the in situ simulation continues along the pipeline to generate power spectrum with εppsilon. This is a relatively unique concept; only a few studies have investigated calibration effects on the power spectrum (e.g. [111, 101]).

Historically, calibration effects have been tested primarily on their ability to converge. Variance and convergence statistics of calibration in image space have been studied in detail, including the application of ionospheric changes [106, 56, 115, 17], variation of diffuse structure scales [56], addition of polarized components [56], inclusion of small source position offsets [64, 115, 17], and the imperfection of source models [17, 18]. However, the EoR measurement is to occur in power spectrum space, so the standard of the effectiveness for calibration should also be established in power spectrum space. While variance statistics can assess the *precision* of a calibration procedure, the power spectrum in a calibration simulation can assess the *accuracy*.

5.2 Signal loss

The original motivation for creating the in situ calibration simulation was to investigate signal loss. It is imperative to provide evidence that the EoR signal is not being subtracted in an EoR analysis. Removing signal can create false limits. Therefore, we set up an in situ simulation with a typical calibration scheme to validate our pipeline.

5.2.1 Gaussian EoR

For this test, a model of over 11,000 compact sources seen by the MWA and compiled in the KGS catalog [12] were used as simulated input data, along with the addition of a simulated Gaussian EoR signal in the visibilities. This approach is completely noiseless, and contains no information about ionospheric effects or diffuse galactic emission.

In addition to the simulated input data, a model of the sky is generated for sky-based calibration. Antenna gain solutions that minimize the differences in visibilities between the input data and the calibration model are calculated through an iterative, least-squares approach using all cross-correlated visibilities [90] (see §3.1.5). The final result constitutes our calibration solutions, which captures maximal per-frequency spectral structure. In essence, this is the worst case scenario, a concept that is explored in depth in §6.

When we allow the simulation to use all of the input catalog sources as a model from

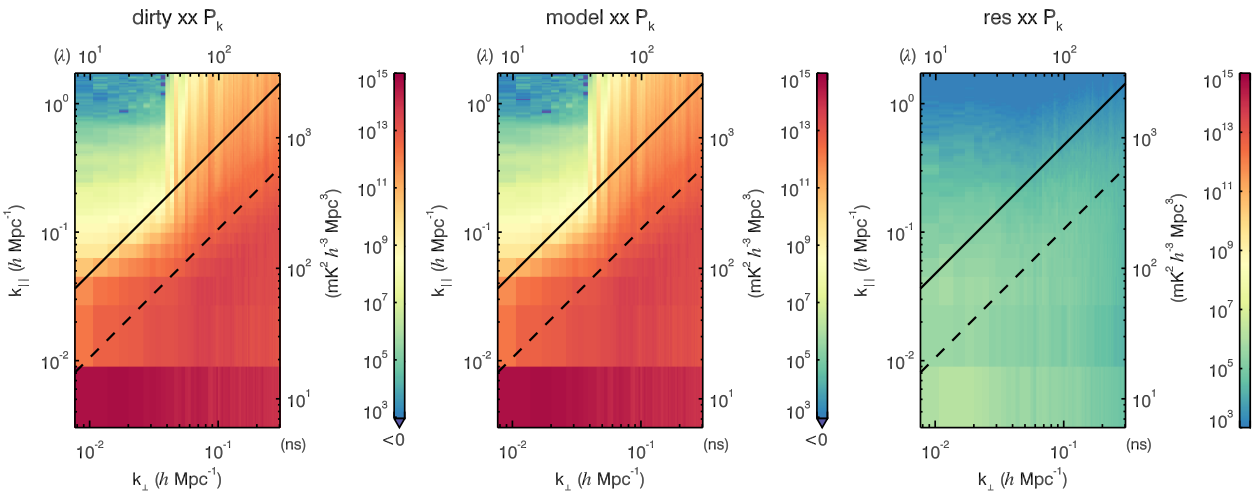


Figure 5.1: Result of the calibration simulation pipeline using a Gaussian EoR with a perfect sky model. Modeling, calibrating, and subtracting all of the 11,000+ KGS sources using FHD and εppsilon recovers the added simulated EoR signal. The EoR signal that we recover does not experience signal loss, regardless of calibration technique used. The color scale has been fixed throughout this chapter and the next to provide order-of-magnitude reference.

which to calibrate and subtract, all foreground sources are removed perfectly. This reveals the simulated EoR signal in the residual power spectrum with no foregrounds or chromaticity effects, as seen in Figure 5.1 for one snapshot. The color scale has been fixed to highlight the order of magnitude difference between the EoR signal peaked at $10^6\,\mathrm{mK}^2\,h^{-3}\,\mathrm{Mpc}^3$ and the intrinsic and mode-mixed foregrounds peaked at $10^{14}\,\mathrm{mK}^2\,h^{-3}\,\mathrm{Mpc}^3$.

We compare the residual signal to the power that was used to generate the Gaussian EoR (along with image-based EoR, §5.2.2) in Figure 5.3 for a more quantitative estimate. The binning of the input Gaussian EoR power is coarser, but otherwise there is very little difference.

Retrieving the EoR power spectrum demonstrates consistency within the simulation, and will provide a magnitude scale for simulation outputs with unmodeled, faint sources. This also demonstrates that the pipeline effectively recovers the EoR signal with very lit-

tle contamination or signal loss if the foreground model is perfect, regardless of additional calibration techniques from §4.

5.2.2 Image-based EoR

The same tests can be run with a more realistic EoR simulation. A Gaussian EoR simulation is convenient and correct in the power spectrum space, but is not accurate in image space. We can incorporate EoR bubble simulations to improve the validity of the simulation.

We use a simulation generated from [45], made specifically in the regions of interest for the MWA. It's a large N-body simulation ($130\,\mathrm{Mpc\,h^{-1}}$ comoving) in image space at a high resolution (1024^3 underlying dark matter simulation, with a 512^3 radiative transfer simulation from [53]). However, this is not large enough for our widefield requirements in the data pipeline. We tile the simulation, satisfying boundary conditions, until it fills our field of view. The value at each image pixel is Fourier-transformed to the uv-plane, where we estimate the instrumental measurement by convolving each baseline with the beam to make visibilities.

The resulting bubble visibilities have some inherent assumptions. Of course, there were assumptions made during the simulation that generated them in [45], but the process to turn them into visibilities introduced more assumptions. First, our tiling operation inhibits any meaningful structure at modes larger than the input simulation ($130\,\mathrm{Mpc\,h^{-1}}$ comoving). Second, the finite gridding resolution erases meaningful structure smaller than the gridding size. Third, the gridding may introduce biases dependent on the gridding selection. However, we are not interested in modes that are not in the EoR window, and thus the consequences of these assumptions are minor.

Using the same setup as §5.2, we can investigate signal loss with image-based EoR structure. This arguably has more chance of potential signal loss, since real image-based structure could be absorbed into point sources through calibration. The 2D residual power spectrum from calibrating and subtracting all point sources with an image-based EoR is shown in Figure 5.2.

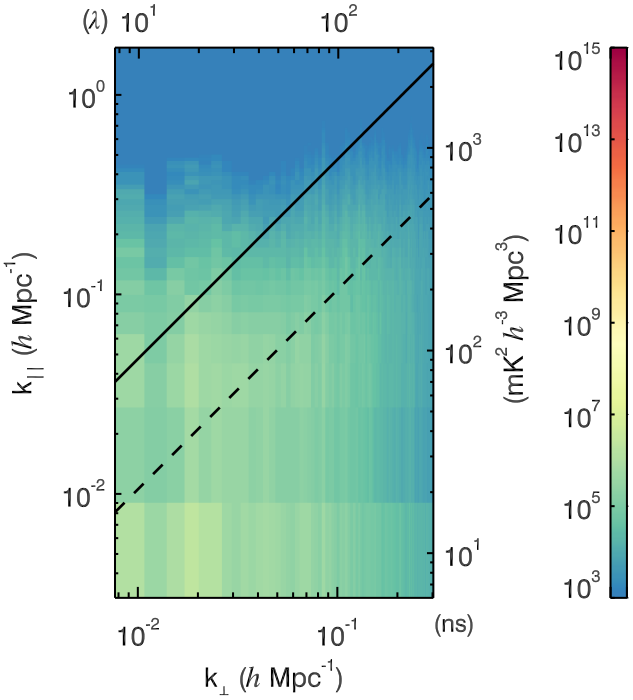


Figure 5.2: Result of the calibration simulation pipeline using an image-based EoR with a perfect sky model. Modeling, calibrating, and subtracting all of the 11,000+ KGS sources using FHD and εppsilon recovers the added image-based EoR signal. Neither Gaussian EoR signal nor image-based bubble EoR simulations experience signal loss in our pipeline.

For this case, we need to create a reference EoR power spectrum. We input the image-based EoR visibilities without foregrounds and without calibration in our pipeline and generate a power spectrum. This will provide comparison for a signal loss simulation that focuses specifically on loss through foreground calibration and removal. It is significantly lower at higher $k_{||}$ than the Gaussian EoR signal in §5.2.1, as presented in a 1D power spectrum in Figure 5.3.

When we compare the reference to the recovered image-based EoR in Figure 5.3, we see that there is no signal loss. While there is some difference at large k, there is no obvious structural differences that indicate systematic signal loss. We have recovered the EoR signal.

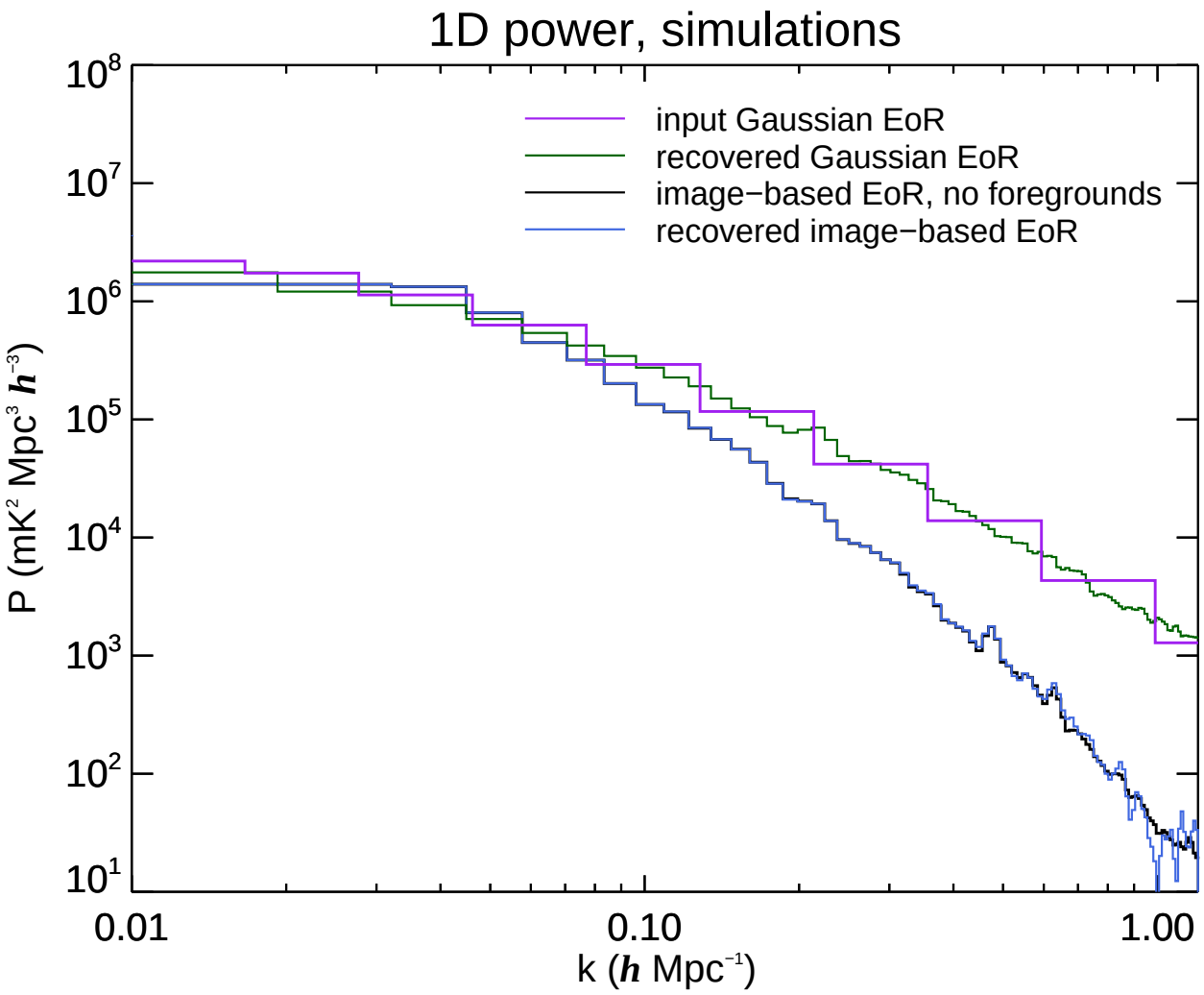


Figure 5.3: 1D power spectra for the signal loss simulations. The input Gaussian EoR signal (purple) uses a different bin size, in part because it is fully independent from our pipeline. We recover the Gaussian EoR signal after foreground calibration and removal (green). The image-based EoR signal is generated from [45], and the power is calculated without foregrounds or calibration (black). Again, we recover the image-based EoR signal after foreground calibration and removal (blue). There are differences at high k, but they are not indicative of systematic signal loss.

5.2.3 Phase II configuration

The MWA has undergone improvements and upgrades which have changed the configuration, described in §2.2.4. Therefore, the sensitive modes in the MWA have changed. The relatively smooth point spread function from Phase I is now coarse, a consequence of redundant measuring of EoR sensitive modes. We would like to validate these new changes in our pipeline using the same signal loss tests as in §5.2.1.

We repeat the test in §5.2.1: input over 11,000 point sources as the foregrounds, add a Gaussian EoR signal, and calibrate and subtract the input point sources. Figure 5.4 shows our full 2D power spectra diagnostic of the calibrated raw, model, and residual for instrumental XX.

There is a significant difference between the resulting power spectra from Phase I in Fig-

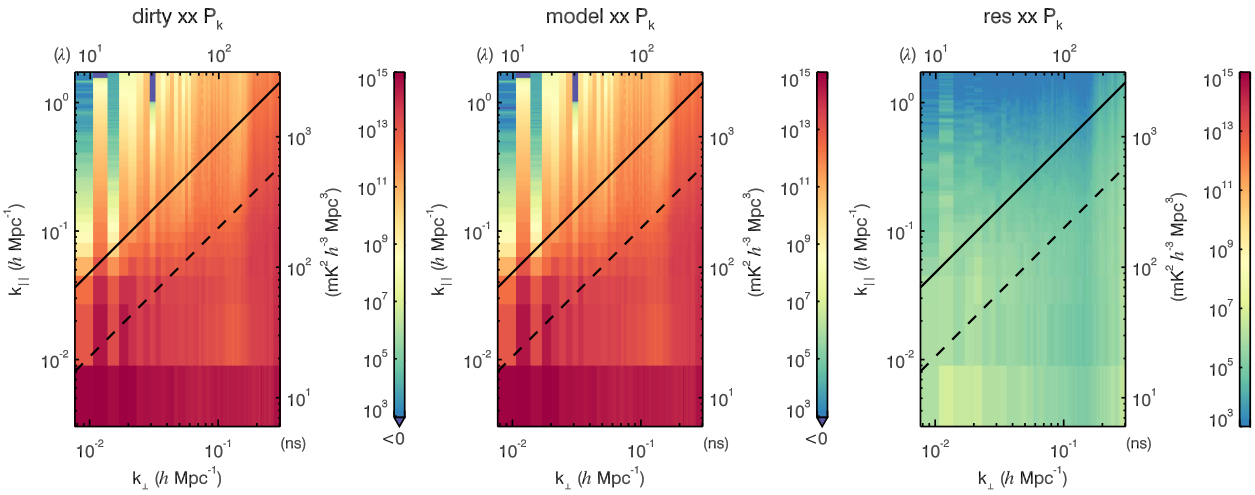


Figure 5.4: Result of the calibration simulation pipeline using a Gaussian EoR with a perfect sky model and the Phase II configuration of the MWA. Modeling, calibrating, and subtracting all of the 11,000+ KGS sources using FHD and εppsilon recovers the added Gaussian EoR signal, even though the uv-coverage of the Phase II configuration is concentrated on only a few modes. Our pipeline is flexible enough to handle multiple configurations and EoR theories.

ure 5.1 and Phase II in Figure 5.4. Notably, there are more regions of incomplete uv-coverage in our typical plots, which manifests as vertical streaks of foreground-coupled regions. However, we still recover the underlying Gaussian EoR signal when all point sources are used in calibration and subtraction. The residual 2D power spectrum is essentially the same as the input EoR signal. Potential aliasing effects may be more prominent in analyzed data from Phase II, resulting in contaminated vertical streaks (see §8.3 for treatment).

Given the evidence presented in this section, our pipeline does not experience signal loss. It is flexible enough to handle different EoR theories and MWA configurations while still providing robust output. While this is no indication of how favorable our limits will be, it gives us confidence in our ability to quote reliable upper limits.

Chapter 6

IN SITU CALIBRATION SIMULATION RESULTS

Traditional radio astronomy calibration techniques involve calculating the gain at every frequency for every antenna from the visibilities using an iterative least-squares solver (see §3.1.5). Historically, this approach was an extension from single-frequency radio astronomy to small-bandwidth multi-frequency instrumentation [102, 98]. While this method involves solving for many variables, the number of degrees of freedom in the data from the MWA is orders of magnitude larger than the number of parameters used and thus theoretically constrained. This calibration method has remained a stalwart in the community as the field has advanced.

Our calibration simulations first examine the traditional radio astronomy calibration technique and how it affects the EoR power spectrum measurement. We then investigate mitigation techniques and create recommendations for future instruments. This chapter is drawn from our published work in [1].

6.1 Calibration errors due to faint, unmodeled sources

We simulate the sky as 6,950 sources and add a Gaussian EoR simulation to the visibilities. To capture realistic differences between the true sky and the calibration catalog, we only use the brightest 4,000 to predict the visibilities for use in calibration. This introduces small differences between the sky and calibration visibilities that can affect the per-frequency antenna calibration solutions. We apply the antenna calibration solutions to the input sky visibilities, and then subtract the brightest 4,000 sources used in the calibration model for one observation. Reasoning and methodology is discussed in §5.

The residual 2D power spectrum is shown as the left plot of Figure 6.1. The 2,950

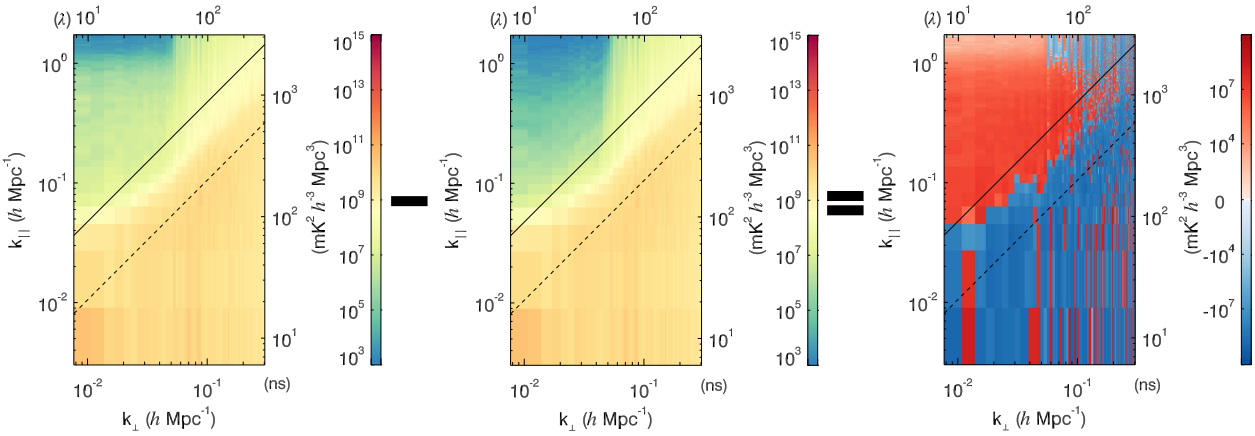


Figure 6.1: The subtraction of a residual 2D power spectrum with traditional per-frequency antenna calibration methods (left) and a reference residual 2D power spectrum without simulated calibration effects (middle) to create a difference 2D power spectrum (right). Red indicates a relative excess of power, and blue indicates a relative depression of power. Spectral contamination power at all modes in the EoR window is evident. The most sensitive, theoretically contaminant-free EoR modes have excess power on levels of $10^7\,\mathrm{mK}^2\,h^{-3}\,\mathrm{Mpc}^3$, making the measurement impossible with reasonable calibration catalog errors and traditional per-frequency antenna calibration.

unmodeled faint sources populate the foreground wedge as expected. We can also calculate a residual 2D power spectrum with a perfect calibration and with the same 4,000 source foreground subtraction to provide a reference for the observation. This is shown in the middle panel of Figure 6.1. Unmodeled faint sources also populate the foreground wedge; however, it is apparent that the 2D power spectrum using traditional per-frequency antenna calibration has relatively high amounts of power in the EoR window.

The calibration simulation can be used to quantify the shape and amount of excess power in power spectrum space. Direct subtraction between the traditional calibration simulation 2D power spectrum (left plot of Figure 6.1) and the reference 2D power spectrum (middle plot of Figure 6.1) provides this information. The result is a difference 2D power spectrum,

where red indicates relative excess power and blue indicates relative depressed power (right plot of Figure 6.1). More information about 2D difference power spectra can be found in §3.4.3.

Evident in the difference 2D power spectrum is the excess power contamination of the entire EoR window by as much as $10^7\,\mathrm{mK}^2\,h^{-3}\,\mathrm{Mpc}^3$. It is important to note that this level of calibration error would make EoR power spectrum measurements impossible. Using traditional per-frequency antenna calibration in a power spectrum measurement would require a calibration catalog with near perfect accuracy.

Qualitatively, allowing independent calibration parameters for each frequency channel and antenna allows small deviations from the true solutions on small spectral scales. These amplitude and phase deviations are caused by the point spread functions (**PSF**) of unmodeled sources, which modify the observed fluxes of true sources, as seen in real data by [67]. This effect is frequency dependent and its magnitude depends on the completeness of the sky model and the natural PSF of the array. The resulting calibration errors are only on the order of 1 part in 10^3 in this simulation. However, this varied spectral structure in the calibration solutions is enough to couple power from the bright, intrinsic foregrounds to the Fourier modes in the EoR window. This fills every possible EoR measurement mode with foreground power.

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

Using the modulation theorem, we can quantitatively associate the level of contamination seen in the power spectrum with the observed calibration errors. Data that is modified by spectrally varied calibration solutions is Fourier-transformed into power spectrum space, and the modulation theorem of Fourier transforms results in mode-mixing between the modes

of the unmodeled spectral structure and the bright foreground wedge. This couples the response of foregrounds with calibration deviations along the frequency axis.

Excess power can be estimated given a modulated signal

$$h(\nu) = f(\nu)\,(1 + \Delta g \cos \eta_0 \nu)\,, \tag{6.1}$$

where $h(\nu)$ is the modulated instrumental response as a function of frequency, $f(\nu)$ is the original instrumental response as a function of frequency, η_0 is the Fourier dual of a mode in the amplitude deviations of the calibration gain, and Δg is the amplitude deviation associated with the frequency mode η_0. The modulation theorem results in the Fourier transform

$$H(\eta) = \frac{\Delta g}{2} F(\eta - \eta_0) + \frac{\Delta g}{2} F(\eta + \eta_0) + F(\eta). \tag{6.2}$$

Fourier transforms of the original signal f construct signal at η, $\eta - \eta_0$, and $\eta + \eta_0$. Equation 6.2 is squared to obtain the power spectrum, and cross-terms between $F(\eta)$ and $F(\eta \pm \eta_0)$ can be neglected since overlap is small for an η_0 in the EoR window. An order of magnitude estimate of the positive power spectrum of this modified signal is

$$\mathcal{O}(|H(\eta)|^2) \approx \mathcal{O}(|F(\eta)|^2) + \mathcal{O}\left(\left|\frac{\Delta g}{2} F(\eta \pm \eta_0)\right|^2\right). \tag{6.3}$$

As a result, the modulated power response $\mathcal{O}(|H(\eta)|^2)$ has power contributions as a function of η and, to a lesser extent, $\eta \pm \eta_0$. When all η and η_0 values are considered, the result is equivalent to the convolution of the foregrounds with the Fourier transform of the calibration deviations.

For small η values, intrinsic foregrounds dominate. Power will be modulated from these intrinsic foregrounds into any frequency mode η_0 captured in the amplitude deviations in calibration. Given simulation values of the intrinsic foregrounds ($\mathcal{O}(P) \approx 10^{14}\,\mathrm{mK}^2\,h^{-3}\,\mathrm{Mpc}^3$) and the amplitude deviations ($\mathcal{O}(|\frac{\Delta g}{2}|^2) \approx 10^{-7}$, or a Δg of order 1 part in 10^3), the excess contamination in frequency mode η_0 of the power spectrum is estimated to be $10^7\,\mathrm{mK}^2\,h^{-3}\,\mathrm{Mpc}^3$. This agrees with the level of contaminated power in Figure 6.1 generated by calibration simulations.

The satisfactory performance of traditional per-frequency antenna calibration depends on a highly accurate calibration catalog. When we use the same sources to generate the sky and calibration models—even with an added EoR signal—the resulting calibration and foreground suppression in the power spectrum is excellent, as seen in Figure 5.1. However, this is not a realistic situation for current and planned EoR observatories. When the calibration catalog is not perfect, traditional per-frequency antenna calibration distributes spectral power and overwhelms the faint cosmological signal as seen in Figure 6.1. This sets very strong constraints on the accuracy of the calibration catalog if the traditional calibration approach is to be used for EoR measurements.

6.2 Mitigation by smooth calibration solutions

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

6.2.1 Constraining smooth instrumental response

If an antenna has a naturally smooth bandpass, its response can be modeled with low-order polynomials or other slowly varying functions. With this restriction, we avoid the fine-scale spectral structure in the calibration solutions that causes the contamination of the EoR window seen in §6.1. Perfect calibration solutions are flat in our simulation, so polynomials applied to the frequency band would only model the level of error expected with polynomial fitting.

We calculate best-fit polynomials over the whole frequency band from the traditional calibration solutions generated in §6.1 with a calibration catalog of the brightest 4,000 of

the 6,950 input sources. Five calibration parameters for the frequency band are allowed and are chosen to represent typical instrumental variation: three amplitude parameters create a second-order bandpass-like polynomial, and two phase parameters create a smooth ramp in frequency.

Figure 6.2 shows the difference 2D power spectrum between the smooth calibration solution power spectrum and the reference power spectrum for one observation. The EoR window from the smooth mitigation technique and the reference are strikingly similar, leading to very little difference. The level of the difference is also noise-like and far below the EoR signal. This will neither affect an EoR measurement to a significant degree nor bias the result.

Power from intrinsic foregrounds is not coupled to the EoR window due to the restriction of smoothness relative to EoR spectral modes. Spectral contamination on the scale of the polynomials still occurs; however, this contamination only occurs within the foreground wedge and will not hinder the EoR measurement. The bold line in Figure 6.2 highlights the highest $k_{||}$ with significant contamination caused by the low-order polynomial fitting, which is below the EoR window.

Differences in power in the region of high $k_{\perp}$ and high $k_{||}$ in Figure 6.2 are also apparent. Poor baseline coverage couples the foreground wedge to this region, as described in §3.4.1[1]. Since spectral contamination did occur in the foreground wedge, power changes occur in the region affected by poor baseline coverage. EoR measurements will not be made in k-space areas with poor baseline coverage, so power changes in this area are not a large concern. The vertical dashed line in Figure 6.2 indicates the largest $k_{\perp}$ with high baseline coverage. A significant foreground-free EoR window to the upper left remains.

By restricting the instrumental response to be smooth with respect to EoR spectral modes, we significantly reduced the excess power in the EoR window caused by spectral contamination. Now, the EoR window has no power bias and what little contamination

[1]This coupling has since been reduced. See §8.3 for treatment.

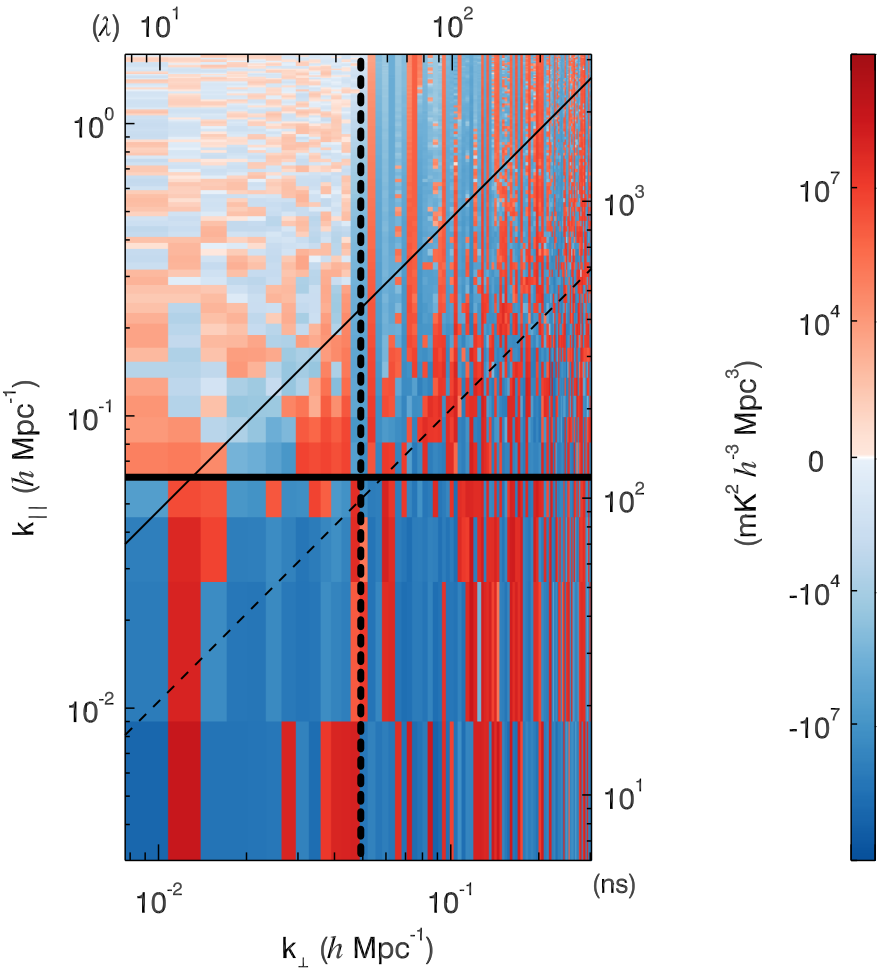


Figure 6.2: The difference 2D power spectrum between a reference 2D power spectrum and a 2D power spectrum where calibration solutions were modeled with a second-order polynomial in amplitude and a linear fit in phase. Red indicates an excess of power, and blue indicates a depression of power. The bold line indicates the largest $k_{||}$ affected by power spectral contamination from low-order polynomial fitting. The vertical dashed line indicates the largest $k_{\perp}$ not affected by the coupling of poor baseline coverage to the foreground wedge. The EoR window is noise-like and unbiased, since smooth mitigation techniques did not capture fine-scale spectral structure caused by unmodeled sources.

there is appears noise-like. Measuring the EoR signal in a 2D power spectrum that utilizes the smooth mitigation technique in the calibration solutions will be essentially unaffected by an imperfect catalog. Instrumentation that is spectrally smooth can avoid contaminating the EoR window using this technique.

6.2.2 Calibration parameters in spectral modes

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

Receiver-to-beamformer cable reflections with amplitudes of $\sim$1% of the signal are apparent in MWA data, creating a characteristic frequency ripple in the antenna gain at the corresponding light travel time [22, 2, 24, 40]. Cable reflections can vary dramatically from antenna to antenna, and must be fit individually (see §4.2 for more details). Our calibration simulation uses three parameters—mode, amplitude, and phase—to describe the spectral ripple from a hypothetical 150 m cable reflection in a subset of the antennas. The 150 m cable reflection does not actually exist in the simulated data, thus the fit responds to only unmodeled spectral structure from faint sources. For clarity, no other calibration parameters are included. The observed error in the reflection calibration parameters is less than 1 part in 10^3.

The effect of coupling the intrinsic foregrounds and the imperfect fitted mode is shown in Figure 6.3. The difference 2D power spectrum between the reference and the power spectrum with fitted cable reflections in calibration shows a clear excess of power at $k_{||} \sim 0.7\,h\,\mathrm{Mpc}^{-1}$, or the $k_{||}$ associated with a 150 m spectral ripple.

The amount of excess power, about $10^6\,\mathrm{mK}^2 h^{-3}\,\mathrm{Mpc}^3$, is not a coincidence. The accuracy levels of the fit and the amplitude deviations in traditional per-frequency calibration in §6.1 were similar. Whether the calibration is described per-frequency (§6.1) or per-spectral-mode ($\propto k_{||}$), the same number of calibration terms are required to cover the bandwidth and the

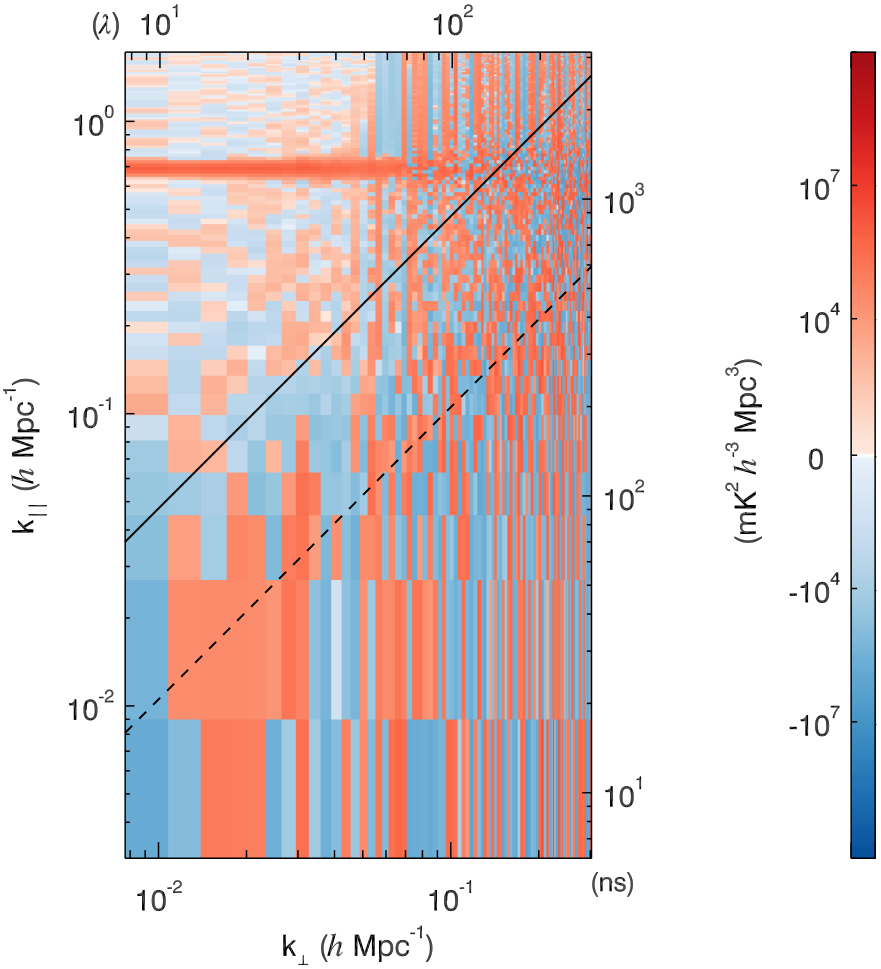


Figure 6.3: The difference 2D power spectrum between the reference 2D power spectrum and a 2D power spectrum where a 150 m cable reflection is fit to within 1 part in 10^3. Red indicates an excess of power, and blue indicates a depression of power. Spectral contamination occurs when there is error in the fit, and mode-mixing occurs between bright, intrinsic foregrounds and the mode associated with the fit error. Excess power at $k_{||} \sim 0.7\,h\,\mathrm{Mpc}^{-1}$ is over $10^6\,\mathrm{mK}^2\,h^{-3}\,\mathrm{Mpc}^3$. This essentially removes that mode from potential EoR measurements.

same level of contamination results.

Fitting for instrumental spectral structure necessarily removes those modes from potential measurement of the EoR. If the number of instrumental spectral features that must be calibrated fills much of the EoR window, measurement of the signal will be infeasible. This necessitates a smooth instrumental response in the modes associated with the EoR window.

As an example of how this affects instrument design, the MWA has chosen to only have a few different cable lengths in the array. This limits the regions of contamination within the the EoR window—having many different cable lengths would contaminate the entire measurement region. This has also led to a hard limit of all analog signal paths being $<35\,\mathrm{m}$ in HERA.

6.3 Mitigation by averaging calibration solutions

Calibrating fine-scale instrumental frequency structure requires accurately modeling the antenna response while avoiding spurious spectral structure from unmodeled sources. This faint, unmodeled structure can be largely incoherent between antennas and can change quickly with Local Sidereal Time (LST) (due to an effect known as uv-rotation). We explore calibration techniques that average over antenna responses and over LST to reduce the spectral contamination identified in this chapter.

Each antenna calibration solution is calculated from the subset of visibilities which include that antenna. The associated "antenna PSF" captures the effect of unmodeled sources on that specific calibration solution. Since each antenna's baseline coverage is largely independent for non-redundant arrays, the spurious spectral structure from unmodeled sources varies from antenna to antenna. If the antennas are identical in manufacture, then averaging their calibration solutions to form a common bandpass can reduce the calibration amplitude and phase deviations that cause spectral contamination. This may not be true for arrays with redundant layouts or layouts with strong symmetries, where the antenna PSFs can be very similar.

Additionally, if the antenna calibrations are very stable in time, subsequent calibration

solutions can be averaged effectively. This relies on rotation of the antenna PSFs with LST to provide semi-independent contamination from unmodeled sources.

We simulate a hypothetical array with the MWA layout (very random distribution, [4, 3, 105]) with mechanically identical and stable antennas. Using traditional calibration techniques, we calculate solutions per frequency channel for each of the 128 antennas in the MWA, every 2 minutes, for a 30-minute pointing traversing zenith. Following Equation 4.5, the resulting 1,920 solutions per frequency (128×15) are then averaged, excluding outliers beyond a 2σ cut. The final averaged per-frequency calibration solution is then applied to all antennas for only one observation, allowing direct comparison to the other simulations in this chapter.

Figure 6.4 presents the difference 2D power spectrum between a reference and the power spectrum with the averaged calibration solution. The amount of excess power in the EoR window has decreased by over three orders of magnitude compared to Figure 6.1 with traditional per-frequency and per-antenna calibration. However, relative excess power still remains and lines of constant $k_{||}$ contaminate the EoR window, indicating that spectral structure from unmodeled faint sources is still present in the average bandpass solution. The excess power level is similar to the expected power of the EoR. Whether or not an EoR measurement is feasible with this fully averaged calibration scheme will depend on the completeness of the sky model, similarity of the antennas across elements and time, and the instrument's design.

Figure 6.5 compares all difference power spectra presented in this chapter to the expected level of the EoR. We average $k_{\perp}$ from 10 to 20 λ to generate a 1D power spectrum as a function of $k_{||}$ in the EoR window. The level of contamination should be significantly below the EoR in order to realistically detect it, as there are no other sources of errors included. We find that the contamination from the maximally averaged calibration solution over all possible LSTs and antennas is at the level of the EoR, using a 4,000-source sky model. Therefore, this is not a practical solution for the MWA. In contrast, we find that using low-order polynomials, as described in §6.2, is our best calibration method; contamination is lower than the expected EoR signal by up to two orders of magnitude. Current efforts to calibrate MWA EoR data use

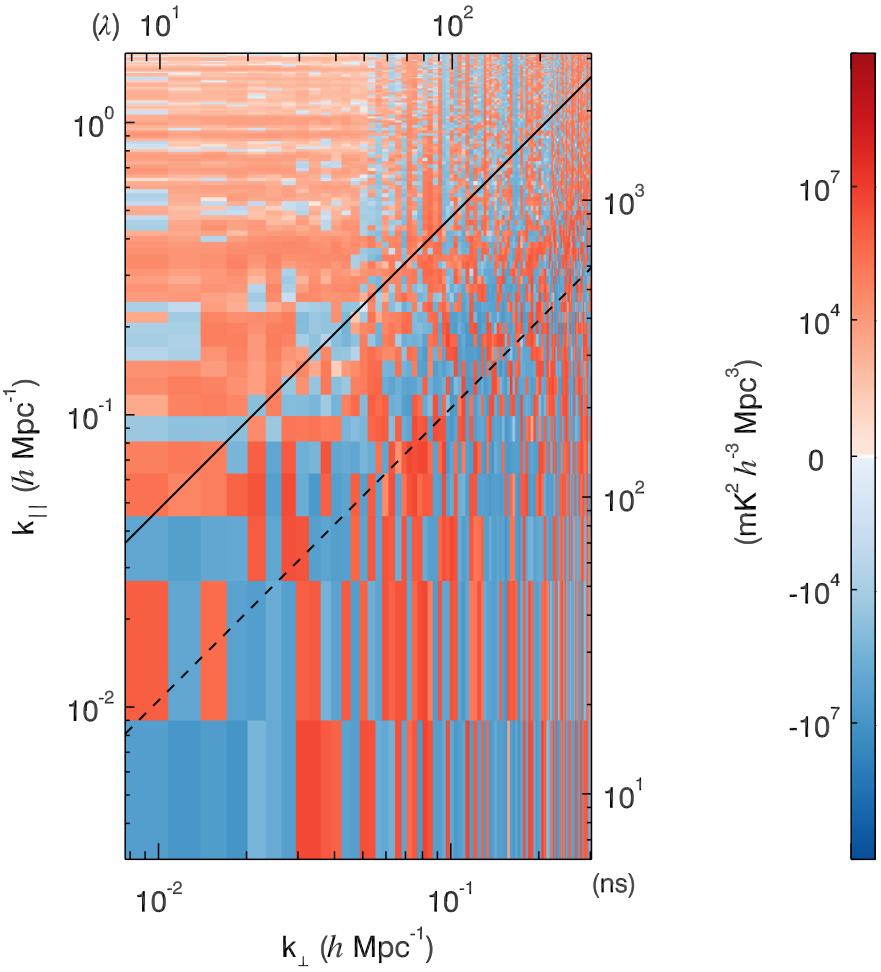


Figure 6.4: The difference 2D power spectrum between a reference 2D power spectrum and a 2D power spectrum with an averaged calibration. Red indicates an excess of power, and blue indicates a depression of power. Calibration parameters per frequency were averaged across all antennas and over 30 minutes of observations for maximal LST coverage, excluding outliers beyond 2σ. Excess power levels are much lower in the EoR window than with traditional calibration in Figure 6.1, but still higher than with smooth mitigation calibration techniques in Figure 6.2.

antenna and time averaging, in conjunction with the smooth characteristics of the antenna, to reduce the level of calibration contamination [2]. Other instruments may be able to achieve practical levels of spectral contamination with only averaged calibration solutions if more LST or antenna samples can be used.

In practice, thermal noise will also affect the power spectrum and the calibration solutions. An additional set of simulations explored the effect of thermal noise on the calibration solutions, and showed the expected additional spectral contamination in the EoR window. The thermal noise contribution is uncorrelated in time, and averaging calibration solutions night to night proves effective in removing this contribution (see [111] for detailed analysis). However, spectral contamination from faint, unmodeled sources still remains as a systematic error due to the limited number of antennas and observing LSTs. The systematic contribution of the calibration due to an imperfect calibration catalog has more potential than the thermal noise contribution in hindering long integrations in the search for the EoR signal.

6.4 Discussion

We explored the impact of instrumental calibration on EoR power spectrum measurements using a precision end-to-end simulation. Our simulation framework includes a precise frequency-dependent instrument model, a foreground model based on the observed compact sources in the MWA EoR0 field, and the full FHD and εppsilon calibration, imaging, and power spectrum estimation pipeline used in EoR analysis [2, 40]. We found that it was crucial to include a calibration catalog that does not exactly match the simulated sky and to propagate all of the calibration effects to the power spectrum where the EoR measurement will be performed.

In Figure 5.1, we showed that the EoR is perfectly recovered if the calibration model is identical to the foreground sources in the sky model. However, it is impossible to have a perfect calibration model—any catalog will miss faint sources and have small errors in the flux, position, and spatial morphology of the included sources. In §6.1, we simulated the effect of an imperfect calibration catalog by using only the brightest 4,000 sources for

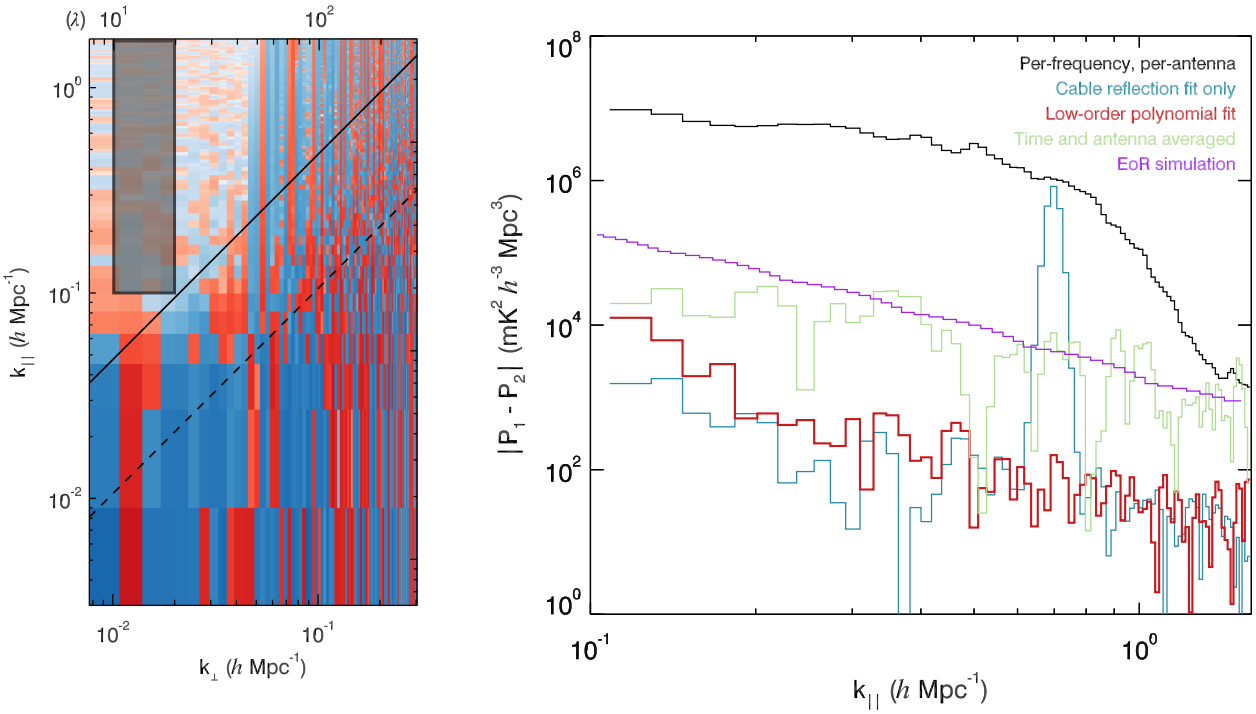


Figure 6.5: A 1D comparison of all power spectrum difference plots using the various calibration techniques detailed in this chapter (Figures 6.1–6.4) in the EoR window. We select the same $k_{||}$ and $k_{\perp}$ region for all power spectrum difference plots, shown for illustration on a copy of Figure 6.2 as a grey box. We then average $k_{\perp}$ within the box from 10 to 20 λ to generate 1D comparisons as a function of $k_{||}$. The plot on the right is the resultant 1D power spectra as a function of $k_{||}$, comparing the traditional per-frequency and per-antenna calibration (black, Fig. 6.1), the 150 m cable reflection fit (blue, Fig. 6.3), the low-order polynomial fit calibration (red, Fig. 6.2), and the maximally averaged calibration, in both LST and antennas (green, Fig. 6.4). The power spectrum of the estimated EoR signal (purple, Fig. 5.1) is also plotted to show where contamination will surpass the desired detection. A maximally averaged calibration at the current level of precision for the MWA is not a practical solution, given that we only explored one of many possible contamination sources, and contamination is already approximately at the level of the EoR signal. The low-order polynomial fit is by far the best solution if the instrument varies smoothly across the entire frequency band.

calibrating out of a total of 6,950 sources in the foreground sky simulation. Using a traditional per-frequency antenna calibration, we showed that the resulting power spectrum has contamination throughout the EoR window—precluding EoR observations.

In §6.1, we explored the source of this contamination, identifying the PSFs of the faint, unmodeled sources as the root cause. The chromatic PSFs of the unmodeled sources lead to small calibration errors on the order of 10^{-3} that couple to the bright foregrounds and distribute spectral contamination throughout the EoR window (Figure 6.1). In §6.2 & §6.3, we then explored more advanced calibration techniques that could mitigate this contamination.

The lesson for SKA, HERA, the MWA upgrade, the LOFAR upgrade, and other future EoR instruments is that any spectral features of the antennas that are calibrated will lead to contamination at the corresponding location in the EoR window. This is most clearly seen in Figure 6.3. This source of contamination via calibration errors places strong constraints on the instruments and planned observational programs. To avoid the contamination identified here, we identify four potential solutions:

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

(2). Use antennas with very smooth spectral responses. It is possible to design an analog and digital system that is naturally very spectrally smooth, which would not require calibration of features within the EoR window. The power spectrum of such a hypothetical antenna is shown in Figure 6.2. In practice, this means there can be no spectral features larger than $\sim 10^{-5}$ in the antenna or receiver system with spectral scales faster than $\sim 8\,\mathrm{MHz}$ ($125\,\mathrm{ns}$). This thinking is driving the spectrally smooth antenna and

receiver designs of HERA and the MWA upgrade [62, 25, 103, 78].

(3). Manufacture physically identical and stable antennas. If all the antennas have the same spectral features and their response is very stable in time, then the antenna–time averaging explored in §6.3 can be used to reduce the spectral smoothness and catalog precision requirements. While these techniques focus on sky-based calibration, redundant calibration methods explicitly depend on the antenna responses being identical [114, 48, 74, 65, 117].

(4). Develop an external calibration system. The coupling to an incomplete sky catalog can be entirely avoided by using an external calibrator such as a drone, satellite, pulsar, or pulse injection system. It is still a challenge to reach the $\sim 10^{-5}$ calibration precision needed, but several groups have been pursuing this path [63, 77, 61].

Which combinations of these four approaches will work the best is jointly dependent on the instrument-specific antenna PSFs and the precision and depth of the calibration catalog. Calibration simulations following the techniques developed in §5 & §6 must be explored for each instrument to calculate the necessary instrument specifications.

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

Spectrally smooth antennas described in approach (2) are the practical solution to avoiding spectral contamination of the EoR window. Figure 6.5 shows that it is the least contaminated calibration method in the sensitive EoR modes explored in this work. We recommend that EoR instruments aim to have no spectral features larger than $\sim 10^{-5}$ on scales faster

than $\sim$8 MHz (125 ns).

It has long been recognized that precision calibration is necessary to perform EoR power spectrum observations. We have identified that traditional per-frequency calibration techniques with an imperfect calibration catalog can lead to significant contamination of the EoR power spectrum window. This insight and the associated simulation techniques can help guide the design of the SKA and other future EoR instruments.

Chapter 7

SPECIALIZED CALIBRATION DEVELOPMENT

We must calibrate to unprecedented levels of precision for an EoR power spectrum measurement. Our goal is to reach the equivalent of 1 part in 10^5 in the calibration solutions, a requirement to detect the EoR with our current instrument [1]. In §4 & §6, we explored both experimentally and theoretically how to manage this, focusing on averaging many variables and axes in the bandpass. However, bandpass averaging may or may not achieve the necessary precision, as discussed in §6.

Now, we focus on more advanced and creative solutions. If we are to detect the EoR, every aspect of the instrument must be investigated—and understood. We explore the non-linearity of the instrument, and the effects that has on our solutions, mitigation techniques, and future improvements. In addition, we study the impact of improvements on the calibration phases through floating references and averaging. We also attempt to augment our understanding of the bandpass in an alternative space, examining potential avenues for the future.

7.1 Bit noise

The data from the MWA is digitized. It is converted into a limited number of bits, and there are specific amplitude regions where those bits can accurately capture the varying signal. The MWA attempts to handle oversaturation and undersaturation by applying digital gain jumps: multiplying data by integer values during conversion, and subsequently dividing out the quantized steps after processing.

In §4, experimentally determined bandpasses had errors that did not scale with the overall structure of the bandpass, in both the digital gain jump and in the edge channels. This nonlinearity indicated a bit noise, otherwise known as a quantization error. For the

digital gain jump, it caused contamination throughout sensitive EoR measurement regions. Whether or not a bandpass solution was successful depended heavily on the modeling of the digital gain jump. In contrast, bit noise on edge channels was not a drastic contribution, since it only contaminated the coarse band harmonic lines. Understanding how bit errors affect the instrument will help to constrain calibration methodologies and techniques.

7.1.1 Coarse band edges

The two-stage PFB creates characteristic coarse band shapes every 1.28 MHz in the MWA (see §2.2.3). We regularly flag, or remove, 80 kHz at each coarse band edge to avoid contamination due to aliasing created during fine channelization. This contributes to our typical bandpass features in §4.1. *All* bandpass solutions shown throughout this work already have a theoretical coarse band shape removed by pre-pipeline packages. However, simulations of the particular coarse band shape do not accurately reflect the actual gain, and thus it still affects all our bandpass solutions.

There may be missing components to the theoretical coarse band shape that could account for the difference. Nevertheless, there is also a unique effect after the digital gain jump that cannot be captured by simple additions to the coarse band simulation. The coarse band shape changes, especially near the edge channels.

Figure 7.1 shows cable-averaged cross-correlation and auto-correlation calibration solutions for the zenith pointing of August 23, 2013. Each coarse band has a separate normalization to remove large-scale bandpass effects, but some smaller-scale ripples remain, such as cable reflections. The coarse bands at frequencies below the digital gain jump all appear to be consistent; they all seem to follow a typical shape. However, the coarse band edges *above* the digital gain jump vary across cable types.

Above the digital gain jump, the cross-correlation calibration solutions have a flatter coarse band shape that appears to depend weakly on cable type. In contrast, the auto-correlation calibration solutions have heightened coarse band edges in that region. This indicates that there is an increase in correlated noise in the coarse band edges. The auto-

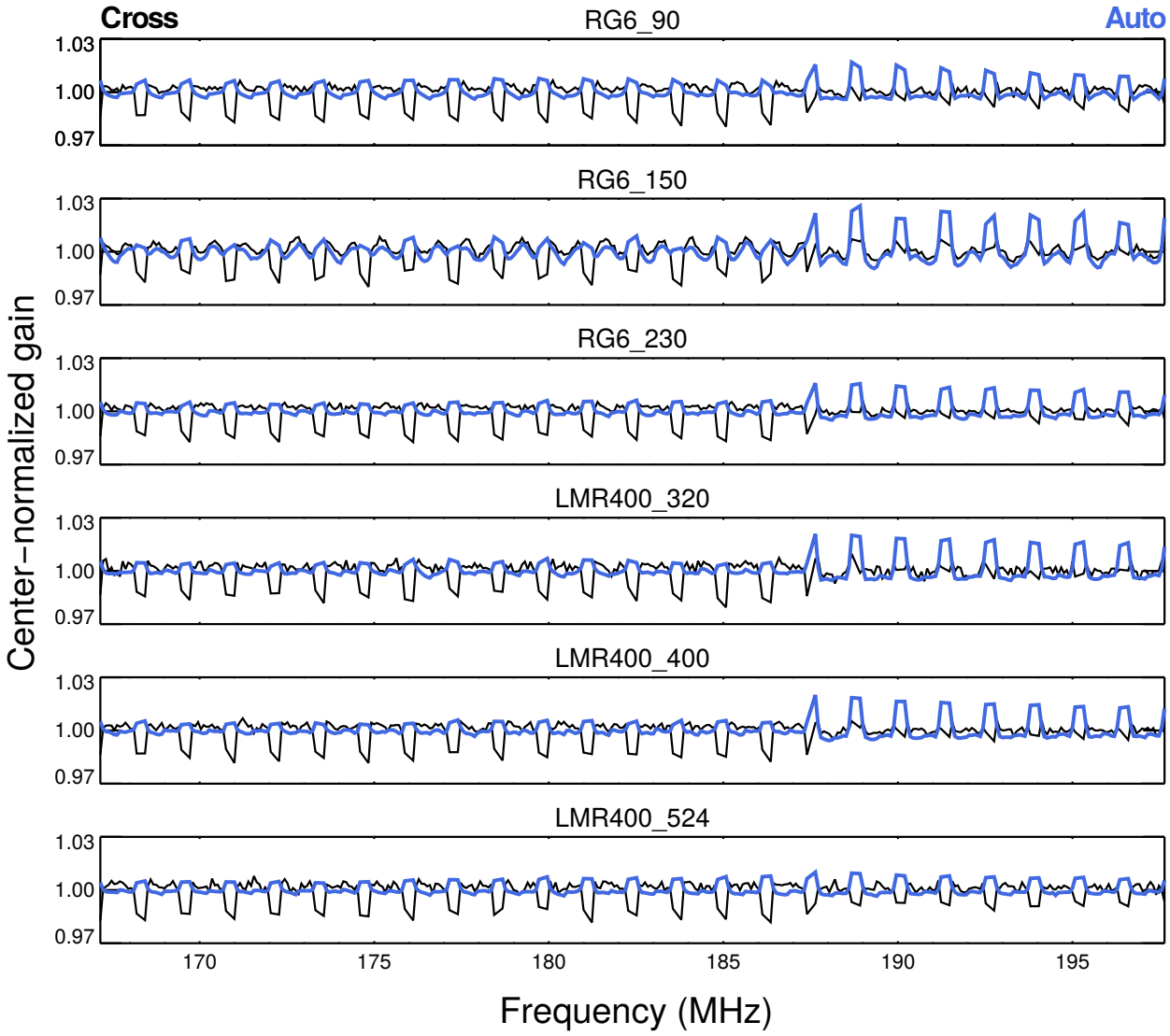


Figure 7.1: Cable-averaged calibration solutions from the cross-correlations (black) and the auto-correlations (blue) for August 23, 2013. Each coarse channel has been normalized individually. The coarse channels at frequencies above the digital gain jump (at about 187.5 MHz) have a different shape that appears to be weakly correlated with cable type. This shape is different between the auto-correlation and the cross-correlation solutions, and is indicative of a change in the bit statistics after the digital gain jump.

correlations square the noise, whereas cross-correlations will have zero contribution from uncorrelated noise. A change in the bit statistics might force a Gaussian data set into a truncated Gaussian, which may be correlated across the array.

In investigating this effect, we found that it evolved over Season 1. The gain in the coarse channel edges at frequencies above the digital gain jump fluctuated. This negatively affects our time-averaging schemes. Upon further inspection, it appeared correlated with the seasonal weather.

We plotted the average edge channel gain in the cross-correlation solutions against the average temperature recorded by the beamformer (essentially a measure of ambient temperature at the beamformer's location). Figure 7.2 shows this relation for each of the cable types and for a variety of days in Season 1. There is a strong correlation with edge depth after the digital gain jump and the temperature at the time of the observation, where lower edge depths happened on hotter days.

The amplifiers in the beamformer are more efficient at lower temperatures, and thus will yield higher gains. The hardware at the receiver then applies a digital gain jump to avoid undersaturation. However, there appears to be some bit truncation. The more gain that originates from the beamformer, the more unlike the rest of the band the edge depth is, indicative of bit error. This suggests that the digital gain jump overcompensates, causing oversaturation. This is especially evident on the coarse band edges—we divide by our gains, so a low calibration edge depth is correcting for high raw amplitude.

Temperature-dependent bit errors only seem to occur in 2013 data, at frequencies above the digital gain jump. Unfortunately, none of our bandpass calibration schemes can account for this behavior. An attempt can be made to fit the temperature-dependence and create an estimated edge depth. However, due to this error and difficulty in fitting the digital gain jump, we choose to avoid using this data in calculating EoR limits. This is a drastic limitation—we reduce our 2013 data by one third. Initial processing of 2014 data indicates that this issue may have been largely mitigated in later years by fine-tuning of the digital gain jump, which bodes well for longer integrations.

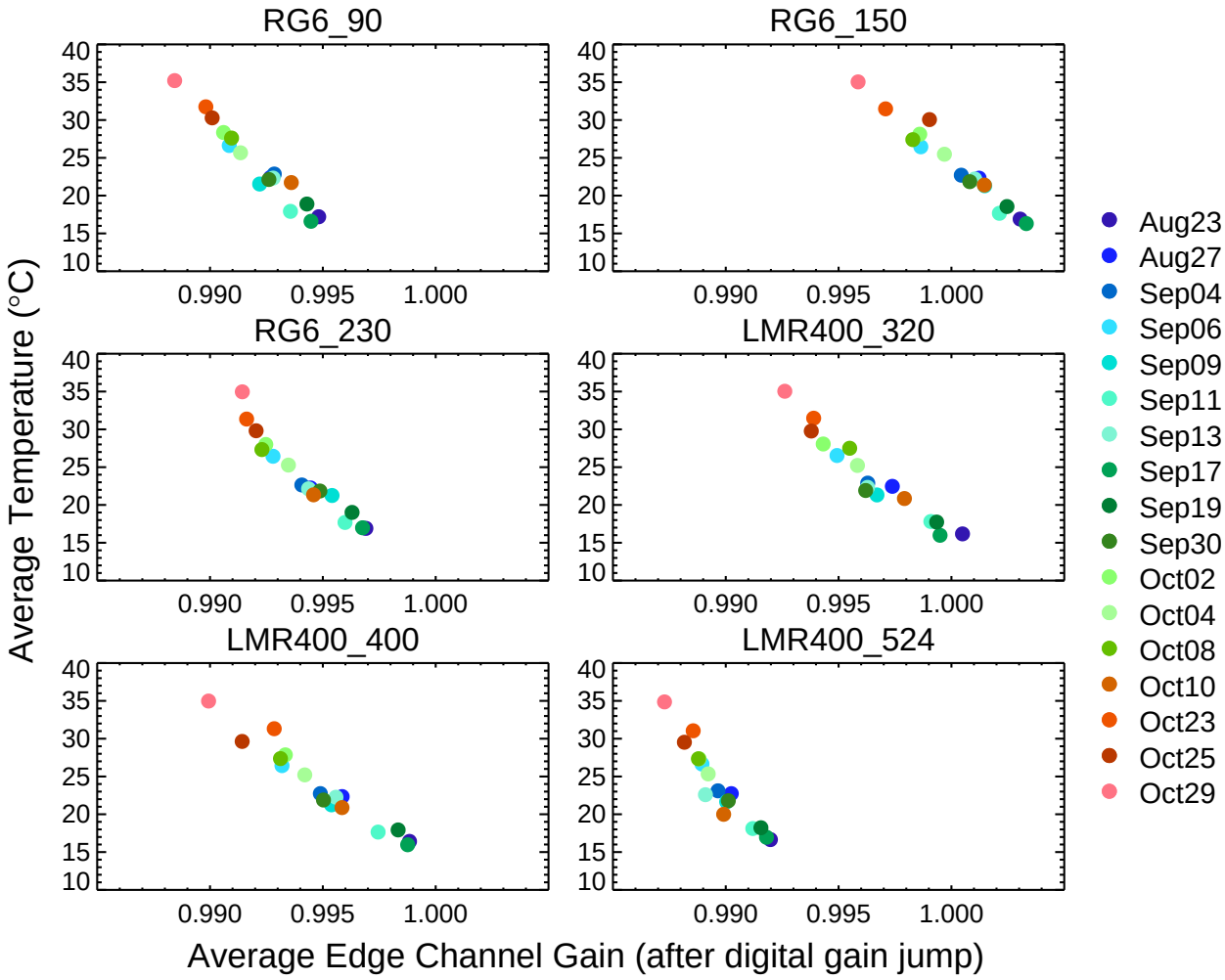


Figure 7.2: Temperature correlation of the average edge channel gain for a subset of the band for Season 1. Hotter days have lower edge channel gain, which aligns with the rest of the coarse band shapes in Figure 7.1. Colder days have higher bit noise, indicating oversaturation caused by the digital gain jump.

7.1.2 Digital gain jump

The digital gain jump was responsible for oversaturation of the top one third of the frequency band for cold, high-gain days in 2013 data. Given how strongly the edge depth correlated with temperature, it was apparent that we needed to investigate the digital gain jump itself for the same effect.

In addition, inaccurately accounting for the digital gain jump is one of the largest sources of contamination in the bandpass, as seen in §4. Averaging the calibration solutions by cable type lowered the contamination, but our best solution was to allow for independent fitting of the digital gain jump. This is consistent with the behavior of the coarse band edges: weakly dependent on cable type, but mostly independent of the averaging schemes.

Figure 7.3 shows the cable-averaged digital gain jump as a function of temperature for a variety of days in Season 1. There is a strong correlation, and it is extremely similar to the dependence for the edge channels in Figure 7.2. This is not surprising; the digital gain jump happens between the edges of the 16^{th} and 17^{th} coarse bands, and therefore is subject to the same effects that dominate the edge channels.

Temperature dependence of the digital gain jump is another reason to exclude the high-frequency portion of the band from EoR power spectrum limits. While we try to account for the variation by using an independent fit in §4, using a free parameter for the correction is almost certainly not at the precision level needed to detect the EoR. Again, fine-tuning of the digital gain jump in later years might have alleviated this issue.

7.2 Improving phases

We sought to reduce spectral structure from unmodeled sources in the gains by averaging the calibration amplitudes, detailed in §4 & §6. That naturally leads to another potential area of improvement: the phases of the calibration solutions. Up until now, we have essentially ignored the possible contamination from the phases.

In Equation 3.9, we fit the phases with a linear function across frequency. This is a decent

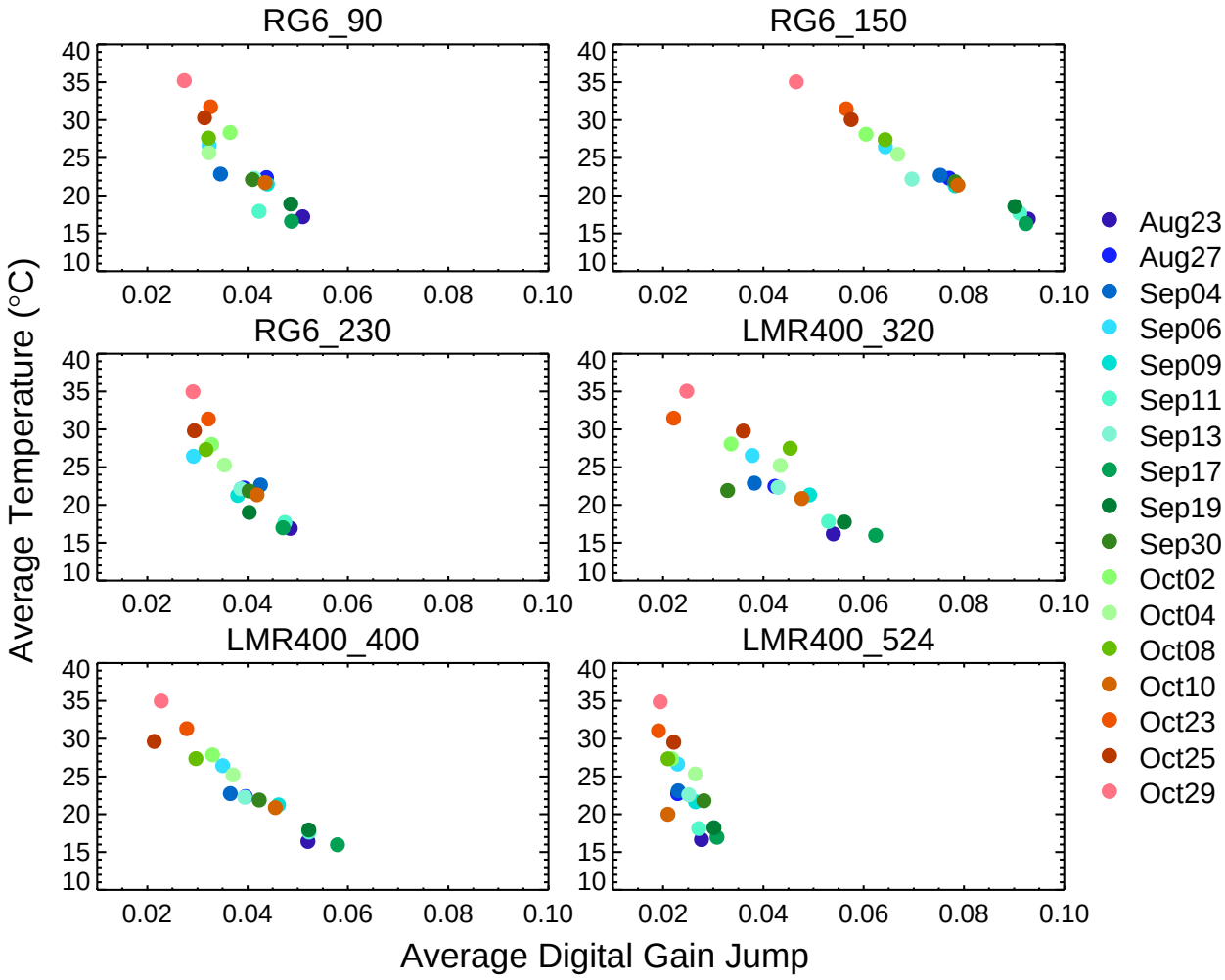


Figure 7.3: Temperature correlation of the average digital gain jump for a variety of days in Season 1 in the zenith pointing. Hotter days have a smaller gain jump, while colder days have a higher gain jump. The same bit errors that affect the edge channels are affecting the digital gain jump.

approximation; the phases are very linear. However, can we improve the power spectrum by averaging together the phases to reduce both noise and spectral contamination, while capturing real structure?

In order to average the phases, we must find a stable reference. There is a degeneracy in the phase calculations that depends on the definition of the initial time of an incoming wave's interaction with the instrument. Normally, we reference all phases to a particular well-behaved tile's phase. If the phase of the reference tile changes, so do all the other phases across the array. While the phases are very linear, they do change from day to day. A changing reference makes averaging impractical.

To counteract this, we allowed the reference to float. We didn't reference the tiles when we calculated their phases. All phases floated in unison, and there was an apparent background reference that could be extracted. This varies drastically day-to-day, and has no physical meaning. We averaged all tiles and observations within a day, and referenced all tiles to that average.

We are left with only fine-scale structure from the phase, usually close to zero mean. This is repeated for many days in Season 1, and all fine-scale structure is then averaged across time for each tile. The fine-scale structure can then be added onto linear function fits of the phases from any reference.

An example of the result is shown in Figure 7.4. Regular unfit phases are plotted in grey with their respective averaged phases in color. Immediately, it is obvious that not much structure is present, neither in the unfit nor the averaged phases. Small ripples can be seen in the averaged phases which may be attributable to cable reflection features. Overall, however, there is not much structure in the phases to extract via an averaging scheme.

We test the average calibration phases in power spectrum space to investigate their contribution in Figure 7.5. Following §4, we integrate over 64 observations from August 23, 2013 and compare to reference power spectra. We keep the calibration amplitude schemes the same between all power spectra, so the only differences should result from changes in the phase calibration.

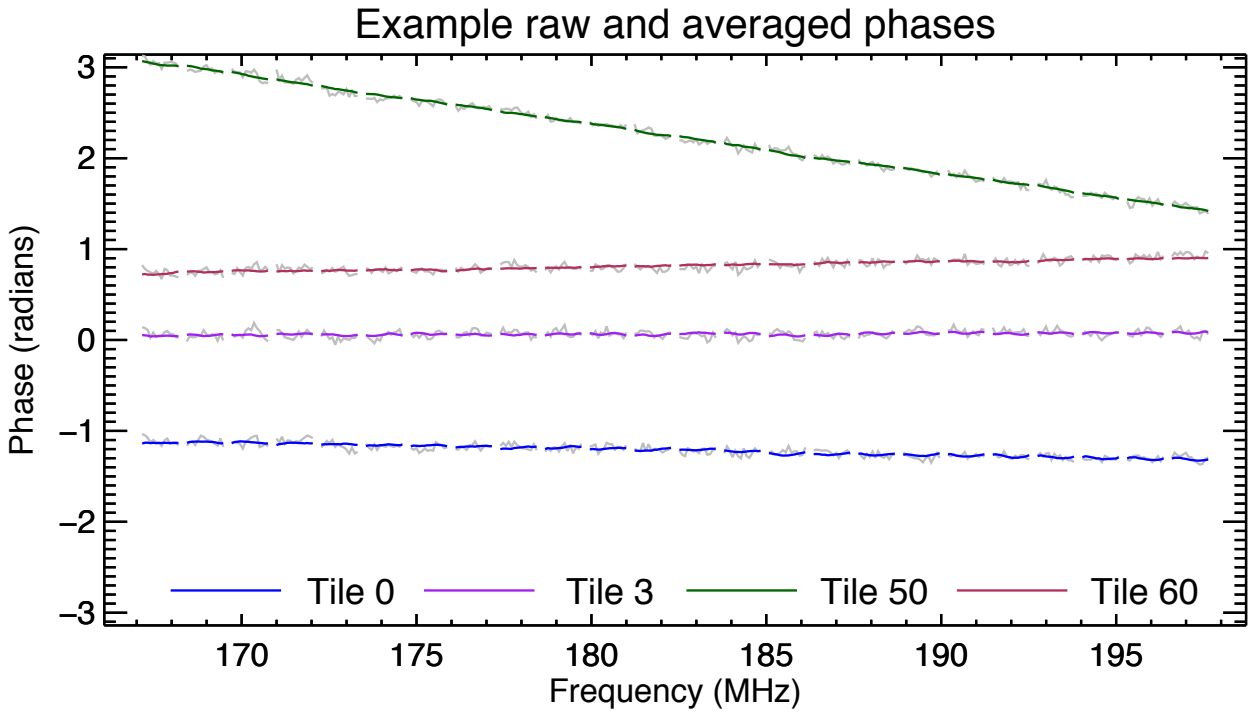


Figure 7.4: An example of the phases for a subset of tiles. The unfit, raw phases are plotted in grey, and the fine-scale-averaged phases are plotted in color.

First, we examine the contributions from the floating reference. The left column of Figure 7.5 is the difference between power spectra from two calibration schemes: one with a floating phase reference and the other with a constrained phase reference. The colors are generally light; the differences are below $10^8\,\mathrm{mK}^2\,h^{-3}\,\mathrm{Mpc}^3$. However, there does appear to be a small amount of structure resulting from a floating reference.

Second, we examine the contributions from the addition of fine-scale averages. The right column of Figure 7.5 is the difference between adding fine-scale structure and no extra additions, both with a linear fit with typical phase reference. There is significant benefit in the mode associated with the 90 m cable, which indicates that the reflection is better accounted for than with the linear fit. There is also similar benefit in the 230 m and 320 m reflection modes, but only at high $k_\perp$. In fact, the transition occurs at approximately $50\,\lambda$—

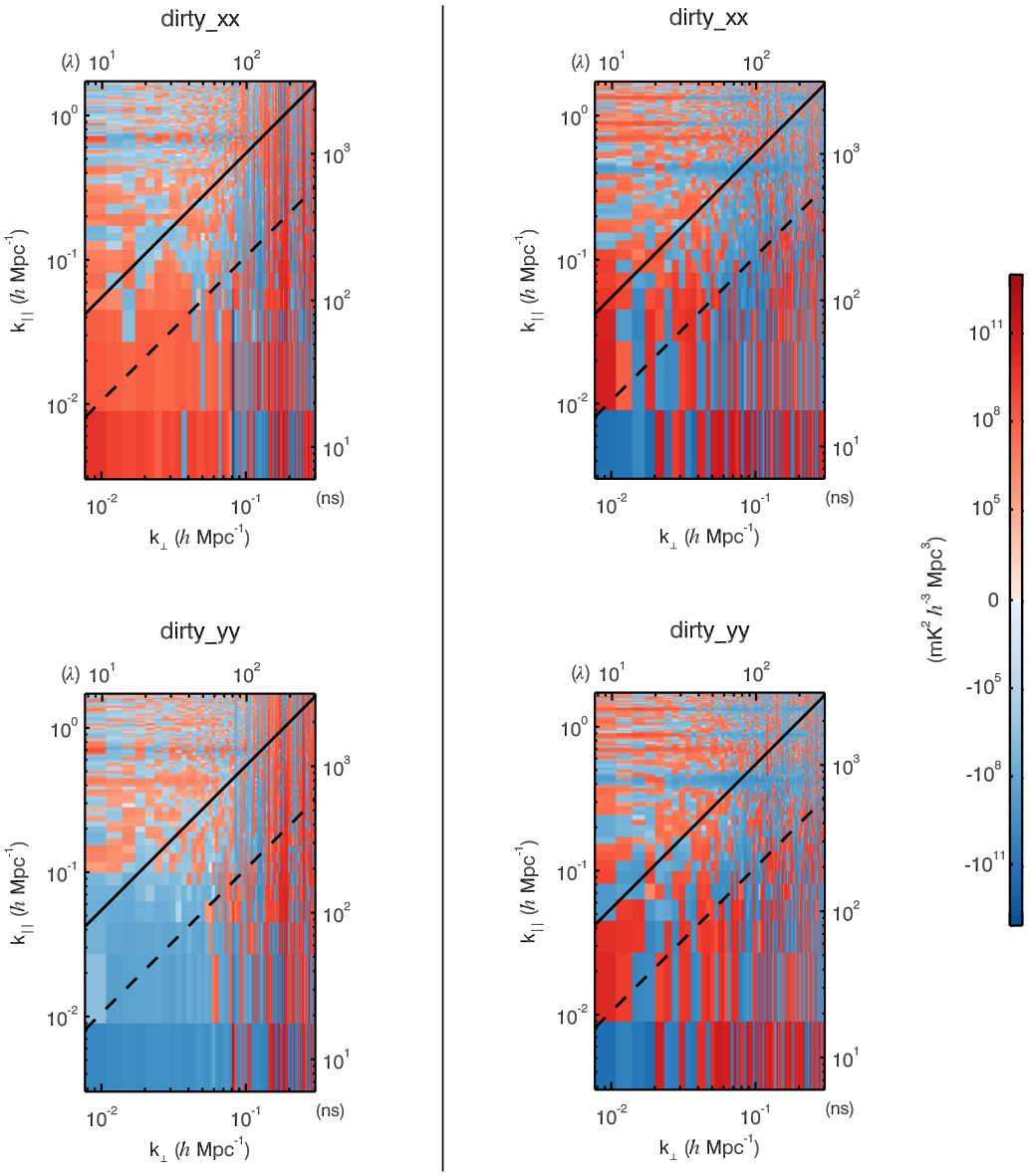


Figure 7.5: 2D power spectrum differences comparing two integrations of 64 observations from August 23, 2013. The floating-phase reference is compared against the typical constrained-phase reference (left column) and the fine-scale phase average is compared against a linear fit (right column). Fine-scale averaging of the phases improves the cable reflections where the calibration is constrained (shown in blue), but this improvement is small and in a high-contamination region.

the smallest baseline used in calibration. These modes are unconstrained and thus have higher noise [76], which highlights the need for a calibration model on all scales.

Unfortunately, phase averaging has very little practical benefit at this time. The improvements in the power spectrum are in regions of low interest or high contamination. When reliable diffuse catalogs become available, more of the EoR window at higher $k_{||}$ will become usable with phase averaging.

7.3 Calibration delay space

In §4, we extensively explored reducing contamination and noise on the bandpass by averaging in various ways. The more contributions to the average, the smoother the resulting bandpass. In general, this led to better power spectra. While averaging is a reliable way to lessen noise, it requires stability and large quantities of data. Therefore, we attempt to smooth the bandpass through alternative means.

We Fourier transform each bandpass calibration solution along frequency, creating solutions as a function of time. This is essentially a spectra of modes, and thus we call it **calibration delay space**. If we are going to work within this space, we have to understand the typical bandpass features described in §4.1:

Regular flagging: A set of flagged channels at regular spacing creates a Dirac comb, and the Fourier transform of a Dirac comb, is, in fact, a Dirac comb. Therefore, we have regular spikes in calibration delay space whose height depends on the DC mode of the bandpass.

Coarse bandpass shape: The characteristic shape of the coarse bandpass repeats on the same frequency scale as the flagging; thus, it broadens the regular flagging modes.

Large-scale ripple and attenuation: Large frequency modes will result in short delay modes. There will be strong response on short delays since there are many large-scale ripples in the bandpass.

Digital gain jump: A sharp feature in one space will be a wide feature in the other. This poses a problem for delay space; there will be contamination on all scales if the digital gain jump is not removed beforehand.

Cable reflections: Since reflections generate an oscillation dependent on the travel time within the cable, there will be a strong mode at the associated delay.

In addition, these features are no longer independent. When we take the Fourier transform, each of these features are convolved with one another. For example, the cable reflection is no longer an isolated mode; convolution with the regular flagging creates signal on the cadence of the comb, away from the original reflection mode. This is particularly harsh for the digital gain jump since signal is expected to spread everywhere during its transform.

We can remove certain features that do not need to be investigated further. First of all, we can remove the higher portion of the band after digital gain jump before transforming to delay space. This will remove its ringing response (see §4). Next, we can remove the overall amplitude to create bandpasses which are zero mean. This reduces extra signal from the convolution of the DC mode, as well as reduces the effect of the regular flagging. Lastly, we can remove large-scale ripples through low-order polynomial fits, which will not affect the EoR window and are thus of little concern.

Figure 7.6 shows this result in calibration delay space for all Season 1 bandpasses for a particular tile as a color histogram. The real and imaginary components have been split (top left and top right panels). Large-scale modes in frequency space are at small delays, and while most of the comb has been removed, there still remains a residual effect. A cable reflection can be seen at about 1025 ns since the example tile has a 150 m beamformer-to-receiver cable. The histogram shows strong stability across Season 1 since the distribution is Gaussian-like for each delay.

The by-tile time average calculated in §4.4.3 for the zenith pointing is overlayed on the histogram in black. The notable features in the histogram are also present in the mean. However, it should be noted that there doesn't appear to be any particular detections of

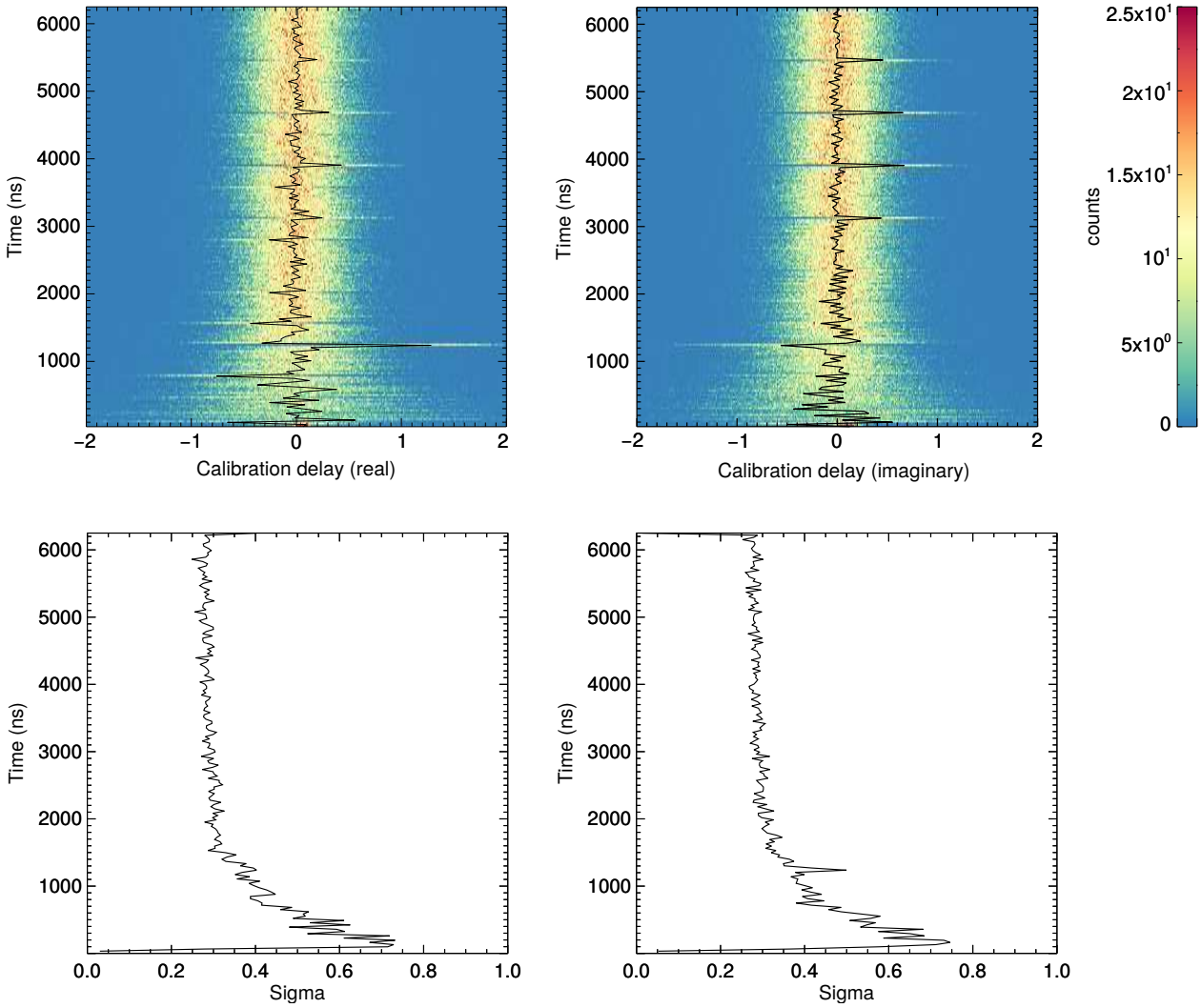


Figure 7.6: Calibration delay space for bandpasses in Season 1 for Tile 4. Histograms (top row) of all Season 1 bandpasses have had their digital gain jump and mean removed before being Fourier-transformed. Large-scale frequency structure is at low delay (labeled as time) and fine-scale structure is at high delay. The by-tile time average from §4.4.3 is overlayed in black. The square root of the variance (sigma) as a function of mode for Season 1 bandpasses (bottom row) show little variation at fine-scale frequency modes, and increased variance at large-scale frequency modes. Real and imaginary components of the transform have been split left and right, respectively.

fine-scale structure modes, other than residual regular flagging peaks.

We can also generate statistics of calibration delay space for Season 1. Again, the real and imaginary components have been split (bottom left and bottom right panels). In general, the large-scale frequency ripples have high variance, while fine-scale frequency structures have low variance. The cable reflection mode is only slightly higher variance compared to the surrounding modes in the imaginary component. A lack of structure indicates that all the fine-scale frequency modes are dominated by noise in the calibration solutions.

Given the lack of definite fine-scale frequency modes, we can attempt to smooth out the bandpass by zeroing those modes. However, we must retain the convolved components in the fine-scale structure. Otherwise, there will be false structure when we inverse Fourier transform back into frequency space.

We keep all structure below the first coarse band in the average bandpass, thereby labeling it as true. We then remove all structure above the first coarse band while retaining modes associated with the regular flagging and cable reflections. We calculate all convolutions that occur due to the remaining modes, and add those back to our zeroed fine-scale modes. This solution is then inverse Fourier-transformed back into frequency space as our new by-tile time-average bandpass, and the result is much smoother.

Sadly, this does not translate well into power spectrum space. Figure 7.7 shows the 2D power spectrum differences between implementing this new bandpass and the global bandpass with a fit to the digital gain jump and the cable reflection (Equation 4.2 with Equation 4.3). The left column compares the full band while the right column compares the frequency band below the digital gain jump. Both differences show a large amount of excess contamination (red) in the EoR window. This may indicate that there are *true* modes in the noise beyond the first coarse band harmonic, and that zeroing them has a detrimental effect. In addition, it could also be the case that the convolved components calculated from the large-scale modes were wrong; if the modes themselves were inaccurate, this would propagate with the convolution. In either case, we do not know enough about our instrument at this point to artificially remove modes from the bandpass.

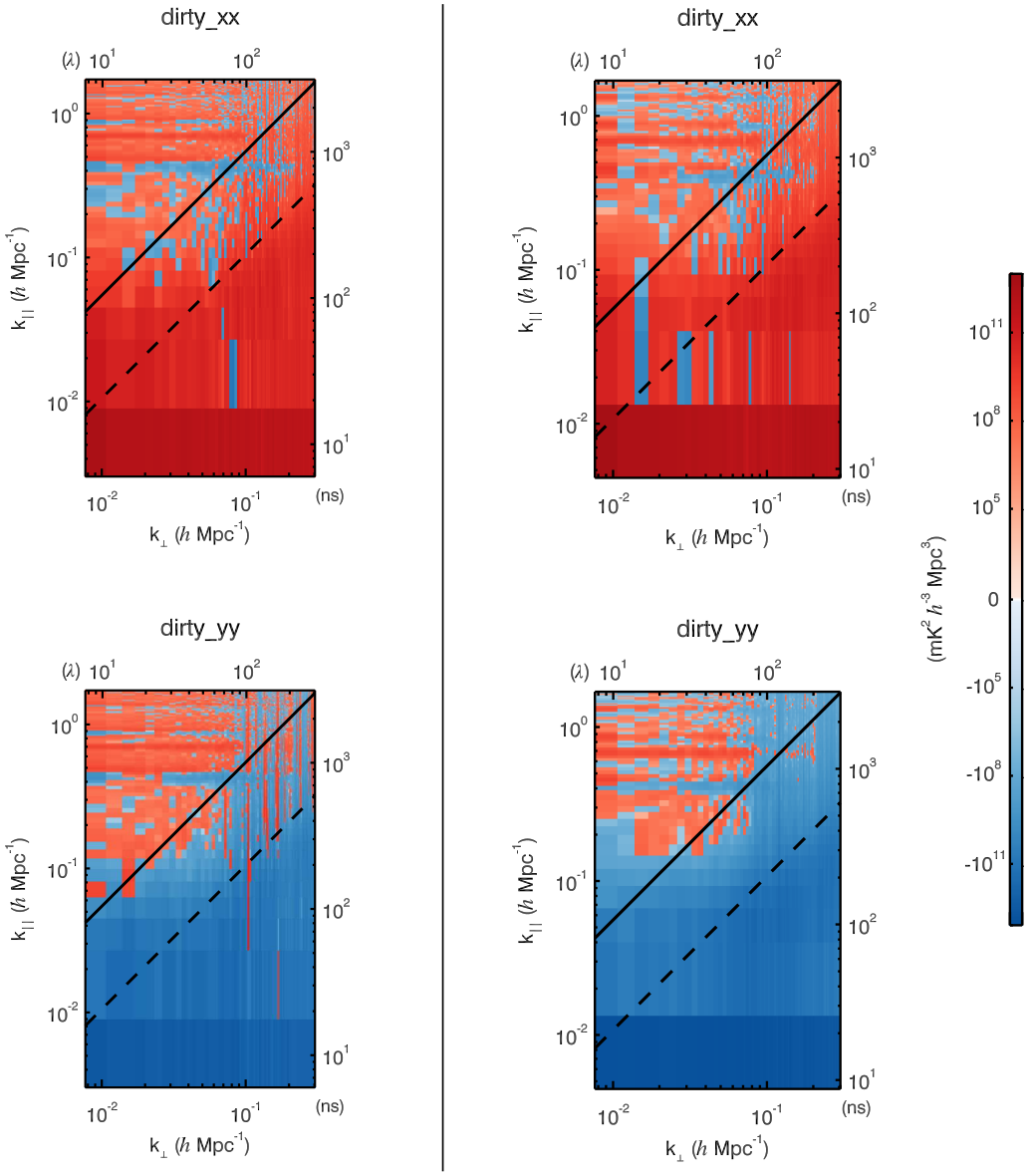


Figure 7.7: 2D power spectrum difference comparing two integrations of 64 observations from August 23, 2013. The bandpass smoothed via calibration delay is compared against the global bandpass with fits to the digital gain jump and the cable reflection. Both the full band (left column) and band before the digital gain jump (right column) are shown. Artificially removing modes from the bandpass has a detrimental effect.

Chapter 8

SIGNIFICANT PROGRESS ON PRECISION TECHNIQUES

In the previous chapters, we have almost exclusively focused on calibration effects on the EoR power spectrum measurement. By understanding the causes and consequences of calibration contamination, we now understand a variety of unrelated errors with similar repercussions.

In order to reduce contamination in the EoR window, we examine issues related to the beam kernel. There are two particular problems that we have found—both are negligible effects until the Fourier transform is taken. We present some solutions to mitigate the problems, but there are indications that a new degridding approach may need to be implemented in the future.

Application of anti-aliasing filters in key transformations has also been vital in efforts to reduce contamination. Significant reduction of aliased effects will not only reduce contamination for EoR power spectrum measurements, but will increase scientific output for future studies in uv-space.

We revisit a common and well-studied problem in EoR science: quality assurance of data. Our new method to determine data condition will be used to explore faint, abnormal signals from the sky and from the instrument. All of the new lines of inquiry examined in this chapter have been important in moving towards an updated limit, and opening up new opportunities for future science.

8.1 Quality statistics via visibility differences

A crucial part of precision measurements is the quality of the data. In the case of the EoR, one poor observation can be the dominant source of systematic error for an integration of

over one thousand observations. Improving our ability to choose only the best observations for creating limits is a constant area of research.

We currently have in place all of the quality assurance selections that were used to create previous limits from the FHD+εppsilon pipeline [5]. This includes:

Pointing cuts: Retaining observations only from pointings near zenith (two pointings before zenith to two pointings after zenith) due to a decrease in beam modeling quality far from zenith (see §2.2.2) and an increase in foreground contamination from the galactic plane in our sidelobes.

Delay spectrum window power: Removing outliers that have excess integrated window power, calculated from an uncalibrated delay spectrum.

Delay spectrum instrumental power ratio: Removing outliers with nonstandard ratios between the instrumental polarizations of the integrated window powers, calculated from an uncalibrated delay spectrum.

High flux density RMS: Discarding snapshots that have a high residual image RMS at sources greater than 0.5 Jy.

Calibration bandpass: Visual inspection of remaining observations' bandpass solutions to remove irregularities.

In addition to this list, we can also include selection cuts on ionospheric activity metrics to remove nights where our foreground removal is expected to be less successful [41].

We have a wide variety of metrics which help to construct a clean data set. However, we continue to pursue new ways to visualize data and data quality in order to create the best data set possible, as well as to learn more about the ways our data can fail.

We introduce a novel way to look for non-thermal signatures, inspired by work detailed in [38]. We generate residual visibilities for all frequencies, baselines, and observations within our Season 1 data set, and bin them in a histogram as a function of amplitude in Figure 8.1.

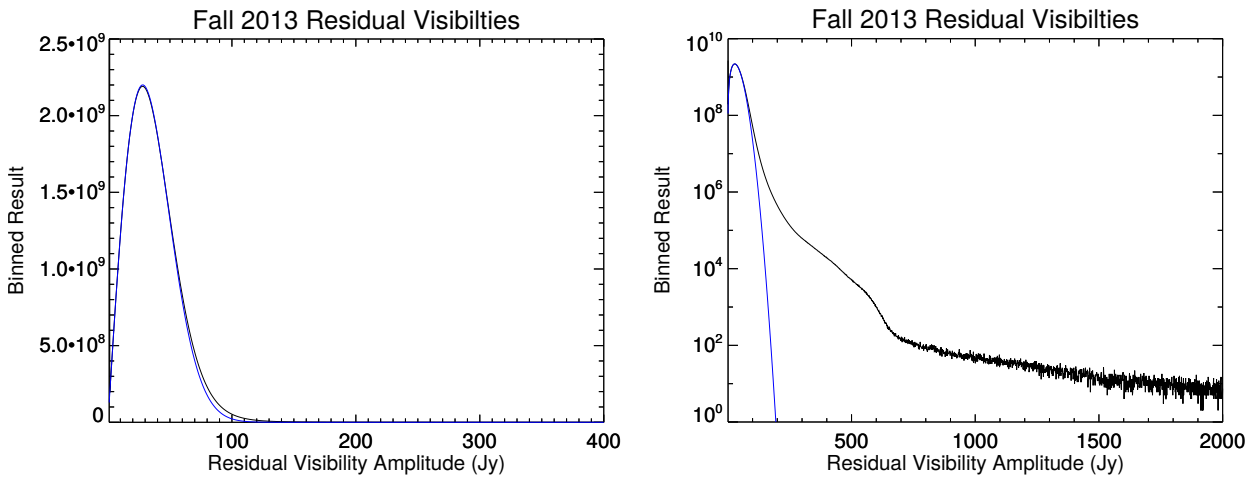


Figure 8.1: Histogram of residual visibilities for Season 1 with an associated Rayleigh fit in blue, plotted in linear-linear space on the left and semi-log space on the right. There is significant deviation at large amplitude in our Season 1 data set.

If a sufficient number of the point sources have been removed, a thermal signature should remain. The expected distribution for the amplitude of complex Gaussian errors is a Rayleigh distribution.

The left panel is the Season 1 histogram in linear-linear space, compared to a Rayleigh fit in blue. There is an extremely large sample of data points: over 150 billion! A large portion of the data is well-fit by the Rayleigh distribution. The right panel is the same data and fit, but plotted in semi-log space. Now, it is evident there is a significant deviation from the fit at large amplitude.

This is most likely *not* entirely an indication of abnormal data. We do not model diffuse structure, and that may have a significant contribution. In addition, the expected noise variance changes as a function of frequency, which requires that a rigorous best-fit be a superposition of separate Rayleigh distributions.

We can further examine the data with this new visualization. The sky signal can be effectively removed by performing differences of sequential two-second time steps. Time steps

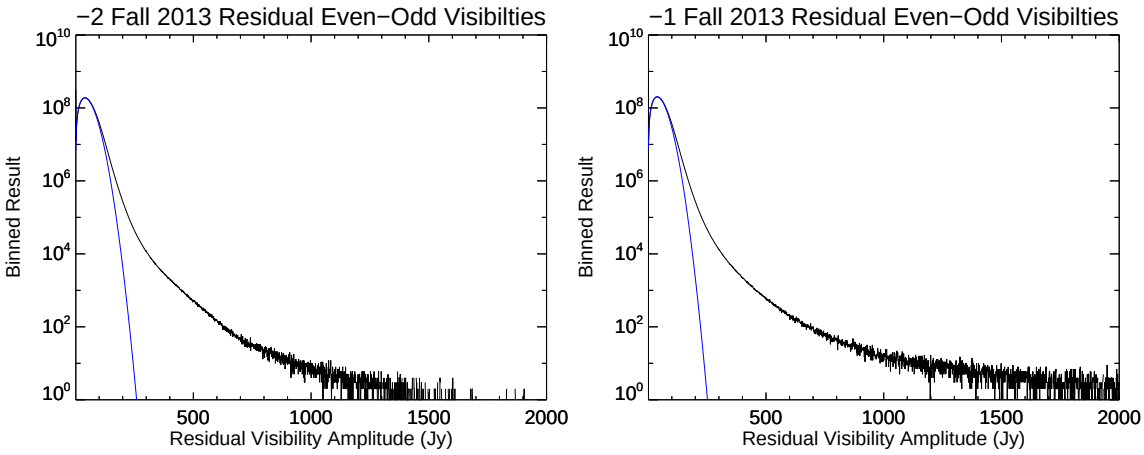
that are odd in the observations are subtracted from time steps which are even, following previous methods to remove the sky to probe the underlying noise signature in power spectrum space (see §3.3.2). This helps to remove unsubtracted foregrounds from our distribution.

Figure 8.2 shows the residual visibilities for instrumental XX where the sky has been removed though interleaved time-step differences, as a function of pointing. Zenith and the two pointings after zenith appear well-behaved, especially since the associated Rayleigh fits do not account for a frequency-varying noise signature. However, the pointings *before* zenith have high-amplitude outliers which do not follow a Rayleigh distribution.

In order to investigate this potential contamination, we can decompose the histograms in Figure 8.2 to reveal their functionality. We take all outliers beyond 300 Jy from each pointing, and histogram them as a function of frequency in Figure 8.3. Large spikes that occur at regular intervals are due to the DC channel of the polyphase filter banks—we flag the DC channel before frequency averaging because it does not contain meaningful data [88]. Thus, the center frequency of each coarse band has only half the data compared to all other frequencies, and is therefore noisier. This is a significant contributor to the outlier tails in all pointings. Interestingly, this is the only outlier feature in most analyzed pointings.

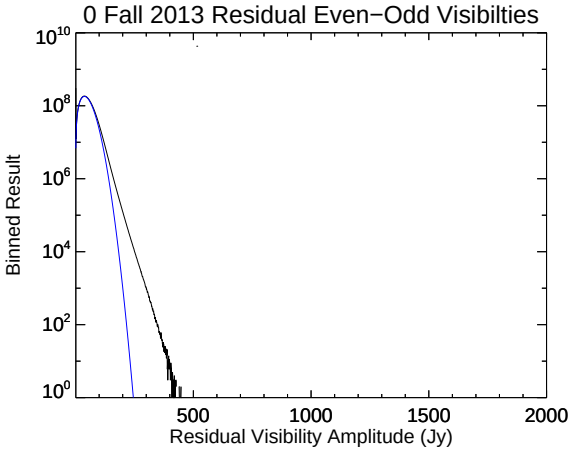
The first pointing of the night that is included in our clean data set has another significant feature towards higher frequencies. It is conceivable that this feature could be a response from the galactic plane increasing the noise threshold of our observations as a function of frequency. Indeed, this contamination seems to become more prevalent in even earlier pointings in power spectrum space [5], confirming a possible connection to the setting of the galactic plane.

In previous limits created using our pipeline, the earliest pointing (two pointings before zenith) was determined to be high-quality data. However, it is evident with our visibility statistics visualization that this pointing is noisier and thus lower quality compared to other included pointings. In the future, we may need to remove this pointing from our clean data integration, in order to lower systematics. We have not reached this level of required precision yet; we are dominated by other systematics.

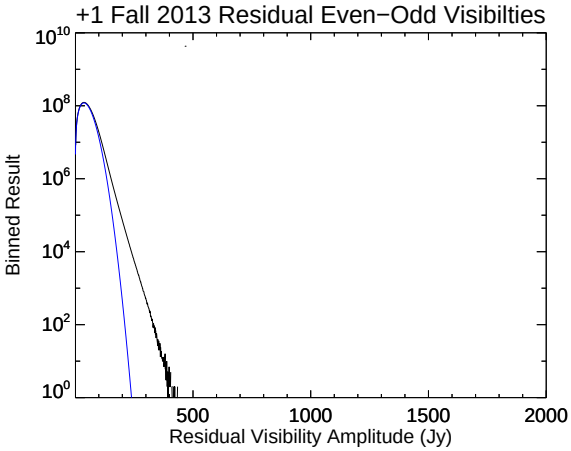


(a)

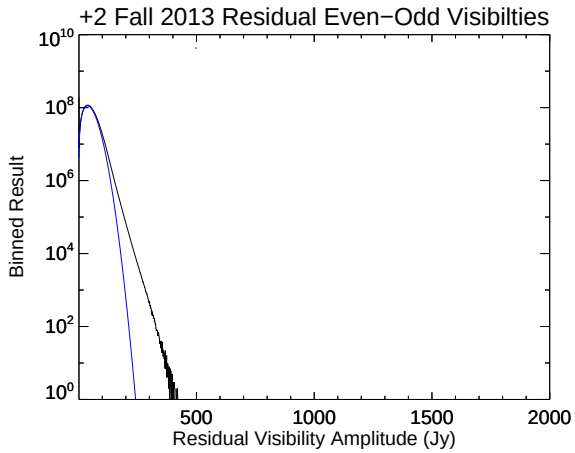
(b)



(c)



(d)



(e)

Figure 8.2: Histograms of residual visibilities over Season 1, where the even and odd time samples have been subtracted to remove sky signal. A χ^2 best-fit Rayleigh distribution has been added to each distribution in blue. Significant deviations occur in the pointings before zenith (8.2a and 8.2b).

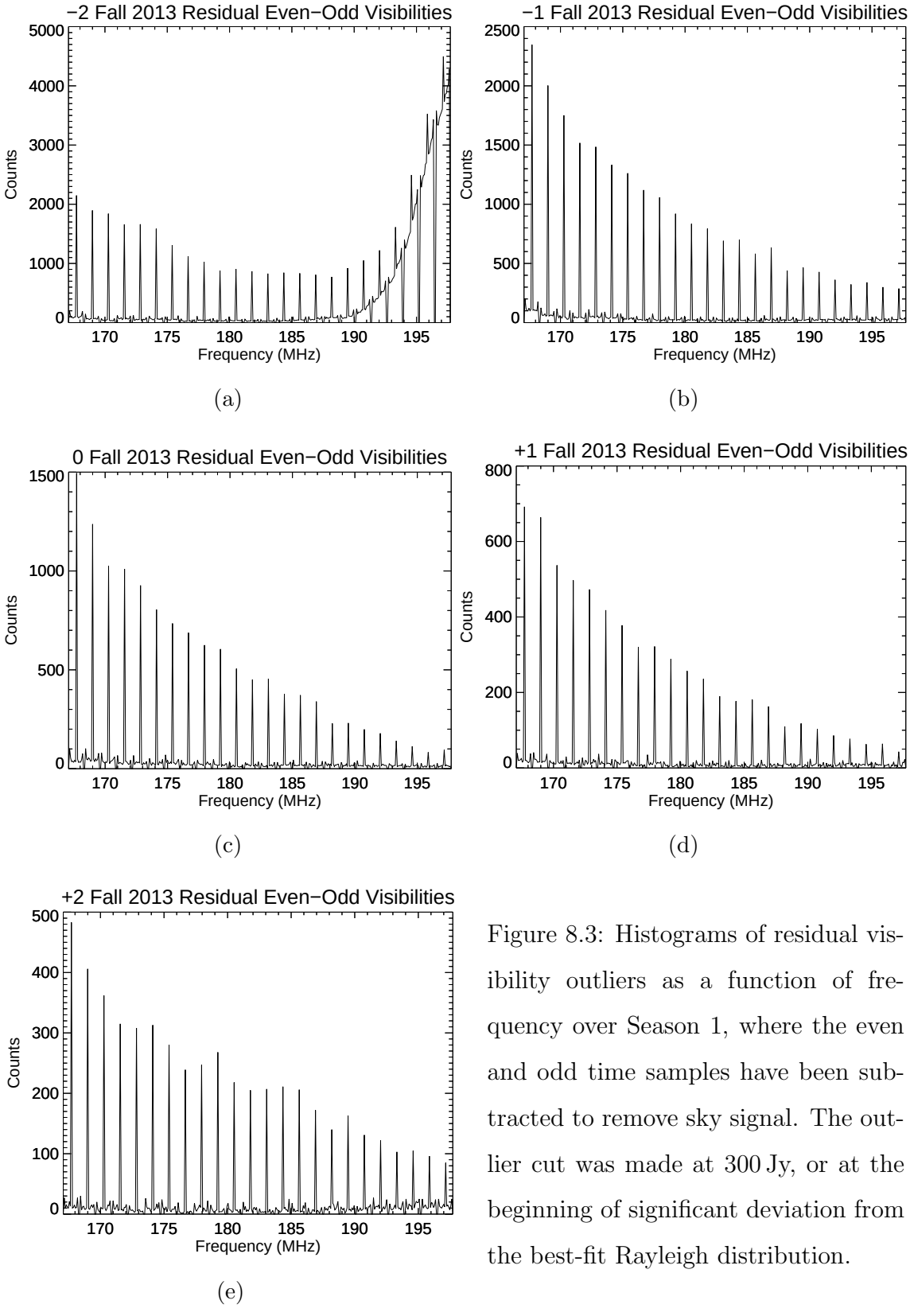


Figure 8.3: Histograms of residual visibility outliers as a function of frequency over Season 1, where the even and odd time samples have been subtracted to remove sky signal. The outlier cut was made at 300 Jy, or at the beginning of significant deviation from the best-fit Rayleigh distribution.

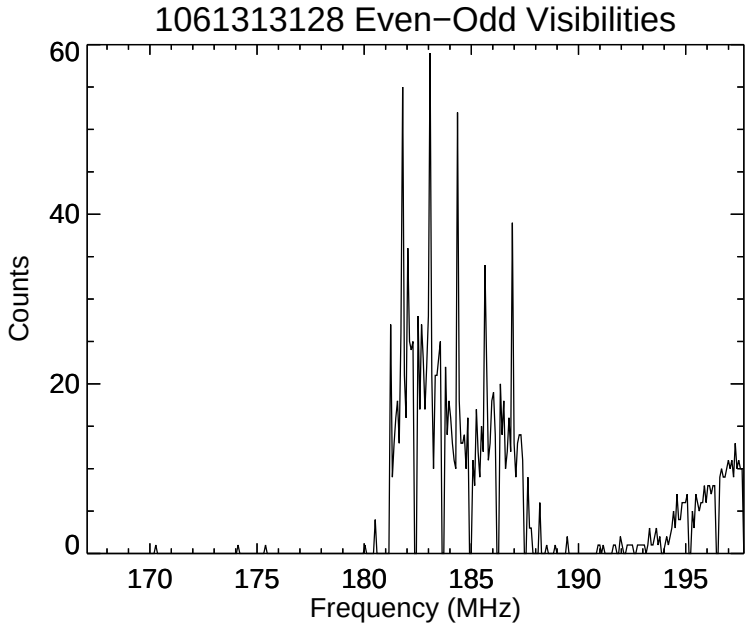


Figure 8.4: Histogram of residual visibility outliers as a function of frequency for observation 1061313128 (GPS time), where the even and odd time samples have been subtracted to remove sky signal. There is a large increase in noise from about 181 MHz to 188 MHz.

Visibility statistics can be used to reveal low-signal effects that otherwise would not be seen. This may illuminate failure modes within our instrument or pipeline. One interesting case is an observation from August 23, 2013 which was deemed low-quality by our usual quality statistics measurements. Its frequency distribution of outliers is shown in Figure 8.4 for instrumental XX. This observation occurs two pointings before zenith, and therefore has a characteristic increase in noise at higher frequencies. However, there is also a large increase in noise in the frequency range 181 MHz to 188 MHz.

A potential candidate for this increase in noise is digital TV. Channel 7 broadcasts with center frequency 184.5 MHz for a width of 6 MHz, which fits nicely into our frequency range of significant outliers. However, this only occurred for two minutes, the length of one observation. While ionospheric weather may allow transmission and can fluctuate often, the signal is more likely a reflection of digital TV off of a passing satellite or airplane.

Investigating anomalies in the visibility statistics is ongoing research. For now, we use this metric to help cut contaminated data that is not caught by other quality assurance strategies.

8.2 Small window power effects

Reducing contamination and systematics in the EoR window is one of the main goals of an EoR analysis. This takes considerable time; some small effects may require long integrations in order to observe, but still be orders of magnitude greater than the EoR signal itself. However, our in situ simulation in §5 can be used to quickly characterize a wide variety of effects.

We have found and mitigated important window power contaminants regarding the beam, or gridding kernel. These issues highlight the extreme precision we need to achieve. It would have taken many seasons' worth of integration in order to reveal them, and thus lowering our window contamination through the in situ simulation saves countless hours. We investigate systematics caused by our pipeline, and solve them to make our analysis more robust.

8.2.1 Beam clipping

We generate the beam from in-depth theoretical modeling [100, 99] and hyperresolve to extreme levels (typically to $\frac{1}{7000}\,\lambda$), as discussed in §3.1.3. There is, however, some inherent error to the input model, and improving it is an area of continual research. Most of the beam is well defined to a high precision; within the half-power range, the beam has been proven to do well for precision catalog science [39, 12].

The beam edges in the *uv*-plane are the source of the most significant errors. The theoretical modeling cannot characterize this, and generates fluctuating error out in the nulls. We account for this by clipping the beam—we only form the beam out to a certain percentage. This percentage is extremely small, usually only at about the 1% level. We then use this beam as a gridding kernel for our sources, and thus it is overlaid onto the *uv*-plane over 10,000 times.

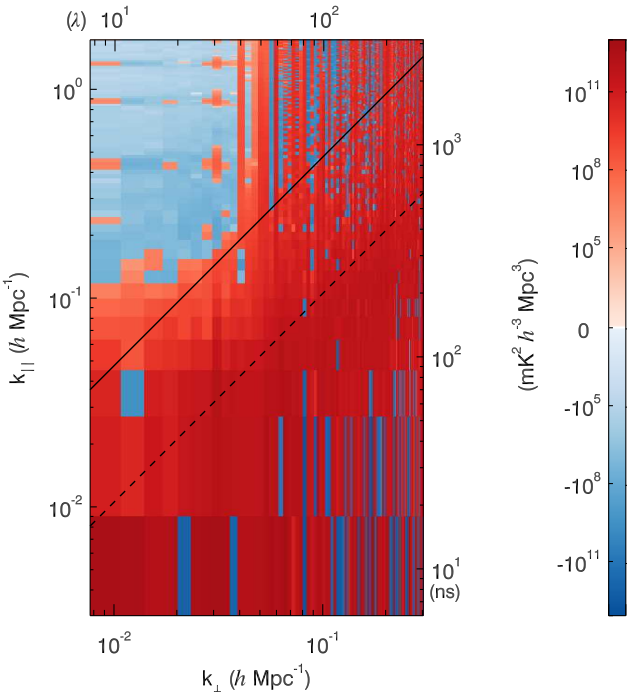


Figure 8.5: 2D difference power spectrum between a reference and an edge-subtracted, renormalized beam. Blue indicates a decrease in power, and red indicates an increase in power relative to the reference. The beam correction lowers window contamination.

A very small edge was made when we clipped the beam this way. A Fourier transform of an edge creates a wide-scale ringing in the Fourier dual space. Our small clip that was gridded over 10,000 times creates a series of small contamination ripples when the uv-space is Fourier-transformed into image space. This happens on all scales, which unfortunately includes the modes in the EoR window.

We tested a wide variety of solutions to the beam clip. The most smooth solution in Fourier space, and thus the best solution for EoR science, was to subtract off the clip edge from the entire beam and renormalize. Figure 8.5 compares a reference to the correction for one observation, where blue indicates a decrease in power and red indicates an increase in power relative to the reference. We have successfully lowered the EoR window with our beam-

clip correction. The subsequent rise in the wedge is not surprising; we are moving power from the point sources' ringing contamination back to where it belongs in the foreground wedge.

8.2.2 Cyclic beam errors

One very important aspect in errors that dominate our EoR measurements is their dependence on frequency. Often, it is not the fact that the error exists that makes it significant, but rather that it varies as a function of frequency in such a way that it contaminates the modes of interest. This concept was explored heavily in the context of calibration errors in §6.

Another form of frequency-dependent errors was discovered in the degridding methodology. A *uv*-plane is generated by Fourier-transforming a catalog of the sky. Then, we integrate within a beam kernel at the baseline location on the *uv*-plane to estimate a model visibility—a process we call degridding. This is done for every frequency since baselines move as a function of frequency.

The *uv*-plane is pixelated, and the process of degridding is effectively a discrete sum at specified points. There is error associated with approximating an integral this way. Even though we have tried to mitigate beam sampling errors with the hyperresolution of the beam kernel, we still have an integral with sample points determined by the resolution of the *uv*-plane.

The gridding resolution of the *uv*-plane determines when the errors repeat. While we have taken steps to reduce the error, its cyclic nature is the real issue. This creates characteristic harmonics in power spectrum space. These harmonics are quite small, only on the 1% level. However, there are *many* harmonics, dependent on all the contributions to the model visibilities. Harmonics can be pushed up past the EoR window by changing the gridding resolution to be finer, but at this time the computational costs are not feasible.

Instead, we have mitigated this issue by applying a delay filter. We transform model visibilities to delay space (Fourier-transforming along frequency) and filter out all structure that

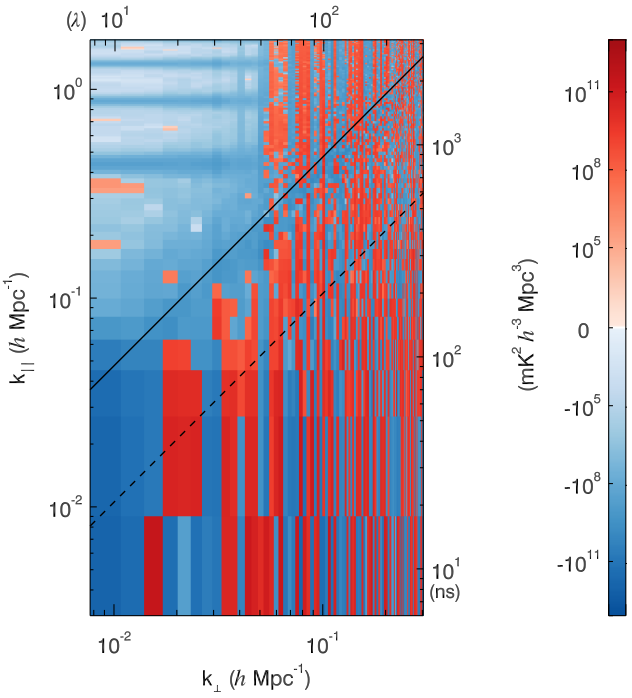


Figure 8.6: 2D difference power spectrum comparing a reference to a power spectrum that was generated from delay-filtered model visibilities. There is a decrease in the EoR window compared to the reference. While this removed the cyclic beam error contamination from the EoR window, the error still remains in the foreground wedge.

occurs beyond the calculated horizon. Then, we reform frequency visibilities by performing an inverse Fourier transform, while using appropriate window functions. This effectively removes all effects, including cyclic beam errors, from the EoR window. We do not expect any real contributions to the EoR window in the model visibilities; if there are any unknown real contributions, we will artificially remove them. At this time, we do not attempt to remove these errors from the foreground wedge.

Figure 8.6 shows the effect of our model delay filter. The 2D difference power spectrum compares a reference power spectrum to a power spectrum generated from delay-filtered visibilities, where blue indicates a decrease in power relative to the reference. We have

successfully removed power from the EoR window. In order to remove the contamination from all regions in power spectrum, we may have to investigate alternatives to degridding itself.

8.3 Anti-aliasing filters

As described in §3.1.6 & §3.2, we choose to integrate observations in HEALPix coordinates to avoid w-projection issues. This requires a multitude of steps, including FFTs, DFTs, and interpolations. In order to find ways to lower our systematics, we investigate the potential errors that may arise from our choice of integration space.

The HEALPix integration procedure can be summarized into four steps:

1. Fast Fourier Transform the uv-plane into the slant orthographic projection of the sky.

2. Interpolate the slant orthographic projection into the specified HEALPix pixels. Pixel area between slant orthographic and HEALPix projections is constrained to be comparable to avoid interpolation effects.

3. Integrate across all observations.

4. Direct Fourier Transform the HEALPix pixels to a uv-plane. Data is not regularly gridded in the HEALPix frame so fast Fourier methods cannot be used.

For step 2, we explicitly calculate our desired HEALPix pixel indices beforehand. We choose a specific subset of pixels that are covered well by observations to avoid excess computation time in step 4. In the past, this has amounted to a square patch on the sky in the center of the beam. The extent of the chosen region must correspond to the pixel size in uv-space to avoid aliasing, and thus we constrain the resolution in εppsilon by choosing the extent in image space.

However, we accidentally introduced a systematic with this method. Our chosen HEALPix set is a *square* on the sky, with well-defined, hard edges. When a Fourier transform is taken

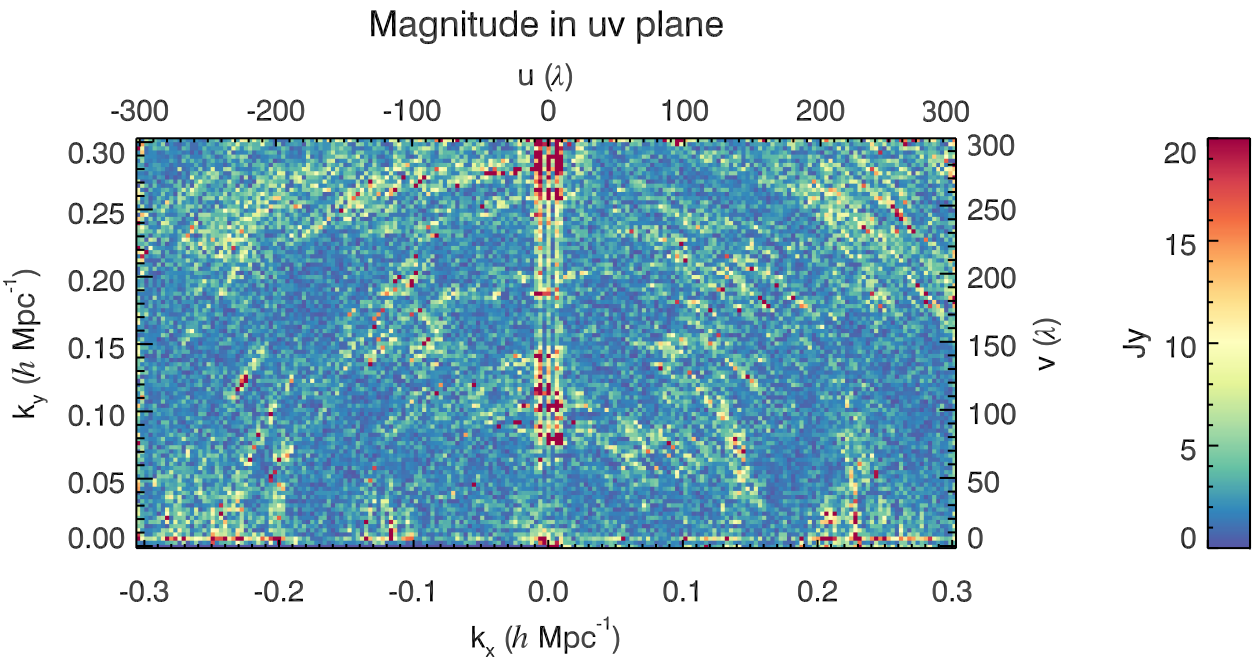


Figure 8.7: Residual uv-plane, integrated over Season 1, from a square selection of the HEALPix image plane in the center of the beam. Aliasing lines of high signal can be seen near $k_x = 0$ for all k_y, and $k_y = 0$ for all k_x.

over data with a hard edge, there is subsequent ringing. This ringing is not inherent to the data and will contaminate our power spectrum.

This systematic can be seen in uv-space. Figure 8.7 shows the residual uv-plane in units of Jy over a Season 1 integration. It is the result of a DFT for every HEALPix pixel to the uv-plane. While most of the structure is true, like the uv-rotation of baselines creating circular bands of high signal, there is false signal as well. This manifests as vertical and horizontal lines of high signal near $k_x = 0$ for all k_y, and $k_y = 0$ for all k_x. True uv-data should not be coupled to our choice of axes; thus, we can be certain that linear features of this nature are false.

In order to remove this systematic, we need to apply an anti-aliasing filter to the HEALPix image plane before the DFT. We expand our HEALPix selection set have a larger extent,

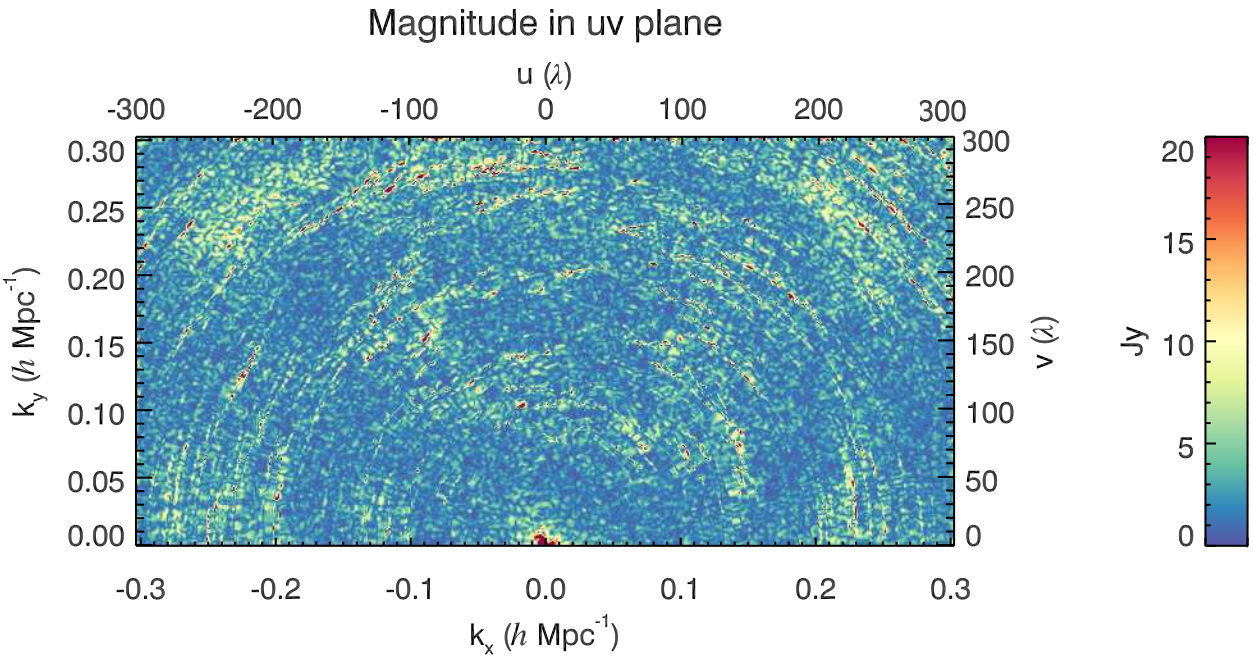


Figure 8.8: Residual *uv*-plane, integrated over Season 1, from a Tukey-filtered HEALPix image plane. The resolution has increased due to an increase in image extent. The aliasing lines from Figure 8.7 have been mitigated by our Tukey filter.

which will subsequently reduce the size of our *uv*-pixels. We then apply a Tukey filter[1], with a flat region comparable to our original square selection and curved cosine slopes reaching zero on all sides. In total, our new selection is triple in extent both in RA and DEC. The anti-aliasing window is applied once for all calibrated data, model data, and sampling map cubes, and is applied as the square for all variance map cubes.

The result of a Tukey-filtered residual *uv*-plane integrated over Season 1 is illustrated in Figure 8.8. Other than the filter and the decrease in pixel size, this *uv*-plane is comparable to Figure 8.7. Now, the vertical and horizontal aliasing is gone, indicating that the Tukey window function removed the aliasing from the hard image cut significantly. Not only is this beneficial for EoR science, it will also increase our scientific output for future cross-correlation

[1]Also known as a tapered cosine window.

studies in the uv-plane.

However, not much improvement was seen in the EoR window in the power spectrum. After further investigation, we found that other image filters perform better. The Tukey filter significantly improves the uv-plane, but residual contamination continues to harm the EoR window.

We examined this predicament in uv-space by reducing the amount of gridded baselines for clarity. In this test case, we only simulated gridded baselines which fell between 40 and $50\,\lambda$, creating an annulus of signal in the uv-plane. We then applied a variety of different image filters to investigate if there was any residual contamination due to the choice of filter. In Figure 8.9, we show three instances of uv-plane weights in units of visibility contributions to each pixel, where visibilities have been spread out by the gridding kernel.

The top uv-plane in Figure 8.9 is the weights for the unfiltered, large HEALPix set. It has the same resolution as the other panels, but has obvious signs of aliasing that we reduced previously in Figure 8.8. In this case, the vertical and horizontal lines fill the entire space, which is particularly noticeable where their shouldn't be any signal at all! There doesn't appear to be significantly higher aliasing along the zeros of the axes like in Figure 8.7, but those were particularly strong due to coupling with the high-signal foregrounds at low k (which are not present in our annulus test). The top panel indicates that there is lower-level aliasing *everywhere* that is proportional to the signal that it intersects.

The middle uv-plane in Figure 8.9 is the weights for an annulus with applied Tukey image filter, like in Figure 8.8. The reduction of aliasing is quite large, but not complete. There still remains some local aliasing near the annulus itself. We attempt to alleviate the residual aliasing by applying a Blackman-Harris window function in the image space instead. This is a much harsher filter. The result is the bottom uv-plane, where the residual aliasing has been dramatically reduced everywhere. Small sidelobes remain, an unavoidable fact in window function application.

As with all other changes to the analysis, we compare the final result in the power spectrum to confirm in the space of the measurement. We perform simulations similar to

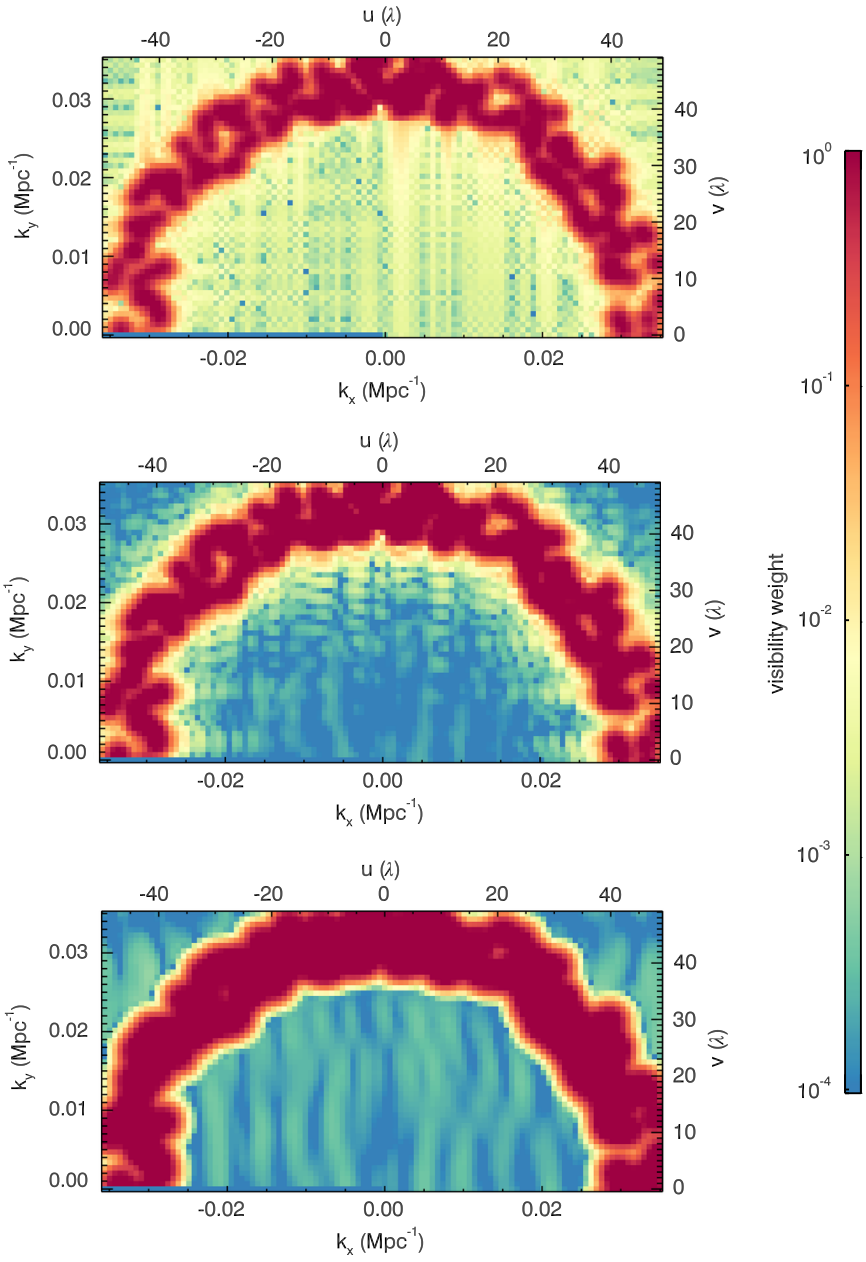


Figure 8.9: *uv*-weights of simulations with baselines in the 40 to 50 λ range for three image filters. The unfiltered set (top) has significant aliasing. The Tukey-filtered set (middle) reduced this aliasing, but still suffers from local bleed near the annulus. The Blackman-Harris window filter (bottom) performs the best, reducing aliasing everywhere. Units are in visibility weight, or the sum of visibilities spread out by the gridding kernel in each pixel.

the process in §5, holding the calibration to be perfect. The only change between the various simulations is the image window function applied.

We compare the three different image filters (square, Tukey, and Blackman-Harris) in Figure 8.10. The 1D space is very similar to a typical limit cut, and thus does not include the wedge or low-coverage regions at large $k_\perp$. In addition, the input Gaussian EoR and the recovered Gaussian EoR are plotted to give context to the level reached by the various image filter tests. The only image filter test that reveals the underlying EoR signal is the test with the Blackman-Harris window function. All k-modes are lower, and there is significant improvement.

While applying a smooth image window function decreases contamination, we are not completely removing it all. As seen in Figure 8.9, we still experience sidelobes, even with the Blackman-Harris image filter. This propagates to the power spectrum, as seen in Figure 8.10 at lower k-modes.

Oddly enough, we can actually *reduce* this contamination by including less data. The uv-plane for the MWA is essentially filled out to approximately $60\,\lambda$, and is sparse beyond that. When we make final limits, we exclude contaminated regions and regions that have poor uv-coverage. Besides diagnostic potential, there is no reason to include baselines out to $300\,\lambda$ like is shown in all 2D power spectra in this work up until this point. If we exclude longer baselines during the transformation from uv-space to HEALPix space, their contributing aliasing effect, showcased in Figure 8.9, will be excluded. This has a significant impact on the EoR window contamination.

Therefore, we only grid out to $100\,\lambda$ when creating the final uv-plane in FHD. Now, the pixel sizes for the slant orthographic and HEALPix projections are no longer matched; we may need to investigate the consequences of this in the future. The result of the smaller gridding extent is compared to the other power spectrum metrics in Figure 8.10. There is a drastic decrease in contamination when we only grid to $100\,\lambda$ while using a Blackman-Harris filter in image space. We now recover a significant portion of the EoR for most modes in the diagram.

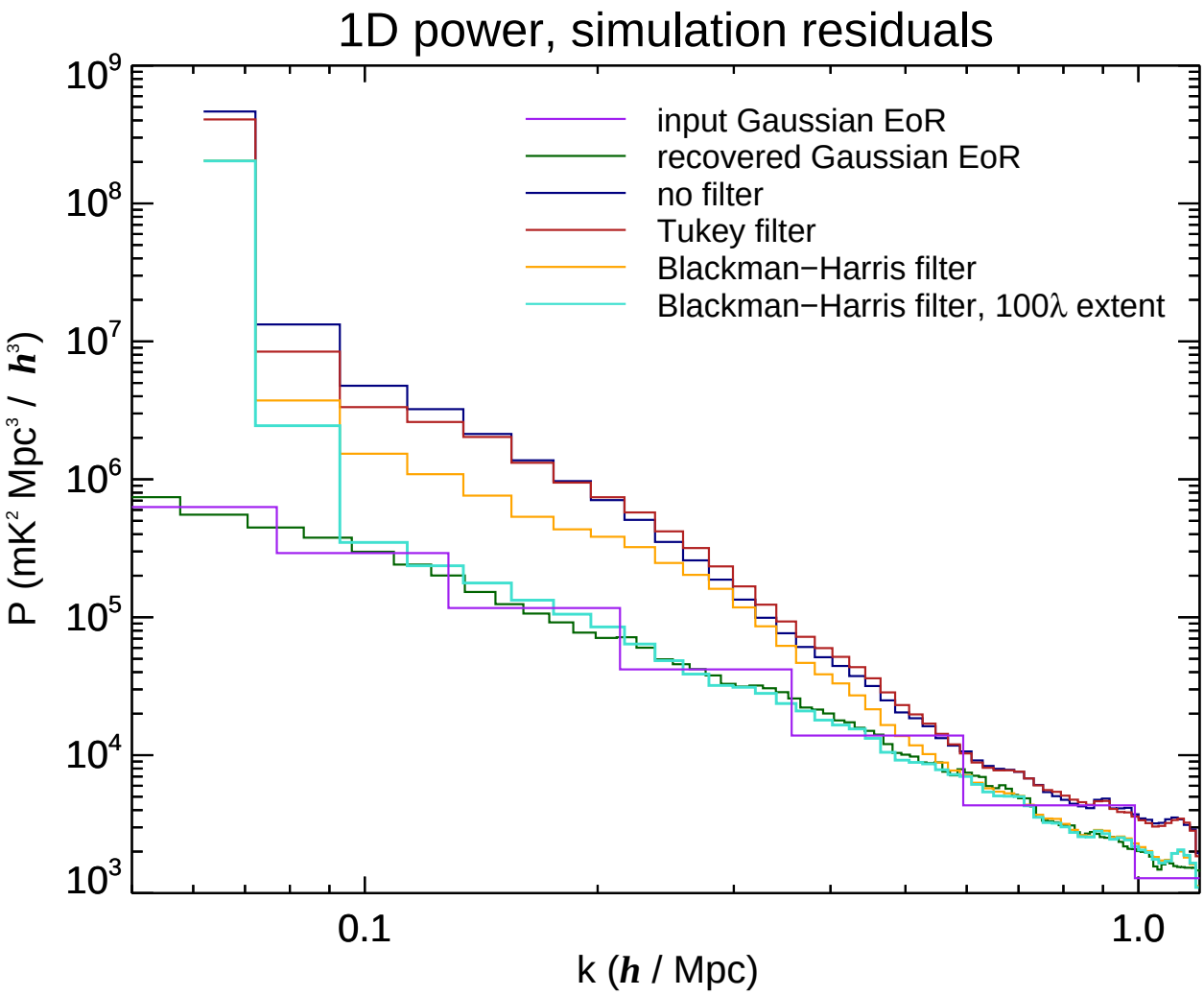


Figure 8.10: 1D power spectrum residuals for various image filter tests, with simulation setup similar to §5, and with no contributing wedge or low-coverage regions at large $k_\perp$. Input Gaussian EoR (purple) and recovered Gaussian EoR (dark green) are provided for context. Applying no image filter (blue) results in the highest level of contamination in general, with the Tukey filter (red) not far below it. We see significant improvement with the Blackman-Harris image filter (orange). If we also limit the uv-gridding extent to $100\,\lambda$ (teal), than we fully recover the EoR for a majority of modes.

We have successfully reduced contamination in the EoR window with the use of anti-aliasing filters and corrections to various beam effects. This decrease has been significant enough for us to recover the underlying signal in most EoR window modes, even with realistic residual foregrounds. The in situ simulations have been vital in finding these systematics, highlighting their importance in removing contamination in our real-data pipeline.

Chapter 9

NEW UPPER LIMIT ON THE EOR SIGNAL

We present upper limits on the EoR as a culmination of our improvements. Incorporating strategies and methodologies from §4 through §8 into a new analysis of a long integration gives a metric on how much we have progressed. In addition to limits on the 1D power spectrum, we provide 2D diagnostic power spectra in order to study locations of contamination and our propagated errors. This motivates our choice of binning scheme for the final calculations.

For this limit, we've decided not to present any new data. Instead, we are showcasing the vast improvements made since the previous MWA limit created with the US pipeline [5]. We use 1,029 high-quality observations from Season 1, chosen by quality assurance cuts summarized in §8.1. Since this data set is the exact same as that used in [5], we can make direct comparisons.

9.1 Analysis pipeline

The analysis methodology is core to the limit, and the main focus of this work has been to make the pipeline more robust and with fewer resulting systematics. We have gone into great detail on all the major adjustments in previous chapters, and we will summarize them here for clarity.

We start with describing our general analysis packages: FHD (§3.1) and εppsilon (§3.3). In FHD, we generate a beam gridding kernel (§3.1.3), create model visibilities from a sky catalog (§3.1.4), calibrate (§3.1.5), and image each observation (§3.1.6). HEALPix images of the calibrated data cubes, model data cubes, sampling map cubes, and variance map cubes are integrated across all observations for each polarization and by each even–odd time sampling set (§3.2). These integrations are imported into εppsilon, where images are trans-

formed into uv-space (§3.3.1), observed noise is calculated (§3.3.2), frequency is transformed into k-space (§3.3.3), and averages of k-space are performed for power spectra (§3.4). Noise is propagated, allowing for calculated error bars rather than estimated error bars. There have been adjustments to many analysis steps since our previous US limits [5], so we will list the major contributors sequentially where they occur in our pipeline.

Beam gridding kernel: We generate a hyperresolved ($\frac{1}{7000}\,\lambda$) beam from a theoretical simulation [100, 99] to an extent determined by numerical limitations of the theory. This left a small uv-clip, which contaminated the EoR window. We now remove this small clip from the entirety of the beam and renormalize (§8.2.1).

Model and calibration catalog: We use KGS [11, 12] as the main catalog in the primary lobe, and use GLEAM [39] in the sidelobes with a brightness factor to match the scale of KGS. Over 11,000 sources are included out to the second sidelobe.

Model delay filter: A cyclic beam error occurs during the degridding process. The error associated with approximating an integral as a finite sum during degridding is small—only about 1%—but it fluctuates as a function of frequency, contaminating the EoR window. When we generate model visibilities, we now remove any signal that extends beyond the horizon in delay space (§8.2.2).

Calibration scheme: We have thoroughly investigated the effects of calibration experimentally (§4) and theoretically (§6), which highlights the importance of smooth calibrations. At this time, we model the amplitude with scaled auto-correlations (§4.5) and independent fits to the digital gain jump (§4.3), and we model the phase as linear ramps with fits to the 150 m cable reflection.

***uv*-gridding extent:** We limit the gridding extent to $100\,\lambda$ in the uv-plane before transforming to slant orthographic projection. Every baseline contributes a small aliasing

feature due to our choice of integration space that varies as a function of frequency (§8.3). By excluding baselines that are purely diagnostic, we remove their aliasing.

HEALPix image filter: When we interpolate to HEALPix pixels, we previously chose a set of pixels well within the primary beam, where all observations contributed. This had a hard edge, and thus contaminated our EoR window. We now triple the extent of the image and apply a Blackman-Harris image filter to remove this aliasing (§8.3).

All of our changes have been implemented in the calculation of our limit, and they constitute the majority of the direct improvements. By analyzing the same Season 1 data from previous limits, we can directly compare our new methodologies, and we can attribute any movement in the limit as direct consequences of updates to our analysis.

9.2 2D diagnostics

Before we calculate limits and binning strategies, we investigate our long integration in 2D power spectra. This builds confidence in the validity of our final result, and highlights potential areas of improvement for the future. By default, we exclude frequencies above the digital gain jump since we know our current calibration schemes cannot capture nonlinear effects (§7.1). In addition, since our analysis only grids out to $100\,\lambda$, our 2D power spectra are limited in $k_\perp$ compared to previous sections.

The 2D power spectra for our long integration are shown in Figure 9.1, where each row is instrumental polarization XX or YY and each column is calibrated data, model data, or residual data. Compared to our single-day integration example in Figure 3.5, there are some notable differences. Most importantly, the region below the first coarse band harmonic and above the foreground wedge at low $k_\perp$ has a detection (yellow). In other words, there are no noise-like pixels in this region. The EoR signal is orders of magnitude smaller, so the detection is that of a systematic. This is our most sensitive region in the EoR window, and thus will give the best upper limit.

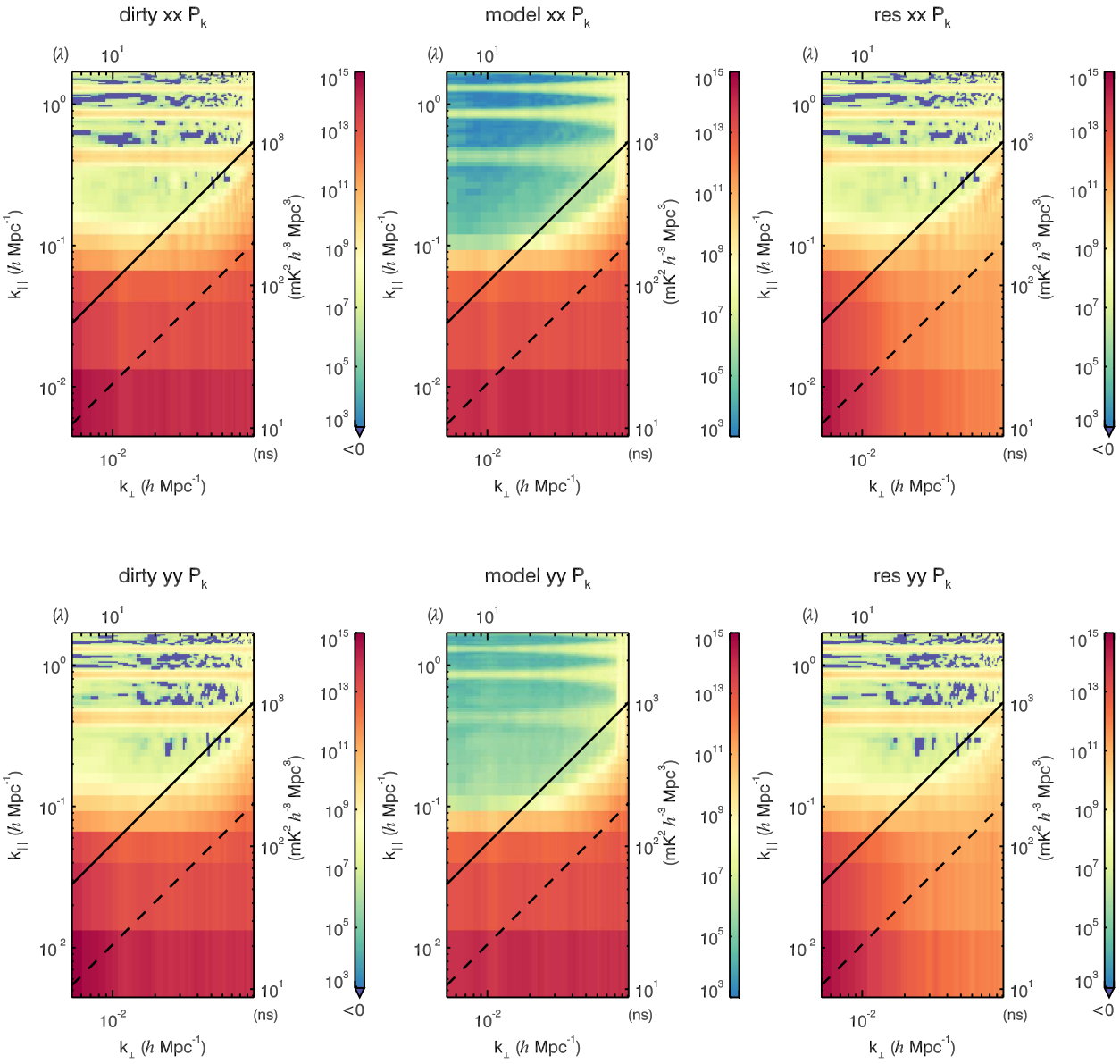


Figure 9.1: The 2D power spectra for the calibrated data, model, and residual for Season 1 (each for instrumental XX and YY) for frequencies smaller than the digital gain jump. We have detected a systematic below the first coarse band harmonic and above the foreground wedge at low $k_\perp$.

However, there are numerous noise-like pixels (dark purple-blue), especially above the first coarse band harmonic. This indicates that there are no dominant systematics that have been detected in those specific regions. Longer integrations will lower the calculated limits for those k bins.

Other notable features in Figure 9.1 include areas of increased contamination. This is expected in the coarse band harmonics; §4.5 & §7.1 show that using a scaled auto-correlation for calibration amplitude does not capture coarse channel edges correctly. However, there is an increase in the model power spectra in the EoR window for YY compared to XX. This is caused by a small beam kernel clip error on the order of 0.1% in off-zenith pointings, and will be corrected in future long integrations.

In addition to 2D *data* power spectra, we can also analyze our 2D *noise* power spectra. Figure 9.2 shows the expected noise, the observed noise, the noise ratio, and the calculated error for instrumental XX. If our propagated noise (top left) is the same as the difference between even and odd data sets (top right), then the noise ratio should be flat and equal to 1.

However, the propagated noise is higher than the observed noise. While it is completely necessary to reduce aliasing, the Blackman-Harris image filter caused large-mode correlations. Since we assume that there are no significant correlations, this leads to a discrepancy between what is expected and what is calculated. As a result, our error bars will be unnecessarily large in our final limit calculations. There are plans to account for this in future work.

9.3 Binning

Limits are calculated from large integrations of the data. When we create 1D power spectra for this purpose, we integrate in spherical shells in 3D k-space. However, we do not want to include all voxels. Some regions are contaminated with foregrounds, which would completely dominate our signal.

We need to make choices on which regions to exclude. To remain unbiased, we only exclude *ranges* of voxels; we do not pick and choose from one voxel to the next. These

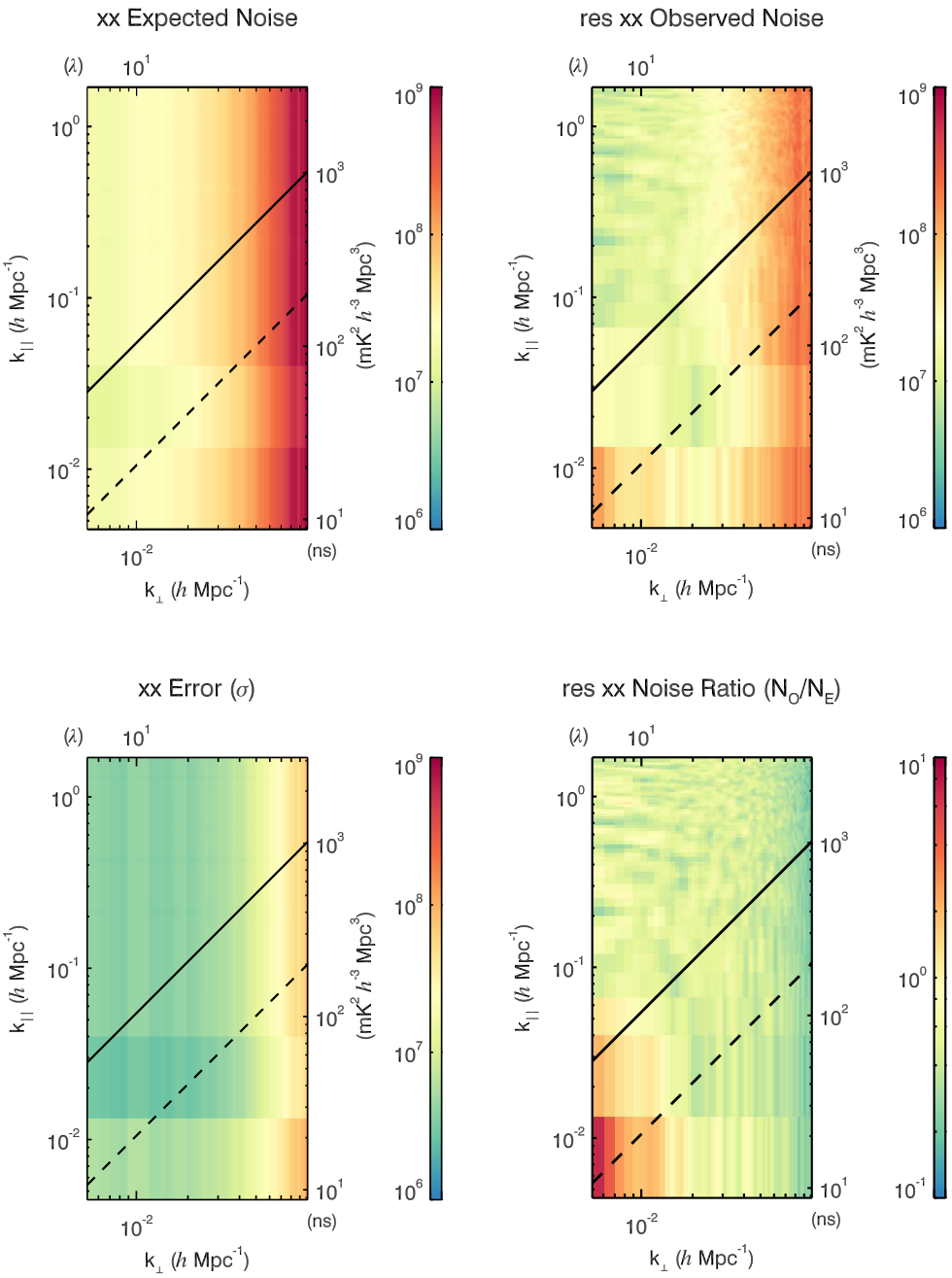


Figure 9.2: The 2D power spectra of expected noise, observed noise, error bars, and noise ratio for Season 1 in instrumental XX for frequencies below the digital gain jump. The expected noise is larger than the observed noise due to the Blackman-Harris window function. As a result, our error bars on calculated limits will be slightly exaggerated.

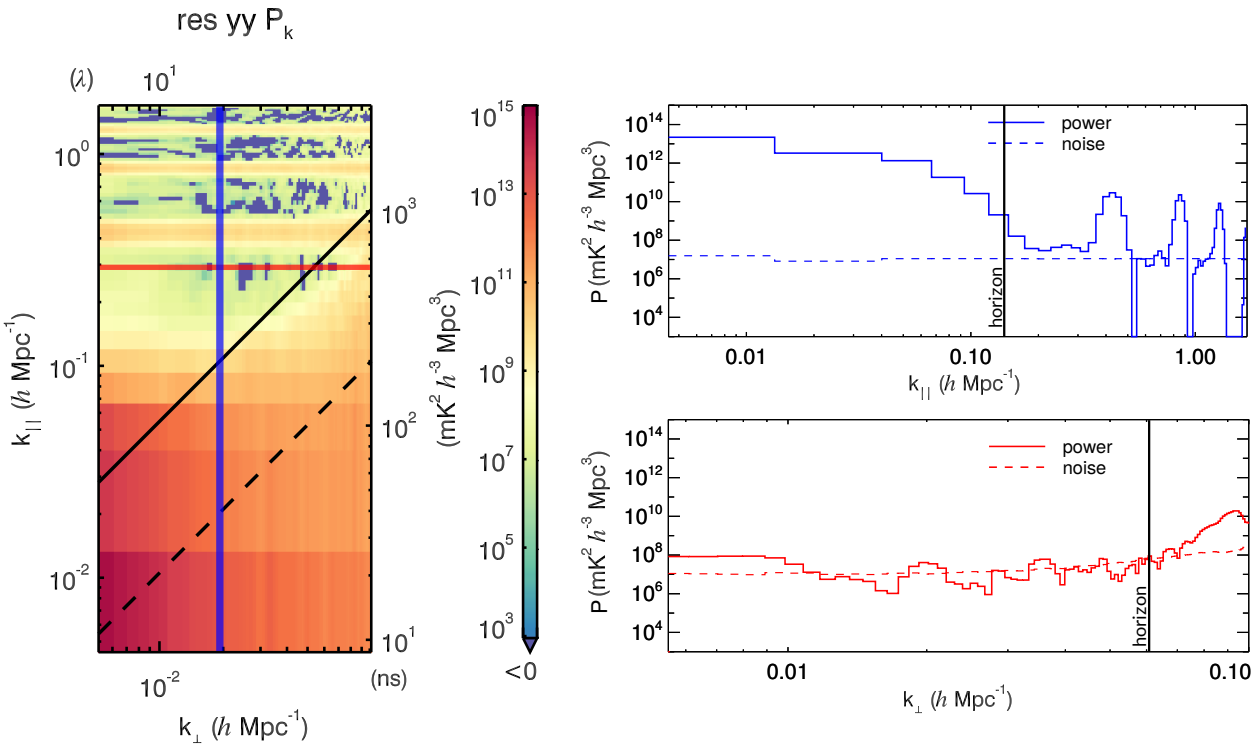


Figure 9.3: 1D power spectrum slices in $k_{||}$ (blue) and $k_\perp$ (red) for specific ranges chosen from the residual 2D power spectrum for instrumental YY. Low $k_{||}$ modes are contaminated by the foreground wedge (and for a few extra modes beyond the horizon), as is also the case for high $k_\perp$ modes. Low $k_\perp$ modes are likely contaminated due to calibration leverage [76].

ranges can have a significant impact on the final result.

To demonstrate how the characteristic regions of contamination affect the 1D power spectra, we show 1D power slices as a function of $k_{||}$ and of $k_\perp$ in Figure 9.3. The slice ranges are shown on the 2D power spectrum—the vertical, blue range corresponds to the power slice as a function of $k_{||}$ (top right) and the horizontal, red range corresponds to the power slice as a function of $k_\perp$ (bottom right).

The foreground wedge dominates low modes in the $k_{||}$ slice, which couples to the coarse band harmonics at higher modes. In addition, the foreground wedge smoothly decreases after the horizon line, contaminating some extra modes. As for the $k_\perp$ slice, both high and

low modes are contaminated. High modes are contaminated due to the foreground wedge for this particular slice, and the noise has increased by an order of magnitude due to a decrease in uv-coverage. Low modes are likely contaminated due to calibration leverage since we do not have any reliable calibration models at low modes (see §7.2 and [76]). Fortunately, this only seems apparent below $20\,\lambda$, but could theoretically contaminate out to $50\,\lambda$.

Given our knowledge of where the affected contamination regions are from the 2D power spectra and the 1D slices, we can choose our integration region. We preclude $k_\perp$ modes greater than $80\,\lambda$ and less than $20\,\lambda$, which removes regions with increased noise and contamination, respectively. As for the wedge, we exclude bins below the horizon line. The remaining voxels are spherically averaged directly from 3D space.

In addition to removing contamination, we must also shorten our frequency band. We cover redshifts from about 6 to 7.5 in our frequency range. The EoR signal could change significantly in that large time period, which would confuse our measurements and interpretation [116]. Therefore, we choose a smaller range to create limits that are more representative of their mean redshift. Our instrument response also forces us to shorten our frequency band; the digital gain jump can cause power contamination (see §4) and thus better limits can be made if it is avoided.

We choose a frequency band from 168.7 MHz to 187.4 MHz, corresponding to a mean redshift of 7. While the frequency band is nominally 18.7 MHz, the application of a Blackman-Harris spectral window (see §3.3.3) effectively reduces the range to only 9.4 MHz. This frequency range excludes the first coarse band, which seems to have higher variance across the tiles of the array.

All of our regional cuts for our chosen frequency range are shown in a 2D power spectrum in Figure 9.4. We are left with a region of low contamination (highlighted) in which to make our 1D limits. Future work will focus heavily on reducing contamination at low $k_\perp$ in order to increase contributing regions.

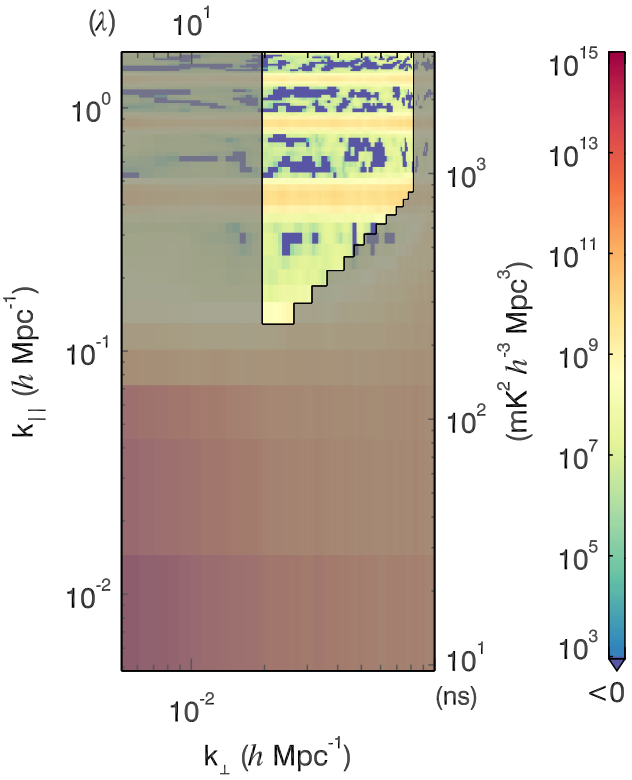


Figure 9.4: 2D region of interest for our 1D limit binning scheme. The power spectrum was made with frequencies 168.7 MHz to 187.4 MHz, corresponding to a mean redshift of 7. The highlighted area has been specifically chosen to exclude contamination.

9.4 Limit calculation

With our improvements to the pipeline and our chosen binning scheme, we are now in a position to calculate upper limits on the EoR. In generating the final results, we must be careful to incorporate errors, include confidence intervals, and manage pixels that fluctuate around zero. Therefore, the 1D power spectrum is not the final result; we must calculate the upper limit *on* the 1D power spectrum.

We employ Bayesian statistics to calculate the upper limit [2][1]. To begin, we describe

[1] [2] covers the same derivation.

the probability of the true data x given the measured data x' as

$$\Pr(x|x') = N\Pr(x'|x)\Pr(x), \tag{9.1}$$

where N is a normalization constant, $\Pr(x'|x)$ is the probability of the measured data given the true data, and $\Pr(x)$ is the prior distribution for the true data.

By using this formalism, we will be able to correctly describe pixels that fluctuate around zero. When the power is near the thermal noise levels of the integration, there is some chance that it will become negative. This is purely due to the thermal noise; we expect the underlying power signal to be positive. Therefore, our prior on the data is

$$\Pr(x) = \begin{cases} 0, & \text{if } x < 0 \\ \\ 1, & \text{if } x \geq 1 \end{cases}. \tag{9.2}$$

The probability of the measured data given the true data, $\Pr(x'|x)$, is the distribution of what we measure. In §3.3, we averaged Erlang-distributed data to generate 1D power, where we assumed that the Central Limit Theorem produced averages that were Gaussian distributed. Thus, we can write

$$\Pr(x'|x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x'-x)^2}{2\sigma^2}}. \tag{9.3}$$

Now, we would like to normalize Equation 9.1 such that the integral of $\Pr(x|x')$ is 1. Given Equation 9.2 & Equation 9.3, we can solve for N:

$$N = \frac{\sqrt{2/(\pi\sigma^2)}}{1 + \text{erf}\left(x/\sqrt{2\sigma}\right)}, \tag{9.4}$$

where *erf* is the error function.

Equation 9.1 is now fully known. In order the calculate an upper limit, we need to integrate $\Pr(x|x')$ until we have accounted for 97.7% of the distribution, which corresponds to a 2σ detection:

$$\int_{-\infty}^{x_{UL}} dx \Pr(x|x') = 0.977. \tag{9.5}$$

Solving for x_{UL}, we get

$$x_{UL} = \sqrt{2\sigma^2}\,\mathrm{erf}^{-1}\left(0.977 - (1-0.977)\,\mathrm{erf}\left(x'/\sqrt{2\sigma^2}\right)\right). \tag{9.6}$$

The only inputs needed to calculate the upper limits are the data and their variances. We inherently assume that the foregrounds and resulting contamination are not coherent with the EoR signal, and thus an upper limit on the data is an upper limit on the EoR signal.

Figure 9.5 is our calculated upper limit on the 1D measured power spectrum using the binning techniques described in §9.3. We describe the final limits in units of mK^2 since the EoR signal is roughly flat in that space. Our thermal noise (dashed black) rises in mK^2 from low k to high k—measurements low k are the most sensitive to the EoR signal.

Our measured power (solid black) is sometimes negative due to thermal noise fluctuations. The 2σ error bars (grey) reach zero when the signal is consistent with zero. However, low-k modes do have detections on systematics. Integrating more observations will decrease the error bars on those modes, but will not lower the systematic floor. Therefore, we must reduce contamination before we integrate further.

There are some distinctions between instrumental XX (top panel) and instrumental YY (bottom panel). Our best limits come from instrumental YY, mostly due to some low-level contamination below the first coarse band harmonic that is present in XX but not YY. This has been seen in past work using our pipeline [5], and may be an indication of sky-related contamination. In addition, a clear detection of the 320 m cable can be seen in YY (near $1\,h\,\mathrm{Mpc}^{-1}$) but not in XX. We will explore these discrepancies in future work.

Our best upper limit is $6.75{\times}10^3\,\mathrm{mK}^2$ at $k = 0.2\,h\,\mathrm{Mpc}^{-1}$ for $z = 7$ in instrumental YY. Slightly higher yet comparable upper limits exist from $k = 0.2\,h\,\mathrm{Mpc}^{-1}$ to $k = 0.3\,h\,\mathrm{Mpc}^{-1}$. As for instrumental XX, the best upper limit is $7.44{\times}10^3\,\mathrm{mK}^2$ at $k = 0.2\,h\,\mathrm{Mpc}^{-1}$ for $z = 7$, and is by far the lowest upper limit for all modes in that polarization.

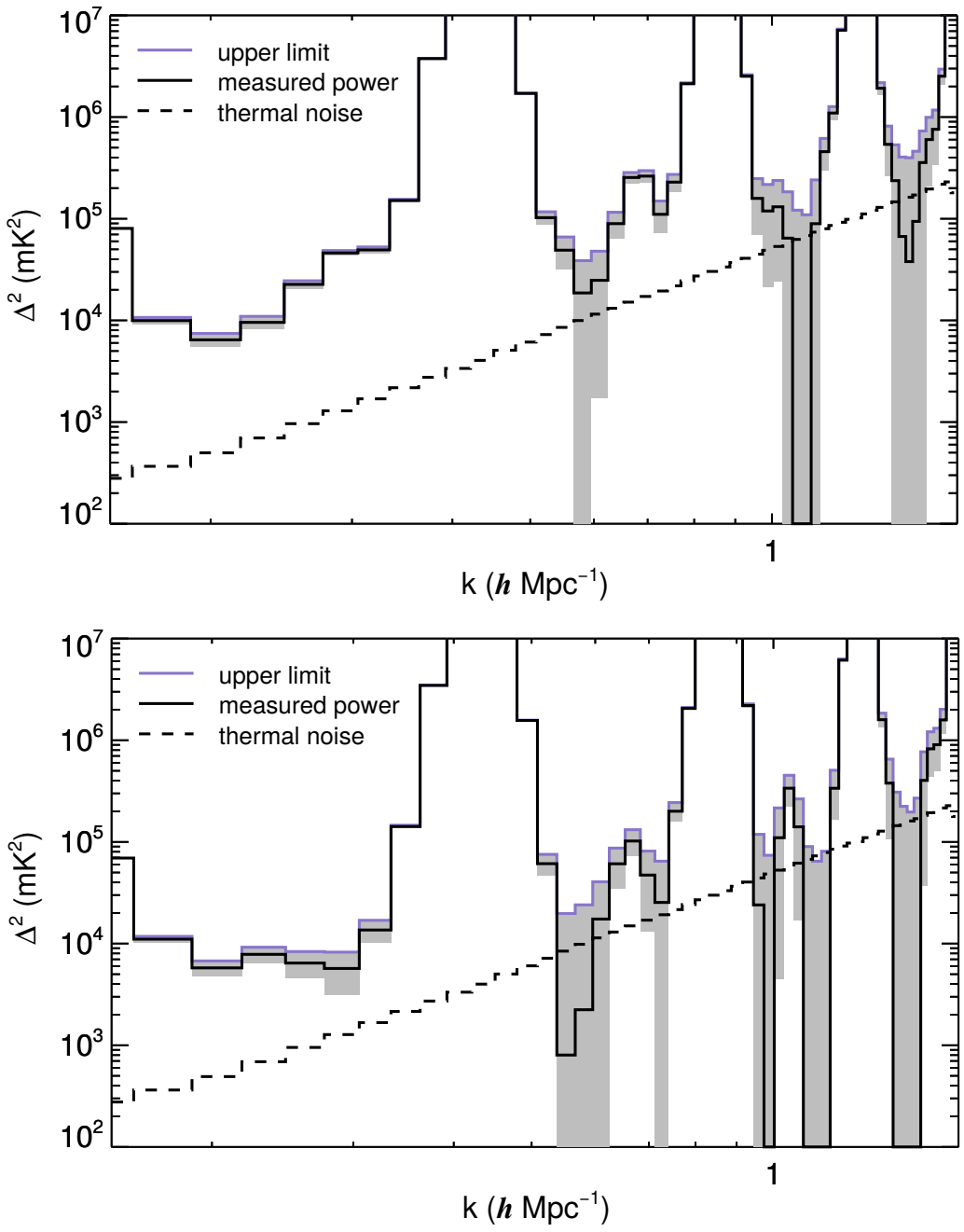


Figure 9.5: Upper limits on the measured 1D power for Season 1. We present upper limits (purple) on the measured power (solid black), with associated 2σ error bars (grey) and thermal noise (dashed black). Instrumental YY (bottom panel) is generally less contaminated than instrumental XX (top panel).

9.5 Comparisons and conclusions

By comparing to other limits, we can put our progress into context for our final conclusions. Fortunately, we can compare to the exact same data set of 1,029 observations from previous work using our pipeline. This will provide the best metric of development, with improvements coming directly from reducing systematics within the analysis, rather than differences in the actual observations.

We take the best limit from [5] for our direct comparison. Their binning scheme is slightly different, using frequencies from the lower third of the band and including more low $k_\perp$ bins for the average. Their resulting limits are calculated for a mean redshift of 7.1. This is not an exact comparison to our binning techniques, but is close enough to draw conclusions. The best limits from this work and from [5] are both from instrumental YY.

The comparison is shown in Figure 9.6. The significant reduction in systematic power on all k modes is the culminating effort of this work[2]. Our resulting final limit is $6.75\times10^3\,\mathrm{mK}^2$ at $k = 0.2\,h\,\mathrm{Mpc}^{-1}$ for $z = 7$, an improvement of a factor of 4.0 from [5]. All measured modes see comparable systematic reduction, except for modes associated with the 320 m cable reflection.

The thermal noise level has also lowered, which is the outcome of an increased image plane extent, a catalog change, and a binning scheme with more frequencies. Noise-like pixels are coupled to this thermal noise reduction and decrease proportionally. Systematic-dominated pixels, like those of low k, are not affected by a change in the thermal noise.

We have significantly reduced our systematic floor, yet some still remains at low k. Before further integration can be beneficial, we must continue to investigate causes and solutions to this contamination. The newly analyzed data set will be explored in depth to evaluate the origin of these systematics, and whether they are related to our analysis or to our data. In particular, we will focus on the conversion from the uv-plane to integrated HEALPix cubes for input into εppsilon, as well as calibration techniques which correct for the coarse channel

[2]Improvements in sky catalogs also contributed to contamination reduction [11, 39].

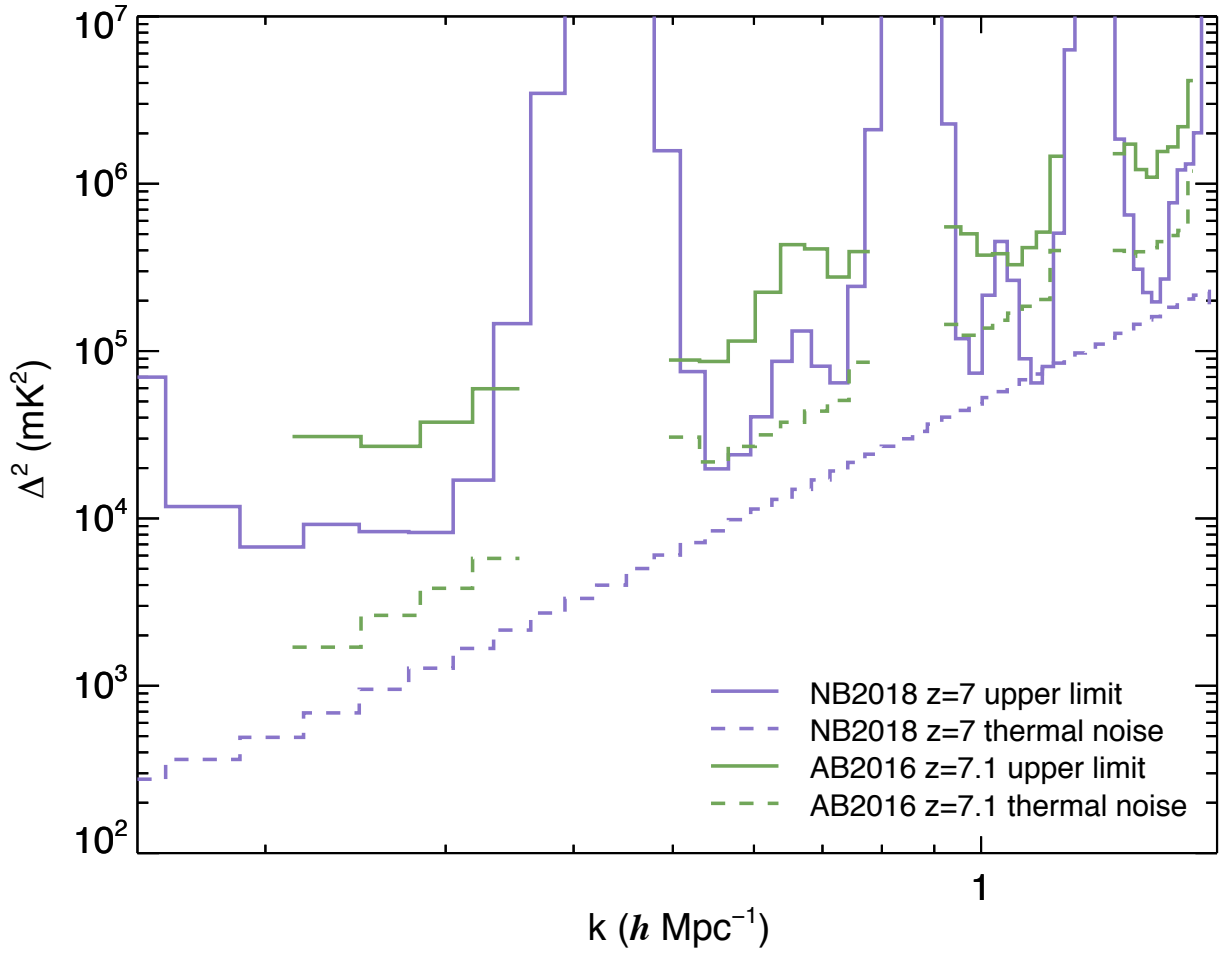


Figure 9.6: EoR upper limits on the measured 1D power for Season 1 compared to previous results from [5]. Both our analysis (purple) and the analysis from [5] (green) are for instrumental YY. We see a reduction of the systematic floor at low k of over $2\times10^4\,\mathrm{mK}^2$. The thermal noise floor has also been reduced due to different binning and image plane extents, but this has no effect on low k detections at this time.

edges and are better constrained for low-k modes.

It is a long process to analyze enough data to see this systematic floor. Unfortunately, we do not have the infrastructure to handle multiple iterations of this long integration. However, with the successful implementation of our end-to-end simulation pipeline, we can quickly analyze a variety of effects. Adding flexibility and improving the scope of the simulation will be essential to reducing our systematic floor.

Going forward, we will continue to study causes of contamination in our analysis, both in real data and in simulation. Creating 2D and 1D diagnostic power spectra, including calibrated data, model data, residual data, errors, expected noise, and observed noise, will be vital to understanding how changes to the analysis propagate to our final measurements. As our pipeline matures, we will continue to provide evidence that our analysis does not experience signal loss. These strategies for improvement will ensure that we continue to be reliable in future EoR measurements.

With these new limits, we are now closer to noise-dominated low k. We will continue to search for potential pipeline systematics through in situ simulations. In addition, we will investigate this new long-integration data set thoroughly to understand data-based and instrumental effects. Given the importance of calibration effects, we will also scrutinize our choice of calibration scheme, and find better ways to achieve highly accurate and smooth calibration solutions.

The scientific merit of this work focuses on the strategies and framework for reducing contamination in precision measurements of the EoR signal in power spectrum space. By providing our methodology, similar simulations and experiments can be reproduced for other instruments and pipelines. Further, our simulations contribute to reliability and robustness in our limits.

We have provided the groundwork for the next analysis stage in the search for the EoR signal, and we will continue to improve our upper limit, pipeline, and calibration.

BIBLIOGRAPHY

[1] 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, September 2016.

[2] A. P. Beardsley. *The Murchison Widefield Array 21 cm Epoch of Reionization Experiment: Design, Construction, and First Season Results*. PhD thesis, University of Washington, 2015.

[3] A. P. Beardsley, B. J. Hazelton, M. F. Morales, W. Arcus, D. Barnes, G. Bernardi, J. D. Bowman, F. H. Briggs, 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, Shankar N. Udaya, 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. The EoR sensitivity of the Murchison Widefield Array. *Monthly Notices of the Royal Astronomical Society: Letters*, 429(1):L5–L9, February 2013.

[4] A. P. Beardsley, B. J. Hazelton, M. F. Morales, R. J. Capallo, R. Goeke, D. Emrich, C. J. Lonsdale, W. Arcus, D. Barnes, G. Bernardi, J. D. Bowman, J. D. Bunton, B. E. Corey, A. Deshpande, L. deSouza, B. M. Gaensler, L. J. Greenhill, D. Herne, J. N. Hewitt, D. L. Kaplan, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, 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. A new layout optimization technique for interferometric arrays, applied to the Murchison Widefield Array. *Monthly Notices of the Royal Astronomical Society*, 425(3):1781–1788, September 2012.

[5] 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. *The Astrophysical Journal*, 833(1):102, December 2016.

[6] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K. M. Smith, R. S. Hill, B. Gold, M. Halpern, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, J. Dunkley, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *The Astrophysical Journal Supplement Series*, 208(2):20, September 2013.

[7] J. S. Bolton, M. G. Haehnelt, S. J. Warren, P. C. Hewett, D. J. Mortlock, B. P. Venemans, R. G. McMahon, and C. Simpson. How neutral is the intergalactic medium surrounding the redshift z = 7.085 quasar ULAS J1120+0641? *Monthly Notices of the Royal Astronomical Society: Letters*, 416(1):L70–L74, September 2011.

[8] J. R. Bond, A. H. Jaffe, and L. Knox. Estimating the power spectrum of the cosmic microwave background. *Physical Review D*, 57(4):2117–2137, February 1998.

[9] Judd D. Bowman, Iver Cairns, David L. Kaplan, Tara Murphy, Divya Oberoi, Lister Staveley-Smith, Wayne Arcus, David G. Barnes, Gianni Bernardi, Frank H. Briggs, Shea Brown, John D. Bunton, Adam J. Burgasser, Roger J. Cappallo, Shami Chatterjee, Brian E. Corey, Anthea Coster, Avinash Deshpande, Ludi deSouza, David Emrich, Philip Erickson, Robert F. Goeke, B. M. Gaensler, Lincoln J. Greenhill, Lisa Harvey-Smith, Bryna J. Hazelton, David Herne, Jacqueline N. Hewitt, Melanie Johnston-Hollitt, Justin C. Kasper, Barton B. Kincaid, Ronald Koenig, Eric Kratzenberg, Colin J. Lonsdale, Mervyn J. Lynch, Lynn D. Matthews, S. Russell McWhirter, Daniel A. Mitchell, Miguel F. Morales, Edward H. Morgan, Stephen M. Ord, Joseph Pathikulangara, Thiagaraj Prabu, Ronald A. Remillard, Timothy Robishaw, Alan E. E. Rogers, Anish A. Roshi, Joseph E. Salah, Robert J. Sault, N. Udaya Shankar, K. S. Srivani, Jamie B. Stevens, Ravi Subrahmanyan, Steven J. Tingay, Randall B. Wayth, Mark Waterson, Rachel L. Webster, Alan R. Whitney, Andrew J. Williams, Christopher L. Williams, and J. Stuart B. Wyithe. Science with the Murchison Widefield Array. *Publications of the Astronomical Society of Australia*, 30:e031, April 2013.

[10] M. R. Calabretta and E. W. Greisen. Representations of celestial coordinates in FITS. *Astronomy & Astrophysics*, 395:1077–1122, December 2002.

[11] P. A. Carroll. *Modelling the Extragalactic Epoch of Reionization Foreground*. PhD thesis, University of Washington, 2016.

[12] P. A. Carroll, J. Line, M. F. Morales, N. Barry, A. P. Beardsley, B. J. Hazelton, D. C. Jacobs, J. C. Pober, I. S. Sullivan, R. L. Webster, G. Bernardi, J. D. Bowman, F. Briggs, R. J. Cappallo, B. E. Corey, A. de Oliveira-Costa, J. 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, HS. 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, D. Oberoi, A. R. Offringa, S. M. Ord, S. Paul, B. Pindor, T. Prabu, P. Procopio, J. Riding, A. E. E. Rogers, A. Roshi, N. Udaya Shankar, S. K. Sethi, K. S. Srivani, R. Subrahmanyan, M. Tegmark, Nithyanandan Thyagarajan, S. J. Tingay, C. M. Trott, M. Waterson, R. B. Wayth, A. R. Whitney, A. Williams, C. L. Williams, C. Wu, and J. S. B. Wyithe. A high reliability survey of discrete Epoch of Reionization foreground sources in the MWA EoR0 field. *Monthly Notices of the Royal Astronomical Society*, 461(4):4151–4175, July 2016.

[13] Joseph Caruana, Andrew J. Bunker, Stephen M. Wilkins, Elizabeth R. Stanway, Mark Lacy, Matt J. Jarvis, Silvio Lorenzoni, and Samantha Hickey. No evidence for Lyman emission in spectroscopy of z $>$ 7 candidate galaxies. *Monthly Notices of the Royal Astronomical Society*, 427(4):3055–3070, December 2012.

[14] Joseph Caruana, Andrew J. Bunker, Stephen M. Wilkins, Elizabeth R. Stanway, Silvio Lorenzoni, Matt J. Jarvis, and Holly Ebert. Spectroscopy of z $\sim$ 7 candidate galaxies: using Lyman α to constrain the neutral fraction of hydrogen in the high-redshift universe. *Monthly Notices of the Royal Astronomical Society*, 443(4):2831–2842, October 2014.

[15] T. J. Cornwell, K. Golap, and S. Bhatnagar. The Noncoplanar Baselines Effect in Radio Interferometry: The W-Projection Algorithm. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):647–657, October 2008.

[16] Glen Cowan. *Statistical Data Analysis*. Oxford University Press, Oxford, New York, June 1998.

[17] A. Datta, S. Bhatnagar, and C. L. Carilli. Detection of Signals from Cosmic Reionization Using Radio Interferometric Signal Processing. *The Astrophysical Journal*, 703(2):1851, September 2009.

[18] A. Datta, J. D. Bowman, and C. L. Carilli. Bright Source Subtraction Requirements For Redshifted 21 cm Measurements. *The Astrophysical Journal*, 724(1):526–538, November 2010.

[19] Mark Dijkstra. Lyα Emitting Galaxies as a Probe of Reionisation. *Publications of the Astronomical Society of Australia*, 31:e040, October 2014.

[20] Joshua S. Dillon, Adrian Liu, and Max Tegmark. A fast method for power spectrum and foreground analysis for 21 cm cosmology. *Physical Review D*, 87(4):043005, February 2013.

[21] Joshua S. Dillon, Adrian Liu, Christopher L. Williams, Jacqueline N. Hewitt, Max Tegmark, Edward H. Morgan, Alan M. Levine, Miguel F. Morales, Steven J. Tingay, Gianni Bernardi, Judd D. Bowman, Frank H. Briggs, Roger C. Cappallo, David Emrich, Daniel A. Mitchell, Divya Oberoi, Thiagaraj Prabu, Randall Wayth, and Rachel L. Webster. Overcoming real-world obstacles in 21 cm power spectrum estimation: A method demonstration and results from early Murchison Widefield Array data. *Physical Review D*, 89(2):023002, January 2014.

[22] Joshua S. Dillon, Abraham R. Neben, Jacqueline N. Hewitt, Max Tegmark, N. Barry, A. P. Beardsley, J. D. Bowman, F. Briggs, P. Carroll, A. de Oliveira-Costa, A. Ewall-Wice, L. Feng, L. J. Greenhill, B. J. Hazelton, L. Hernquist, N. Hurley-Walker, D. C. Jacobs, H. S. Kim, P. Kittiwisit, E. Lenc, J. Line, A. Loeb, B. McKinley, D. A. Mitchell, M. F. Morales, A. R. Offringa, S. Paul, B. Pindor, J. C. Pober, P. Procopio, J. Riding, S. Sethi, N. Udaya Shankar, R. Subrahmanyan, I. Sullivan, Nithyanandan Thyagarajan, S. J. Tingay, C. Trott, R. B. Wayth, R. L. Webster, S. Wyithe, G. Bernardi, R. J. Cappallo, A. A. Deshpande, M. Johnston-Hollitt, D. L. Kaplan, C. J. Lonsdale, S. R. McWhirter, E. Morgan, D. Oberoi, S. M. Ord, T. Prabu, K. S. Srivani, A. Williams, and C. L. Williams. Empirical covariance modeling for 21 cm power spectrum estimation: A method demonstration and new limits from early Murchison Widefield Array 128-tile data. *Physical Review D*, 91(12):123011, June 2015.

[23] Joshua S. Dillon, Max Tegmark, Adrian Liu, Aaron Ewall-Wice, Jacqueline N. Hewitt, Miguel F. Morales, Abraham R. Neben, Aaron R. Parsons, and Haoxuan Zheng. Mapmaking for precision 21 cm cosmology. *Physical Review D*, 91(2):023002, January 2015.

[24] 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, G. Bernardi, Judd D. Bowman, F. Briggs, R. J. Cappallo, P. Carroll, B. E. Corey, A. de Oliveira-Costa, D. Emrich, L. Feng, B. M. Gaensler, R. Goeke, L. J. Greenhill, B. J. Hazelton, N. Hurley-Walker, M. Johnston-Hollitt, Daniel C. Jacobs, D. L. Kaplan, J. C. Kasper, H. S.

Kim, E. Kratzenberg, E. Lenc, J. Line, C. J. Lonsdale, M. J. Lynch, B. McKinley, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, Nithyanandan Thyagarajan, D. Oberoi, S. M. Ord, S. Paul, B. Pindor, J. C. Pober, T. Prabu, P. Procopio, J. Riding, A. E. E. Rogers, A. Roshi, N. Udaya Shankar, Shiv K. Sethi, K. S. Srivani, R. Subrahmanyan, I. S. Sullivan, S. J. Tingay, C. M. Trott, M. Waterson, R. B. Wayth, R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams, C. Wu, and J. S. B. Wyithe. 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, August 2016.

[25] 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. *The Astrophysical Journal*, 831(2):196, November 2016.

[26] H. I. Ewen and E. M. Purcell. Observation of a Line in the Galactic Radio Spectrum: Radiation from Galactic Hydrogen at 1,420 Mc./sec. *Nature*, 168(4270):356–356, September 1951.

[27] Xiaohui Fan, Michael A. Strauss, Robert H. Becker, Richard L. White, James E. Gunn, Gillian R. Knapp, Gordon T. Richards, Donald P. Schneider, J. Brinkmann, and Masataka Fukugita. Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with z $\sim$ 6 Quasars. II. A Sample of 19 Quasars. *The Astronomical Journal*, 132(1):117, June 2006.

[28] Steven R. Furlanetto, S. Peng Oh, and Frank H. Briggs. Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. *Physics Reports*, 433:181–301, October 2006.

[29] Steven R. Furlanetto, Matias Zaldarriaga, and Lars Hernquist. The effects of reionization on Lyα galaxy surveys. *Monthly Notices of the Royal Astronomical Society*, 365(3):1012–1020, January 2006.

[30] Hisanori Furusawa, Nobunari Kashikawa, Masakazu A. R. Kobayashi, James S. Dunlop, Kazuhiro Shimasaku, Tadafumi Takata, Kazuhiro Sekiguchi, Yoshiaki Naito, Junko Furusawa, Masami Ouchi, Fumiaki Nakata, Naoki Yasuda, Yuki Okura, Yoshiaki Taniguchi, Toru Yamada, Masaru Kajisawa, Johan P. U. Fynbo, and Olivier Le

Fèvre. A New Constraint on the Lyα Fraction of UV Very Bright Galaxies at Redshift 7. *The Astrophysical Journal*, 822(1):46, May 2016.

[31] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *The Astrophysical Journal*, 622:759–771, April 2005.

[32] J. Greiner, T. Krühler, J. P. U. Fynbo, A. Rossi, R. Schwarz, S. Klose, S. Savaglio, N. R. Tanvir, S. McBreen, T. Totani, B. B. Zhang, X. F. Wu, D. Watson, S. D. Barthelmy, A. P. Beardmore, P. Ferrero, N. Gehrels, D. A. Kann, N. Kawai, A. Kpc Yolda, P. Mészáros, B. Milvang-Jensen, S. R. Oates, D. Pierini, P. Schady, K. Toma, P. M. Vreeswijk, A. Yolda, B. Zhang, P. Afonso, K. Aoki, D. N. Burrows, C. Clemens, R. Filgas, Z. Haiman, D. H. Hartmann, G. Hasinger, J. Hjorth, E. Jehin, A. J. Levan, E. W. Liang, D. Malesani, T.-S. Pyo, S. Schulze, G. Szokoly, K. Terada, and K. Wiersema. GRB 080913 at Redshift 6.7. *The Astrophysical Journal*, 693(2):1610, March 2009.

[33] J. P. Hamaker, J. D. Bregman, and R. J. Sault. Understanding radio polarimetry. I. Mathematical foundations. *Astronomy & Astrophysics Supplement Series*, 117:137–147, May 1996.

[34] Fredric J. Harris. On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform. *IEEE Proceedings*, 66:51–83, January 1978.

[35] Bryna J. Hazelton, Miguel F. Morales, and Ian S. Sullivan. The Fundamental Multi-Baseline Mode-Mixing Foreground in 21 cm Epoch of Reionization Observations. *Astrophysical Journal*, 770(2):156, June 2013.

[36] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. *The Astrophysical Journal Supplement Series*, 208(2):19, September 2013.

[37] David W. Hogg. Distance measures in cosmology. *arXiv:astro-ph/9905116*, May 1999. arXiv: astro-ph/9905116.

[38] Paul Horowitz and Carl Sagan. Five years of Project META: an all-sky narrow-band radio search for extraterrestrial signals. *The Astrophysical Journal*, 415:218–235, September 1993.

[39] N. Hurley-Walker, J. R. Callingham, P. J. Hancock, T. M. O. Franzen, L. Hindson, A. D. Kapińska, J. Morgan, A. R. Offringa, R. B. Wayth, C. Wu, Q. Zheng, T. Murphy, M. E. Bell, K. S. Dwarakanath, B. For, B. M. Gaensler, M. Johnston-Hollitt, E. Lenc, P. Procopio, L. Staveley-Smith, R. Ekers, J. D. Bowman, F. Briggs, R. J. Cappallo, A. A. Deshpande, L. Greenhill, B. J. Hazelton, D. L. Kaplan, C. J. Lonsdale, S. R. McWhirter, D. A Mitchell, M. F. Morales, E. Morgan, D. Oberoi, S. M. Ord, T. Prabu, N. Udaya Shankar, K. S. Srivani, R. Subrahmanyan, S. J. Tingay, R. L. Webster, A. Williams, and C. L. Williams. GaLactic and Extragalactic All-sky Murchison Widefield Array (GLEAM) survey–I. A low-frequency extragalactic catalogue. *Monthly Notices of the Royal Astronomical Society*, 464(1):1146–1167, September 2017.

[40] Daniel C. Jacobs, B. J. Hazelton, C. M. Trott, Joshua S. Dillon, B. Pindor, I. S. Sullivan, J. C. Pober, N. Barry, A. P. Beardsley, G. Bernardi, Judd D. Bowman, F. Briggs, R. J. Cappallo, P. Carroll, B. E. Corey, A. de Oliveira-Costa, 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, J. Line, A. Loeb, C. J. Lonsdale, M. J. Lynch, B. McKinley, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, A. R. Neben, N. 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. The Murchison Widefield Array 21 cm Power Spectrum Analysis Methodology. *The Astrophysical Journal*, 825(2):114, July 2016.

[41] 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. *Monthly Notices of the Royal Astronomical Society*, 471(4):3974–3987, November 2017.

[42] Nobunari Kashikawa, Kazuhiro Shimasaku, Yuichi Matsuda, Eiichi Egami, Linhua Jiang, Tohru Nagao, Masami Ouchi, Matthew A. Malkan, Takashi Hattori, Kazuaki Ota, Yoshiaki Taniguchi, Sadanori Okamura, Chun Ly, Masanori Iye, Hisanori Furusawa, Yasuhiro Shioya, Takatoshi Shibuya, Yoshifumi Ishizaki, and Jun Toshikawa. Completing the Census of Lyα Emitters at the Reionization Epoch. *The Astrophysical Journal*, 734(2):119, June 2011.

[43] Akira Konno, Masami Ouchi, Yoshiaki Ono, Kazuhiro Shimasaku, Takatoshi Shibuya, Hisanori Furusawa, Kimihiko Nakajima, Yoshiaki Naito, Rieko Momose, Suraphong Yuma, and Masanori Iye. Accelerated Evolution of the Lyα Luminosity Function at $z \gtrsim 7$ Revealed by the Subaru Ultra-deep Survey for Lyα Emitters at $z = 7.3$. *The Astrophysical Journal*, 797(1):16, November 2014.

[44] E. Lenc, B. M. Gaensler, X. H. Sun, E. M. Sadler, A. G. Willis, N. Barry, A. P. Beardsley, M. E. Bell, G. Bernardi, J. D. Bowman, F. Briggs, J. R. Callingham, R. J. Cappallo, P. Carroll, B. E. Corey, A. de Oliveira-Costa, A. A. Deshpande, J. S. Dillon, K. S. Dwarkanath, D. Emrich, A. Ewall-Wice, L. Feng, B.-Q. For, R. Goeke, L. J. Greenhill, P. Hancock, B. J. Hazelton, J. N. Hewitt, L. Hindson, N. Hurley-Walker, M. Johnston-Hollitt, D. C. Jacobs, A. D. Kapińska, D. L. Kaplan, J. C. Kasper, H.-S. Kim, E. Kratzenberg, J. Line, A. Loeb, C. J. Lonsdale, M. J. Lynch, B. McKinley, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, J. Morgan, T. Murphy, A. R. Neben, D. Oberoi, A. R. Offringa, S. M. Ord, S. Paul, B. Pindor, J. C. Pober, T. Prabu, P. Procopio, J. Riding, A. E. E. Rogers, A. Roshi, N. Udaya Shankar, S. K. Sethi, K. S. Srivani, L. Staveley-Smith, R. Subrahmanyan, I. S. Sullivan, M. Tegmark, Nithyanandan Thyagarajan, S. J. Tingay, C. Trott, M. Waterson, R. B. Wayth, R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams, C. Wu, J. S. B. Wyithe, and Q. Zheng. Low-frequency Observations of Linearly Polarized Structures in the Interstellar Medium near the South Galactic Pole. *The Astrophysical Journal*, 830(1):38, October 2016.

[45] Adam Lidz, Oliver Zahn, Matthew McQuinn, Matias Zaldarriaga, and Lars Hernquist. Detecting the Rise and Fall of 21 cm Fluctuations with the Murchison Widefield Array. *The Astrophysical Journal*, 680:962–974, June 2008.

[46] Adrian Liu, Aaron R. Parsons, and Cathryn M. Trott. Epoch of reionization window. I. Mathematical formalism. *Physical Review D*, 90(2):023018, July 2014.

[47] Adrian Liu and Max Tegmark. A method for 21 cm power spectrum estimation in the presence of foregrounds. *Physical Review D*, 83(10):103006, May 2011.

[48] Adrian Liu, Max Tegmark, Scott Morrison, Andrew Lutomirski, and Matias Zaldarriaga. Precision calibration of radio interferometers using redundant baselines. *Monthly Notices of the Royal Astronomical Society*, 408(2):1029–1050, October 2010.

[49] N. R. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39:447–462, February 1976.

[50] Sangeeta Malhotra and James E. Rhoads. Luminosity Functions of Lyα Emitters at Redshifts z = 6.5 and z = 5.7: Evidence against Reionization at z $\leq$ 6.5. *The Astrophysical Journal Letters*, 617(1):L5, November 2004.

[51] Tiziana Di Matteo, Rosalba Perna, Tom Abel, and Martin J. Rees. Radio Foregrounds for the 21 Centimeter Tomography of the Neutral Intergalactic Medium at High Redshifts. *The Astrophysical Journal*, 564(2):576, January 2002.

[52] Ian D. McGreer, Andrei Mesinger, and Xiaohui Fan. The first (nearly) model-independent constraint on the neutral hydrogen fraction at z $\sim$ 5–6. *Monthly Notices of the Royal Astronomical Society*, 415(4):3237–3246, August 2011.

[53] Matthew McQuinn, Adam Lidz, Oliver Zahn, Suvendra Dutta, Lars Hernquist, and Matias Zaldarriaga. The morphology of HII regions during reionization. *Monthly Notices of the Royal Astronomical Society*, 377:1043–1063, May 2007.

[54] Matthew McQuinn and Martin White. On using angular cross-correlations to determine source redshift distributions. *Monthly Notices of the Royal Astronomical Society*, 433(4):2857–2883, August 2013.

[55] Andrei Mesinger. Was reionization complete by z $\sim$ 5–6? *Monthly Notices of the Royal Astronomical Society*, 407(2):1328–1337, September 2010.

[56] D. A. Mitchell, L. J. Greenhill, R. B. Wayth, R. J. Sault, C. J. Lonsdale, R. J. Cappallo, M. F. Morales, and S. M. Ord. Real-Time Calibration of the Murchison Widefield Array. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):707–717, October 2008.

[57] 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.

[58] Miguel F. Morales and Jacqueline Hewitt. Toward Epoch of Reionization Measurements with Wide-Field Radio Observations. *The Astrophysical Journal*, 615(1):7, November 2004.

[59] 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.

[60] Miguel F. Morales and J. Stuart B. Wyithe. Reionization and Cosmology with 21-cm Fluctuations. *Annual Review of Astronomy and Astrophysics*, 48(1):127–171, August 2010.

[61] A. R. Neben, R. F. Bradley, J. N. Hewitt, G. Bernardi, J. D. Bowman, F. Briggs, R. J. Cappallo, A. A. Deshpande, R. Goeke, L. J. Greenhill, B. J. Hazelton, M. Johnston-Hollitt, D. L. Kaplan, C. J. Lonsdale, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, D. Oberoi, S. M. Ord, T. Prabu, N. Udaya Shankar, K. S. Srivani, R. Subrahmanyan, S. J. Tingay, R. B. Wayth, R. L. Webster, A. Williams, and C. L. Williams. Measuring phased-array antenna beampatterns with high dynamic range for

the Murchison Widefield Array using 137 MHz ORBCOMM satellites. *Radio Science*, 50(7):2015RS005678, July 2015.

[62] 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. *The Astrophysical Journal*, 826(2):199, July 2016.

[63] Laura B. Newburgh, Graeme E. Addison, Mandana Amiri, Kevin Bandura, J. Richard Bond, Liam Connor, Jean-François Cliche, Greg Davis, Meiling Deng, Nolan Denman, Matt Dobbs, Mateus Fandino, Heather Fong, Kenneth Gibbs, Adam Gilbert, Elizabeth Griffin, Mark Halpern, David Hanna, Adam D. Hincks, Gary Hinshaw, Carolin Höfer, Peter Klages, Tom Landecker, Kiyoshi Masui, Juan Mena Parra, Ue-Li Pen, Jeff Peterson, Andre Recnik, J. Richard Shaw, Kris Sigurdson, Micheal Sitwell, Graeme Smecher, Rick Smegal, Keith Vanderlinde, and Don Wiebe. Calibrating CHIME: a new radio interferometer to probe dark energy. In *Ground-based and Airborne Telescopes V*, volume 9145, July 2014.

[64] Boon Chong Ng and Chong Meng Samson See. Sensor-array calibration using a maximum-likelihood approach. *IEEE Transactions on Antennas and Propagation*, 44(6):827–835, June 1996.

[65] P. Noorishad, S. J. Wijnholds, A. van Ardenne, and J. M. van der Hulst. Redundancy calibration of phased-array stations. *Astronomy & Astrophysics*, 545:A108, September 2012.

[66] A. R. Offringa, A. G. de Bruyn, S. Zaroubi, G. van Diepen, O. Martinez-Ruby, P. Labropoulos, M. A. Brentjens, B. Ciardi, S. Daiboo, G. Harker, V. Jelić, S. Kazemi, L. V. E. Koopmans, G. Mellema, V. N. Pandey, R. F. Pizzo, J. Schaye, H. Vedantham, V. Veligatla, S. J. Wijnholds, S. Yatawatta, P. Zarka, A. Alexov, J. Anderson, A. Asgekar, M. Avruch, R. Beck, M. Bell, M. R. Bell, M. Bentum, G. Bernardi, P. Best, L. Birzan, A. Bonafede, F. Breitling, J. W. Broderick, M. Brüggen, H. Butcher, J. Conway, M. de Vos, R. J. Dettmar, J. Eisloeffel, H. Falcke, R. Fender, W. Frieswijk, M. Gerbers, J. M. Griessmeier, A. W. Gunst, T. E. Hassall, G. Heald, J. Hessels, M. Hoeft, A. Horneffer, A. Karastergiou, V. Kondratiev, Y. Koopman, M. Kuniyoshi,

G. Kuper, P. Maat, G. Mann, J. McKean, H. Meulman, M. Mevius, J. D. Mol, R. Nijboer, J. Noordam, M. Norden, H. Paas, M. Pandey, R. Pizzo, A. Polatidis, D. Rafferty, S. Rawlings, W. Reich, H. J. A. Röttgering, A. P. Schoenmakers, J. Sluman, O. Smirnov, C. Sobey, B. Stappers, M. Steinmetz, J. Swinbank, M. Tagger, Y. Tang, C. Tasse, A. van Ardenne, W. van Cappellen, A. P. van Duin, M. van Haarlem, J. van Leeuwen, R. J. van Weeren, R. Vermeulen, C. Vocks, R. a. M. J. Wijers, M. Wise, and O. Wucknitz. The LOFAR radio environment. *Astronomy & Astrophysics*, 549:A11, January 2013.

[67] A. R. Offringa, C. M. Trott, N. Hurley-Walker, M. Johnston-Hollitt, B. McKinley, N. Barry, A. P. Beardsley, J. D. Bowman, F. Briggs, P. Carroll, J. S. Dillon, A. Ewall-Wice, L. Feng, B. M. Gaensler, L. J. Greenhill, B. J. Hazelton, J. N. Hewitt, D. C. Jacobs, H.-S. Kim, P. Kittiwisit, E. Lenc, J. Line, A. Loeb, D. A. Mitchell, M. F. Morales, A. R. Neben, S. Paul, B. Pindor, J. C. Pober, P. Procopio, J. Riding, S. K. Sethi, N. U. Shankar, R. Subrahmanyan, I. S. Sullivan, M. Tegmark, N. Thyagarajan, S. J. Tingay, R. B. Wayth, R. L. Webster, and J. S. B. Wyithe. Parametrising Epoch of Reionization foregrounds: A deep survey of low-frequency point-source spectra with the MWA. *Monthly Notices of the Royal Astronomical Society*, 458:1057–1070, May 2016.

[68] A. R. Offringa, R. B. Wayth, N. Hurley-Walker, D. L. Kaplan, N. Barry, A. P. Beardsley, M. E. Bell, G. Bernardi, J. D. Bowman, F. Briggs, J. R. Callingham, R. J. Cappallo, P. Carroll, A. A. Deshpande, J. S. Dillon, K. S. Dwarakanath, A. Ewall-Wice, L. Feng, B.-Q. For, B. M. Gaensler, L. J. Greenhill, P. Hancock, B. J. Hazelton, J. N. Hewitt, L. Hindson, D. C. Jacobs, M. Johnston-Hollitt, A. D. Kapińska, H.-S. Kim, P. Kittiwisit, E. Lenc, J. Line, A. Loeb, C. J. Lonsdale, B. McKinley, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, J. Morgan, A. R. Neben, D. Oberoi, S. M. Ord, S. Paul, B. Pindor, J. C. Pober, T. Prabu, P. Procopio, J. Riding, N. Udaya Shankar, S. Sethi, K. S. Srivani, L. Staveley-Smith, R. Subrahmanyan, I. S. Sullivan, M. Tegmark, N. Thyagarajan, S. J. Tingay, C. M. Trott, R. L. Webster, A. Williams, C. L. Williams, C. Wu, J. S. Wyithe, and Q. Zheng. The Low-Frequency Environment of the Murchison Widefield Array: Radio-Frequency Interference Analysis and Mitigation. *Publications of the Astronomical Society of Australia*, 32:e008, March 2015.

[69] Yoshiaki Ono, Masami Ouchi, Bahram Mobasher, Mark Dickinson, Kyle Penner, Kazuhiro Shimasaku, Benjamin J. Weiner, Jeyhan S. Kartaltepe, Kimihiko Nakajima, Hooshang Nayyeri, Daniel Stern, Nobunari Kashikawa, and Hyron Spinrad. Spectroscopic Confirmation of Three z-dropout Galaxies at z = 6.844–7.213: Demographics of Lyα Emission in z $\sim$ 7 Galaxies. *The Astrophysical Journal*, 744(2):83, December 2012.

[70] S. M. Ord, B. Crosse, D. Emrich, D. Pallot, R. B. Wayth, M. A. Clark, S. E. Tremblay, W. Arcus, D. Barnes, M. Bell, G. Bernardi, N. D. R. Bhat, J. D. Bowman, F. Briggs, J. D. Bunton, R. J. Cappallo, B. E. Corey, A. A. Deshpande, L. deSouza, A. Ewall-Wice, L. Feng, R. Goeke, L. J. Greenhill, B. J. Hazelton, D. Herne, J. N. Hewitt, L. Hindson, N. Hurley-Walker, D. Jacobs, M. Johnston-Hollitt, D. L. Kaplan, J. C. Kasper, B. B. Kincaid, R. Koenig, E. Kratzenberg, N. Kudryavtseva, E. Lenc, C. J. Lonsdale, M. J. Lynch, B. McKinley, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, D. Oberoi, A. Offringa, J. Pathikulangara, B. Pindor, T. Prabu, P. Procopio, R. A. Remillard, J. Riding, 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, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams, and J. S. B. Wyithe. The Murchison Widefield Array Correlator. *Publications of the Astronomical Society of Australia*, 32:e006, March 2015.

[71] S. M. Ord, D. A. Mitchell, R. B. Wayth, L. J. Greenhill, G. Bernardi, S. Gleadow, R. G. Edgar, M. A. Clark, G. Allen, W. Arcus, L. Benkevitch, J. D. Bowman, F. H. Briggs, J. D. Bunton, S. Burns, R. J. Cappallo, W. A. Coles, B. E. Corey, L. deSouza, S. S. Doeleman, M. Derome, A. Deshpande, D. Emrich, R. Goeke, M. R. Gopalakrishna, D. Herne, J. N. Hewitt, P. A. Kamini, D. L. Kaplan, J. C. Kasper, B. B. Kincaid, J. Kocz, E. Kowald, E. Kratzenberg, D. Kumar, C. J. Lonsdale, M. J. Lynch, S. R. McWhirter, S. Madhavi, M. Matejek, M. F. Morales, E. Morgan, D. Oberoi, J. Pathikulangara, T. Prabu, A. E. E. Rogers, A. Roshi, J. E. Salah, A. Schinkel, N. Udaya Shankar, K. S. Srivani, J. Stevens, S. J. Tingay, A. Vaccarella, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, and C. Williams. Interferometric Imaging with the 32 Element Murchison Wide-Field Array. *Publications of the Astronomical Society of the Pacific*, 122(897):1353, October 2010.

[72] Kazuaki Ota, Masanori Iye, Nobunari Kashikawa, Akira Konno, Fumiaki Nakata, Tomonori Totani, Masakazu A. R. Kobayashi, Yoshinobu Fudamoto, Akifumi Seko, Jun Toshikawa, Akie Ichikawa, Takatoshi Shibuya, and Masafusa Onoue. A New Constraint on Reionization from the Evolution of the Lyα Luminosity Function at z $\sim$ 6–7 Probed by a Deep Census of z = 7.0 Lyα Emitter Candidates to 0.3 L*. *The Astrophysical Journal*, 844(1):85, July 2017.

[73] Masami Ouchi, Kazuhiro Shimasaku, Hisanori Furusawa, Tomoki Saito, Makiko Yoshida, Masayuki Akiyama, Yoshiaki Ono, Toru Yamada, Kazuaki Ota, Nobunari Kashikawa, Masanori Iye, Tadayuki Kodama, Sadanori Okamura, Chris Simpson, and Michitoshi Yoshida. Statistics of 207 Lyα Emitters at a Redshift Near 7: Constraints on Reionization and Galaxy Formation Models. *The Astrophysical Journal*, 723(1):869, October 2010.

[74] Aaron Parsons, Jonathan Pober, Matthew McQuinn, Daniel Jacobs, and James

Aguirre. A Sensitivity and Array-configuration Study for Measuring the Power Spectrum of 21 cm Emission from Reionization. *The Astrophysical Journal*, 753(1):81, June 2012.

[75] 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.

[76] Ajinkya H. Patil, Sarod Yatawatta, Saleem Zaroubi, Léon V. E. Koopmans, De Bruyn, A. G, Vibor Jelić, Benedetta Ciardi, Ilian T. Iliev, Maaijke Mevius, Vishambhar N. Pandey, and Bharat K. Gehlot. Systematic biases in low-frequency radio interferometric data due to calibration: the LOFAR-EoR case. *Monthly Notices of the Royal Astronomical Society*, 463(4):4317–4330, December 2016.

[77] Nipanjana Patra, Justin D. Bray, Paul Roberts, and Ron D. Ekers. Bandpass calibration of a wideband spectrometer using coherent pulse injection. *Experimental Astronomy*, 43(2):119–129, April 2017.

[78] Nipanjana Patra, Aaron R. Parsons, David R. DeBoer, Nithyanandan Thyagarajan, Aaron Ewall-Wice, Gilbert Hsyu, Tsz Kuk Leung, Cherie K. Day, Eloy de Lera Acedo, James E. Aguirre, Paul Alexander, Zaki S. Ali, Adam P. Beardsley, Judd D. Bowman, Richard F. Bradley, Chris L. Carilli, Carina Cheng, Joshua S. Dillon, Gcobisa Fadana, Nicolas Fagnoni, Randall Fritz, Steve R. Furlanetto, Brian Glendenning, Bradley Greig, Jasper Grobbelaar, Bryna J. Hazelton, Daniel C. Jacobs, Austin Julius, MacCalvin Kariseb, Saul A. Kohn, Anna Lebedeva, Telalo Lekalake, Adrian Liu, Anita Loots, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary Martinot, Nathan Mathison, Eunice Matsetela, Andrei Mesinger, Miguel F. Morales, Abraham R. Neben, Samantha Pieterse, Jonathan C. Pober, Nima Razavi-Ghods, Jon Ringuette, James Robnett, Kathryn Rosie, Raddwine Sell, Craig Smith, Angelo Syce, Max Tegmark, Peter K. G. Williams, and Haoxuan Zheng. The hydrogen epoch of reionization array dish III: measuring chromaticity of prototype element with reflectometry. *Experimental Astronomy*, March 2018.

[79] Sourabh Paul, Shiv K. Sethi, Miguel F. Morales, K. S. Dwarkanath, N. Udaya Shankar, Ravi Subrahmanyan, N. Barry, A. P. Beardsley, Judd D. Bowman, F. Briggs, P. Carroll, A. De Oliveira-Costa, Joshua S. Dillon, A. Ewall-Wice, L. Feng, L. J. Greenhill, B. M. Gaensler, B. J. Hazelton, J. N. Hewitt, N. Hurley-Walker, D. J. Jacobs, Han-Seek Kim, P. Kittiwisit, E. Lenc, J. Line, A. Loeb, B. Mckinley, D. A. Mitchell, A. R. Neben, A. R. Offringa, B. Pindor, J. C. Pober, P. Procopio, J. Riding, I. S. Sullivan, M. Tegmark, Nithyanandan Thyagarajan, S. J. Tingay, C. M. Trott, R. B. Wayth, R. L. Webster, J. S. B. Wyithe, Roger Cappallo, M. Johnston-Hollitt, D. L. Kaplan,

C. J. Lonsdale, S. R. McWhirter, E. Morgan, D. Oberoi, S. M. Ord, T. Prabu, K. S. Srivani, A. Williams, and C. L. Williams. Delay Spectrum with Phase-Tracking Arrays: Extracting the H I Power Spectrum from the Epoch of Reionization. *Astrophysical Journal*, 833(2):213, December 2016.

[80] P. J. E. Peebles. *Principles of Physical Cosmology.* Princeton University Press, 1993. pp 310–321.

[81] S. Peng Oh and Katherine J. Mack. Foregrounds for 21-cm observations of neutral gas at high redshift. *Monthly Notices of the Royal Astronomical Society*, 346(3):871–877, December 2003.

[82] L. Pentericci, A. Fontana, E. Vanzella, M. Castellano, A. Grazian, M. Dijkstra, K. Boutsia, S. Cristiani, M. Dickinson, E. Giallongo, M. Giavalisco, R. Maiolino, A. Moorwood, D. Paris, and P. Santini. Spectroscopic Confirmation of z $\sim$ 7 Lyman Break Galaxies: Probing the Earliest Galaxies and the Epoch of Reionization. *The Astrophysical Journal*, 743(2):132, November 2011.

[83] L. Pentericci, E. Vanzella, A. Fontana, M. Castellano, T. Treu, A. Mesinger, M. Dijkstra, A. Grazian, M. Bradač, C. Conselice, S. Cristiani, J. Dunlop, A. Galametz, M. Giavalisco, E. Giallongo, A. Koekemoer, R. McLure, R. Maiolino, D. Paris, and P. Santini. New Observations of z $\sim$ 7 Galaxies: Evidence for a Patchy Reionization. *The Astrophysical Journal*, 793(2):113, September 2014.

[84] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *The Astrophysical Journal*, 142:419–421, July 1965.

[85] Planck Collaboration, Adam, R., Aghanim, N., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N., Basak, S., Battye, R., Benabed, K., Bernard, J.-P., Bersanelli, M., Bielewicz, P., Bock, J. J., Bonaldi, A., Bonavera, L., Bond, J. R., Borrill, J., Bouchet, F. R., Boulanger, F., Bucher, M., Burigana, C., Calabrese, E., Cardoso, J.-F., Carron, J., Chiang, H. C., Colombo, L. P. L., Combet, C., Comis, B., Couchot, F., Coulais, A., Crill, B. P., Curto, A., Cuttaia, F., Davis, R. J., de Bernardis, P., de Rosa, A., de Zotti, G., Delabrouille, J., Di Valentino, E., Dickinson, C., Diego, J. M., Doré, O., Douspis, M., Ducout, A., Dupac, X., Elsner, F., Enßlin, T. A., Eriksen, H. K., Falgarone, E., Fantaye, Y., Finelli, F., Forastieri, F., Frailis, M., Fraisse, A. A., Franceschi, E., Frolov, A., Galeotta, S., Galli, S., Ganga, K., Génova-Santos, R. T., Gerbino, M., Ghosh, T., González-Nuevo, J., Górski, K. M., Gruppuso, A., Gudmundsson, J. E., Hansen, F. K., Helou, G., Henrot-Versillé, S., Herranz, D., Hivon, E., Huang, Z., Ilić, S., Jaffe, A. H., Jones, W. C., Keihänen, E., Keskitalo, R., Kisner, T. S., Knox, L., Krachmalnicoff, N., Kunz, M., Kurki-Suonio, H., Lagache, G., Lähteenmki, A., Lamarre, J.-M., Langer, M., Lasenby, A., Lattanzi,

M., Lawrence, C. R., Le Jeune, M., Levrier, F., Lewis, A., Liguori, M., Lilje, P. B., López-Caniego, M., Ma, Y.-Z., Macías-Pérez, J. F., Maggio, G., Mangilli, A., Maris, M., Martin, P. G., Martínez-González, E., Matarrese, S., Mauri, N., McEwen, J. D., Meinhold, P. R., Melchiorri, A., Mennella, A., Migliaccio, M., Miville-Deschênes, M.-A., Molinari, D., Moneti, A., Montier, L., Morgante, G., Moss, A., Naselsky, P., Natoli, P., Oxborrow, C. A., Pagano, L., Paoletti, D., Partridge, B., Patanchon, G., Patrizii, L., Perdereau, O., Perotto, L., Pettorino, V., Piacentini, F., Plaszczynski, S., Polastri, L., Polenta, G., Puget, J.-L., Rachen, J. P., Racine, B., Reinecke, M., Remazeilles, M., Renzi, A., Rocha, G., Rossetti, M., Roudier, G., Rubiño-Martín, J. A., Ruiz-Granados, B., Salvati, L., Sandri, M., Savelainen, M., Scott, D., Sirri, G., Sunyaev, R., Suur-Uski, A.-S., Tauber, J. A., Tenti, M., Toffolatti, L., Tomasi, M., Tristram, M., Trombetti, T., Valiviita, J., Van Tent, F., Vielva, P., Villa, F., Vittorio, N., Wandelt, B. D., Wehus, I. K., White, M., Zacchei, A., and Zonca, A. Planck intermediate results - XLVII. Planck constraints on reionization history. *Astronomy & Astrophysics*, 596:A108, December 2016.

[86] Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., Bartlett, J. G., Bartolo, N., Battaner, E., Battye, R., Benabed, K., Benoît, A., Benoit-Lévy, A., Bernard, J.-P., Bersanelli, M., Bielewicz, P., Bock, J. J., Bonaldi, A., Bonavera, L., Bond, J. R., Borrill, J., Bouchet, F. R., Boulanger, F., Bucher, M., Burigana, C., Butler, R. C., Calabrese, E., Cardoso, J.-F., Catalano, A., Challinor, A., Chamballu, A., Chary, R.-R., Chiang, H. C., Chluba, J., Christensen, P. R., Church, S., Clements, D. L., Colombi, S., Colombo, L. P. L., Combet, C., Coulais, A., Crill, B. P., Curto, A., Cuttaia, F., Danese, L., Davies, R. D., Davis, R. J., de Bernardis, P., de Rosa, A., de Zotti, G., Delabrouille, J., Désert, F.-X., Di Valentino, E., Dickinson, C., Diego, J. M., Dolag, K., Dole, H., Donzelli, S., Doré, O., Douspis, M., Ducout, A., Dunkley, J., Dupac, X., Efstathiou, G., Elsner, F., Enßlin, T. A., Eriksen, H. K., Farhang, M., Fergusson, J., Finelli, F., Forni, O., Frailis, M., Fraisse, A. A., Franceschi, E., Frejsel, A., Galeotta, S., Galli, S., Ganga, K., Gauthier, C., Gerbino, M., Ghosh, T., Giard, M., Giraud-Héraud, Y., Giusarma, E., Gjerløw, E., González-Nuevo, J., Górski, K. M., Gratton, S., Gregorio, A., Gruppuso, A., Gudmundsson, J. E., Hamann, J., Hansen, F. K., Hanson, D., Harrison, D. L., Helou, G., Henrot-Versillé, S., Hernández-Monteagudo, C., Herranz, D., Hildebrandt, S. R., Hivon, E., Hobson, M., Holmes, W. A., Hornstrup, A., Hovest, W., Huang, Z., Huffenberger, K. M., Hurier, G., Jaffe, A. H., Jaffe, T. R., Jones, W. C., Juvela, M., Keihänen, E., Keskitalo, R., Kisner, T. S., Kneissl, R., Knoche, J., Knox, L., Kunz, M., Kurki-Suonio, H., Lagache, G., Lähteenmäki, A., Lamarre, J.-M., Lasenby, A., Lattanzi, M., Lawrence, C. R., Leahy, J. P., Leonardi, R., Lesgourgues, J., Levrier, F., Lewis, A., Liguori, M., Lilje, P. B., Linden-Vørnle, M., López-Caniego, M., Lubin, P. M., Macías-Pérez, J. F., Maggio, G., Maino, D., Mandolesi, N., Mangilli, A., Marchini, A., Maris, M., Martin, P. G., Martinelli, M., Martínez-González, E.,

Masi, S., Matarrese, S., McGehee, P., Meinhold, P. R., Melchiorri, A., Melin, J.-B., Mendes, L., Mennella, A., Migliaccio, M., Millea, M., Mitra, S., Miville-Deschênes, M.-A., Moneti, A., Montier, L., Morgante, G., Mortlock, D., Moss, A., Munshi, D., Murphy, J. A., Naselsky, P., Nati, F., Natoli, P., Netterfield, C. B., Nørgaard-Nielsen, H. U., Noviello, F., Novikov, D., Novikov, I., Oxborrow, C. A., Paci, F., Pagano, L., Pajot, F., Paladini, R., Paoletti, D., Partridge, B., Pasian, F., Patanchon, G., Pearson, T. J., Perdereau, O., Perotto, L., Perrotta, F., Pettorino, V., Piacentini, F., Piat, M., Pierpaoli, E., Pietrobon, D., Plaszczynski, S., Pointecouteau, E., Polenta, G., Popa, L., Pratt, G. W., Prézeau, G., Prunet, S., Puget, J.-L., Rachen, J. P., Reach, W. T., Rebolo, R., Reinecke, M., Remazeilles, M., Renault, C., Renzi, A., Ristorcelli, I., Rocha, G., Rosset, C., Rossetti, M., Roudier, G., Rouillé dOrfeuil, B., Rowan-Robinson, M., Rubiño-Martín, J. A., Rusholme, B., Said, N., Salvatelli, V., Salvati, L., Sandri, M., Santos, D., Savelainen, M., Savini, G., Scott, D., Seiffert, M. D., Serra, P., Shellard, E. P. S., Spencer, L. D., Spinelli, M., Stolyarov, V., Stompor, R., Sudiwala, R., Sunyaev, R., Sutton, D., Suur-Uski, A.-S., Sygnet, J.-F., Tauber, J. A., Terenzi, L., Toffolatti, L., Tomasi, M., Tristram, M., Trombetti, T., Tucci, M., Tuovinen, J., Türler, M., Umana, G., Valenziano, L., Valiviita, J., Van Tent, F., Vielva, P., Villa, F., Wade, L. A., Wandelt, B. D., Wehus, I. K., White, M., White, S. D. M., Wilkinson, A., Yvon, D., Zacchei, A., and Zonca, A. Planck 2015 results - XIII. Cosmological parameters. *Astronomy & Astrophysics*, 594:A13, December 2016.

[87] Jonathan C. Pober, Aaron R. Parsons, James E. Aguirre, Zaki Ali, Richard F. Bradley, Chris L. Carilli, Dave DeBoer, Matthew Dexter, Nicole E. Gugliucci, Daniel C. Jacobs, Patricia J. Klima, Dave MacMahon, Jason Manley, David F. Moore, Irina I. Stefan, and William P. Walbrugh. Opening the 21 cm Epoch of Reionization Window: Measurements of Foreground Isolation with PAPER. *The Astrophysical Journal Letters*, 768(2):L36, April 2013.

[88] Thiagaraj Prabu, K. S. Srivani, D. Anish Roshi, P. A. Kamini, S. Madhavi, David Emrich, Brian Crosse, Andrew J. Williams, Mark Waterson, Avinash A. Deshpande, N. Udaya Shankar, Ravi Subrahmanyan, Frank H. Briggs, Robert F. Goeke, Steven J. Tingay, Melanie Johnston-Hollitt, Gopalakrishna M. R, Edward H. Morgan, Joseph Pathikulangara, John D. Bunton, Grant Hampson, Christopher Williams, Stephen M. Ord, Randall B. Wayth, Deepak Kumar, Miguel F. Morales, Ludi deSouza, Eric Kratzenberg, D. Pallot, Russell McWhirter, Bryna J. Hazelton, Wayne Arcus, David G. Barnes, Gianni Bernardi, T. Booler, Judd D. Bowman, Roger J. Cappallo, Brian E. Corey, Lincoln J. Greenhill, David Herne, Jacqueline N. Hewitt, David L. Kaplan, Justin C. Kasper, Barton B. Kincaid, Ronald Koenig, Colin J. Lonsdale, Mervyn J. Lynch, Daniel A. Mitchell, Divya Oberoi, Ronald A. Remillard, Alan E. Rogers, Joseph E. Salah, Robert J. Sault, Jamie B. Stevens, S. Tremblay, Rachel L. Webster, Alan R. Whitney, and Stuart B. Wyithe. A digital-receiver for the murchison widefield array. *Experimental Astronomy*, 39(1):73–93, March 2015.

[89] Adam G. Riess, Alexei V. Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M. Garnavich, Ron L. Gilliland, Craig J. Hogan, Saurabh Jha, Robert P. Kirshner, B. Leibundgut, M. M. Phillips, David Reiss, Brian P. Schmidt, Robert A. Schommer, R. Chris Smith, J. Spyromilio, Christopher Stubbs, Nicholas B. Suntzeff, and John Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *The Astronomical Journal*, 116(3):1009, September 1998.

[90] Stefano Salvini and Stefan J. Wijnholds. Fast gain calibration in radio astronomy using alternating direction implicit methods: Analysis and applications. *Astronomy & Astrophysics*, 571:A97, November 2014.

[91] J. D. Scargle. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835–853, December 1982.

[92] Matthew A. Schenker, Richard S. Ellis, Nick P. Konidaris, and Daniel P. Stark. Line-emitting Galaxies beyond a Redshift of 7: An Improved Method for Estimating the Evolving Neutrality of the Intergalactic Medium. *The Astrophysical Journal*, 795(1):20, October 2014.

[93] Matthew A. Schenker, Daniel P. Stark, Richard S. Ellis, Brant E. Robertson, James S. Dunlop, Ross J. McLure, Jean-Paul Kneib, and Johan Richard. Keck Spectroscopy of Faint $3 < z < 8$ Lyman Break Galaxies: Evidence for a Declining Fraction of Emission Line Sources in the Redshift Range $6 < z < 8$. *The Astrophysical Journal*, 744(2):179, December 2012.

[94] Joshua Schroeder, Andrei Mesinger, and Zoltán Haiman. Evidence of GunnPeterson damping wings in high-z quasar spectra: strengthening the case for incomplete reionization at $z \sim 6$–7. *Monthly Notices of the Royal Astronomical Society*, 428(4):3058–3071, February 2013.

[95] D. N. Spergel, R. Bean, O. Doré, M. R. Nolta, C. L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H. V. Peiris, L. Verde, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology. *The Astrophysical Journal Supplement Series*, 170(2):377, June 2007.

[96] Daniel P. Stark, Richard S. Ellis, Kuenley Chiu, Masami Ouchi, and Andrew Bunker. Keck spectroscopy of faint $3 < z < 7$ Lyman break galaxies–I. New constraints on cosmic reionization from the luminosity and redshift-dependent fraction of Lyman α emis-

sion. *Monthly Notices of the Royal Astronomical Society*, 408(3):1628–1648, November 2010.

[97] 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. *The Astrophysical Journal*, 759:17, November 2012.

[98] Woodruff T. Sullivan, III. *Cosmic Noise: A History of Early Radio Astronomy*. Cambridge University Press, November 2009.

[99] A. Sutinjo, J. O'Sullivan, E. Lenc, R. B. Wayth, S. Padhi, P. Hall, and S. J. Tingay. Understanding instrumental Stokes leakage in Murchison Widefield Array polarimetry. *Radio Science*, 50(1):2014RS005517, January 2015.

[100] A. T. Sutinjo and P. J. Hall. Intrinsic Cross-Polarization Ratio of Dual-Linearly Polarized Antennas for Low-Frequency Radio Astronomy. *IEEE Transactions on Antennas and Propagation*, 61(5):2852–2856, May 2013.

[101] E. R. Switzer, T.-C. Chang, K. W. Masui, U.-L. Pen, and T. C. Voytek. Interpreting the Unresolved Intensity of Cosmologically Redshifted Line Radiation. *Astrophysical Journal*, 815(1):51, December 2015.

[102] Greg B. Taylor, C. L. Carilli, Richard A. Perley, and National Radio Astronomy Observatory (U.S.). *Synthesis imaging in radio astronomy II: a collection of lectures from the Sixth NRAO/NMIMT Synthesis Imaging Summer School, held at Socorro, New Mexico, USA, 17–23 June, 1998*. Astronomical Society of the Pacific, 1999.

[103] Nithyanandan Thyagarajan, Aaron R. Parsons, David R. DeBoer, Judd D. Bowman, Aaron M. Ewall-Wice, Abraham R. Neben, and Nipanjana Patra. Effects of Antenna Beam Chromaticity on Redshifted 21 cm Power Spectrum and Implications for Hydrogen Epoch of Reionization Array. *The Astrophysical Journal*, 825(1):9, June 2016.

[104] Nithyanandan Thyagarajan, N. Udaya Shankar, Ravi Subrahmanyan, Wayne Arcus, Gianni Bernardi, Judd D. Bowman, Frank Briggs, John D. Bunton, Roger J. Cappallo, Brian E. Corey, Ludi deSouza, David Emrich, Bryan M. Gaensler, Robert F. Goeke, Lincoln J. Greenhill, Bryna J. Hazelton, David Herne, Jacqueline N. Hewitt, Melanie Johnston-Hollitt, David L. Kaplan, Justin C. Kasper, Barton B. Kincaid, Ronald Koenig, Eric Kratzenberg, Colin J. Lonsdale, Mervyn J. Lynch, S. Russell McWhirter, Daniel A. Mitchell, Miguel F. Morales, Edward H. Morgan, Divya Oberoi, Stephen M. Ord, Joseph Pathikulangara, Ronald A. Remillard, Alan E. E. Rogers, D. Anish Roshi, Joseph E. Salah, Robert J. Sault, K. S. Srivani, Jamie B. Stevens, Prabu Thiagaraj, Steven J. Tingay, Randall B. Wayth, Mark Waterson, Rachel L. Webster, Alan R. Whitney, Andrew J. Williams, Christopher L. Williams, and J. Stuart B. Wyithe. A Study of Fundamental Limitations to Statistical Detection of Redshifted H I from the Epoch of Reionization. *The Astrophysical Journal*, 776(1):6, September 2013.

[105] 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. *Publications of the Astronomical Society of Australia*, 30:e007, January 2013.

[106] S. van der Tol, B. D. Jeffs, and A. J. van der Veen. Self-Calibration for the LOFAR Radio Astronomical Array. *IEEE Transactions on Signal Processing*, 55(9):4497–4510, September 2007.

[107] Tomonori Totani, Kentaro Aoki, Takashi Hattori, and Nobuyuki Kawai. High-precision analyses of Lyα damping wing of gamma-ray bursts in the reionization era: On the controversial results from GRB 130606A at z = 5.91. *Publications of the Astronomical Society of Japan*, 68(1):15, February 2016.

[108] Tomonori Totani, Nobuyuki Kawai, George Kosugi, Kentaro Aoki, Toru Yamada, Masanori Iye, Kouji Ohta, and Takashi Hattori. Implications for Cosmic Reionization from the Optical Afterglow Spectrum of the Gamma-Ray Burst 050904 at z = 6.3. *Publications of the Astronomical Society of Japan*, 58(3):485–498, June 2006.

[109] Tommaso Treu, Michele Trenti, Massimo Stiavelli, Matthew W. Auger, and Larry D. Bradley. Inferences on the Distribution of Lyα Emission of z $\sim$ 7 and z $\sim$ 8 Galaxies. *The Astrophysical Journal*, 747(1):27, February 2012.

[110] C. M. Trott, B. Pindor, P. Procopio, R. B. Wayth, D. A. Mitchell, B. McKinley, S. J. Tingay, N. Barry, A. P. Beardsley, G. Bernardi, Judd D. Bowman, F. Briggs, R. J. Cappallo, P. Carroll, A. de Oliveira-Costa, Joshua S. Dillon, A. Ewall-Wice, L. Feng, L. J. Greenhill, B. J. Hazelton, J. N. Hewitt, N. Hurley-Walker, M. Johnston-Hollitt, Daniel C. Jacobs, D. L. Kaplan, H. S. Kim, E. Lenc, J. Line, A. Loeb, C. J. Lonsdale, M. F. Morales, E. Morgan, A. R. Neben, Nithyanandan Thyagarajan, D. Oberoi, A. R. Offringa, S. M. Ord, S. Paul, J. C. Pober, T. Prabu, J. Riding, N. Udaya Shankar, Shiv K. Sethi, K. S. Srivani, R. Subrahmanyan, I. S. Sullivan, M. Tegmark, R. L. Webster, A. Williams, C. L. Williams, C. Wu, and J. S. B. Wyithe. CHIPS: The Cosmological H I Power Spectrum Estimator. *Astrophysical Journal*, 818(2):139, February 2016.

[111] Cathryn M. Trott and Randall B. Wayth. Spectral Calibration Requirements of Radio Interferometers for Epoch of Reionisation Science with the SKA. *Publications of the Astronomical Society of Australia*, 33:e019, May 2016.

[112] 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.

[113] Harish Vedantham, N. Udaya Shankar, and Ravi Subrahmanyan. Imaging the Epoch of Reionization: Limitations from Foreground Confusion and Imaging Algorithms. *The Astrophysical Journal*, 745(2):176, January 2012.

[114] Mark H. Wieringa. An investigation of the telescope based calibration methods 'redundancy' and 'self-cal'. *Experimental Astronomy*, 2(4):203–225, July 1992.

[115] Stefan J. Wijnholds and Alle-Jan van der Veen. Multisource Self-Calibration for Sensor Arrays. *IEEE Transactions on Signal Processing*, 57(9):3512–3522, May 2009.

[116] Matias Zaldarriaga, Steven R. Furlanetto, and Lars Hernquist. 21 Centimeter Fluctuations from Cosmic Gas at High Redshifts. *The Astrophysical Journal*, 608(2):622, June 2004.

[117] H. Zheng, M. Tegmark, V. Buza, J. S. Dillon, H. Gharibyan, J. Hickish, E. Kunz, A. Liu, J. Losh, A. Lutomirski, S. Morrison, S. Narayanan, A. Perko, D. Rosner, N. Sanchez, K. Schutz, S. M. Tribiano, M. Valdez, H. Yang, K. Zarb Adami, I. Zelko,

K. Zheng, R. P. Armstrong, R. F. Bradley, M. R. Dexter, A. Ewall-Wice, A. Magro, M. Matejek, E. Morgan, A. R. Neben, Q. Pan, R. F. Penna, C. M. Peterson, M. Su, J. Villasenor, C. L. Williams, and Y. Zhu. MITEoR: a scalable interferometer for precision 21 cm cosmology. *Monthly Notices of the Royal Astronomical Society*, 445(2):1084–1103, December 2014.

[118] Daniel Zwillinger and Stephen Kokoska. *CRC Standard Probability and Statistics Tables and Formulae*. CRC Press, Boca Raton, 1 edition, December 1999.

Appendix A
INVERSE COVARIANCE WEIGHTING

Inverse covariance weighting along the frequency transform has the potential to remove contamination due to flagged frequency channels. Harmonic coarse band lines in the power spectrum become a dominant systematic for lower frequency bands [24], and thus we aim to incorporate inverse covariance weighting into εppsilon. The formalism we follow was developed from an analysis of the CMB; it is a quadratic estimator method which estimates the maximum likelihood [8]. This estimator was applied to the EoR interferometer problem in a variety of publications [47, 20, 46, 23], and has been used before to analyze MWA data [21, 22, 110].

We reproduce the basics of this method, both for reference and for clarity. In later publications, many of the assumptions are understated, so we explicitly list all known. We also link multiple papers together, where key interconnecting mathematical steps have been omitted previously.

We provide the mathematical setup for incorporating this analysis into εppsilon. By making the correct assumptions, we can reduce the estimator into a simple and efficient iterator problem. This will help to provide quick testing of inverse covariance weighting along the frequency dimension for future implementation.

A.1 The power estimation equation

Following Bond, Jaffe, and Knox [8], hereafter known as BJK, we require the signal and noise to be independent and zero mean. The data can be described as (Eq. 2.1 of BJK)

$$\Delta_i = s_i + n_i, \tag{A.1}$$

where Δ are the data pixels, s is the signal, n is the noise, and i iterates through every pixel. Our assumption dictates the variance should be (Eq. 2.3 of BJK)

$$\langle \Delta_i \Delta_{i'} \rangle = C_{Tii'} + C_{nii'} \tag{A.2}$$

where $\langle \ldots \rangle$ is the ensemble average, and (Eq. 2.2 of BJK)

$$C_{Tii'} = \langle s_i s_{i'} \rangle \quad C_{nii'} = \langle n_i n_{i'} \rangle. \tag{A.3}$$

Since the theoretical covariance matrix ($C_{Tii'}$) depends on the sky signal, it is also related to the power spectrum. We can decompose the power spectrum to depend on general parameters a_p[1], and with the assumption of Gaussianity, they constitute a full statistical description.

With our assumptions, the likelihood function for the parameters a_p is equal to the probability of the data for given a_p, and can be written as (Eq. 2.9 of BJK)

$$\mathcal{L}_\Delta(a_p) = P(\mathbf{\Delta}|a_p) \propto \exp\left(-\frac{1}{2}\mathbf{\Delta}^T \left(\mathbf{C}_T(a_p) + \mathbf{C}_N\right)^{-1} \mathbf{\Delta}\right). \tag{A.4}$$

In this case, $\mathbf{C}_N$ is the more general noise covariance. Equation A.4 can be maximized with proper choice of a_p. By maximizing the likelihood, the parameters a_p can then be described as the most probable parameters that build the power spectrum, and this is what we would like to do. However, solving for the parameters this way is not computationally feasible.

Instead, we can apply a Gaussian approximation to the log of the likelihood function. This allows us to treat the peak of the log-likelihood function as a Gaussian, which can then be expanded simply as a Taylor expansion about each parameter. We then have (Eq. 2.13 of BJK)

$$\ln \mathcal{L}(a + \delta a) = \ln \mathcal{L}(a) + \sum_p \frac{\partial \ln \mathcal{L}(a)}{\partial a_p} \delta a_p + \frac{1}{2} \sum_{pp'} \frac{\partial^2 \ln \mathcal{L}(a)}{\partial a_p \partial a_{p'}} \delta a_p \delta a_{p'}, \tag{A.5}$$

and by maximizing this function and solving for the variation in the a_p parameter, we get (Eq. 2.14 of BJK)

$$\delta a_p = -\sum_{p'} \left(\frac{\partial^2 \ln \mathcal{L}(a)}{\partial a_p \partial a_{p'}}\right)^{-1} \frac{\partial \ln \mathcal{L}(a)}{\partial a_{p'}}. \tag{A.6}$$

---

[1]For the CMB, the a_p parameters would be the cosmological model, like the baryon density, optical depth, Hubble constant, etc.

Once the variation δa_p is solved for, we modify the original parameter by this amount and Equation A.6 is solved for again with the update. This is iterated until convergence is reached. This methodology is the same as Newton's method, and is much faster than solving for Equation A.4 explicitly.

The single derivative term can be described as (Eq. 2.15 of BJK)

$$\frac{\partial \ln \mathcal{L}(a)}{\partial a_{p'}} = \frac{1}{2}\mathrm{Tr}[\left(\mathbf{\Delta\Delta}^T - \mathbf{C}\right)\left(\mathbf{C}^{-1}\mathbf{C}_{T,p'}\mathbf{C}^{-1}\right)], \tag{A.7}$$

where $\mathbf{C} = \mathbf{C}_T + \mathbf{C}_N$ and $\mathbf{C}_{T,p'}$ is shorthand for $\partial\mathbf{C}_T/\partial a_{p'}$. The second derivative term, called the curvature matrix, is much longer and much more computationally expensive (Eq. 2.16 of BJK):

$$-\frac{\partial^2 \ln \mathcal{L}(a)}{\partial a_p \partial a_{p'}} = \mathrm{Tr}[\left(\mathbf{\Delta\Delta}^T - \mathbf{C}\right)\left(\mathbf{C}^{-1}\mathbf{C}_{T,p}\mathbf{C}^{-1}\mathbf{C}_{T,p'}\mathbf{C}^{-1} - \frac{1}{2}\mathbf{C}^{-1}\mathbf{C}_{T,pp'}\mathbf{C}^{-1}\right)] + \frac{1}{2}\mathrm{Tr}[\mathbf{C}^{-1}\mathbf{C}_{T,p}\mathbf{C}^{-1}\mathbf{C}_{T,p'}]. \tag{A.8}$$

Here is where the quadratic estimator intersects with the maximum likelihood method; we estimate the curvature matrix with the *ensemble average* of the curvature matrix. This significantly reduces the terms in the calculation of the curvature (Eq. 2.17 of BJK):

$$\hat{\mathbf{F}}^{(a)}_{pp'} = \frac{1}{2}\mathrm{Tr}[\mathbf{C}^{-1}\mathbf{C}_{T,p}\mathbf{C}^{-1}\mathbf{C}_{T,p'}]. \tag{A.9}$$

This new quantity, called the Fisher matrix[2] $\hat{\mathbf{F}}^{(a)}_{pp'}$, explicitly assumes that $\langle \Delta\Delta^T \rangle = \mathbf{C}$. In other words, it assumes that the theory describing the data is correct and complete. Our final estimate for the variation is thus (Eq. 2.18 of BJK)

$$\delta a_p = \frac{1}{2}\sum_{p'} (\hat{\mathbf{F}}^{(a)})^{-1}_{pp'} \mathrm{Tr}[\left(\mathbf{\Delta\Delta}^T - \mathbf{C}\right)\left(\mathbf{C}^{-1}\mathbf{C}_{T,p'}\mathbf{C}^{-1}\right)]. \tag{A.10}$$

This estimate is called a quadratic estimator because it is quadratic in the data. While this is an estimate, BJK proves that this formulation will reach the exact location of the likelihood

[2]The notation $\hat{\mathbf{F}}^{(a)}_{pp'}$ has been specifically chosen. The circumflex is introduced to distinguish from a Fourier transform, and a Fisher matrix is calculated for each (a).

peak. The maximum likelihood parameters can be obtained directly, however the errors will incorporate our assumptions and differ from their true value. BJK notes that after the value of the maximum likelihood is obtained through iteration, the full curvature matrix formalism can be used to obtain correct errors.

This equation is the starting point for much of the work regarding inverse covariance weighting with the MWA. The assumptions that we have made thus far are:

- Signal and noise are independent with zero mean.
- The data are Gaussian.
- The log-likelihood function is Gaussian near its maximum and the starting point is near the maximum.
- The curvature matrix can be approximated by its ensemble average.
- Our theory is correct and complete, or fairly close.
- Equation A.10 will be iterated over until convergence is reached.

A.1.1 Linking to other implementations

We can rewrite Equation A.10 by expanding the trace:

$$\delta a_p = \frac{1}{2}\sum_{p'}(\hat{\mathbf{F}}^{(a)})^{-1}_{pp'}\left(\mathrm{Tr}[\boldsymbol{\Delta}\boldsymbol{\Delta}^T\mathbf{C}^{-1}\mathbf{C}_{T,p'}\mathbf{C}^{-1}] - \mathrm{Tr}[\mathbf{C}\mathbf{C}^{-1}\mathbf{C}_{T,p'}\mathbf{C}^{-1}]\right). \tag{A.11}$$

The first term can then be written as

$$\mathrm{Tr}[\boldsymbol{\Delta}\boldsymbol{\Delta}^T\mathbf{C}^{-1}\mathbf{C}_{T,p'}\mathbf{C}^{-1}] = \boldsymbol{\Delta}^T\mathbf{C}^{-1}\mathbf{C}_{T,p'}\mathbf{C}^{-1}\boldsymbol{\Delta}, \tag{A.12}$$

using the cyclic properties of the trace. A similar trick can be imposed on the second term,

$$\mathrm{Tr}[\mathbf{C}\mathbf{C}^{-1}\mathbf{C}_{T,p'}\mathbf{C}^{-1}] = \mathrm{Tr}[\mathbf{C}^{-1}\mathbf{C}_{T,p'}]. \tag{A.13}$$

Now, the quadratic power estimator is

$$\delta a_p = \frac{1}{2}\sum_{p'}(\hat{\mathbf{F}}^{(a)})^{-1}_{pp'}\left(\boldsymbol{\Delta}^T\mathbf{C}^{-1}\mathbf{C}_{T,p'}\mathbf{C}^{-1}\boldsymbol{\Delta} - \mathrm{Tr}[\mathbf{C}^{-1}\mathbf{C}_{T,p'}]\right), \tag{A.14}$$

which is similar to an implementation used in [54].

The last term can further be reduced if the theory is simple enough. If the signal theory covariance is a series of independent parameters along the diagonal, the derivative with respect to a particular $a_{p'}$ is simple; it is a matrix with one non-zero value along the diagonal,

$$\frac{\partial}{\partial a_1}\begin{bmatrix} a_0 & \dots & \dots & \dots & 0 \\ \vdots & a_1 & & & \vdots \\ \vdots & & a_2 & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & \dots & \dots & a_n \end{bmatrix} = \begin{bmatrix} 0 & \dots & \dots & \dots & 0 \\ \vdots & 1 & & & \vdots \\ \vdots & & 0 & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 \end{bmatrix}. \tag{A.15}$$

Therefore, the last term simplifies into a constant, giving

$$\delta a_p = \frac{1}{2}\sum_{p'}(\hat{\mathbf{F}}^{(a)})^{-1}_{pp'}\left(\boldsymbol{\Delta}^T\mathbf{C}^{-1}\mathbf{C}_{T,p'}\mathbf{C}^{-1}\boldsymbol{\Delta} - b_{p'}\right). \tag{A.16}$$

This implementation is similar to many recent publications, including [47, 20, 46, 22, 23, 110], among others. In some cases, this last term is emitted altogether.

In many of these implementations, the power estimator is only iterated once. Technically, if all of our assumptions are truly correct, then it will converge in one step. However, the ability to iterate *allows* us to relax our assumptions. We expect our data to need more than one iteration, and thus will continue to treat it as an iterative problem.

A.1.2 Choice of normalization

There can be a choice in the weighting and normalization of the quadratic estimator through a window function. This determines the properties of the error bars, which have some flexibility.

In εppsilon, we have previously made the assumption that errors are uncorrelated in our error propagation (see §3.3), thus we will choose a window function that gives us uncorrelated errors. [20] states this to be (their Eq. 19)

$$\mathbf{M}^{\alpha\beta} = \frac{\left(\hat{\mathbf{F}}^{-1/2}\right)^{\alpha\beta}}{\sum_\gamma \left(\hat{\mathbf{F}}^{1/2}\right)^{\alpha\gamma}}. \tag{A.17}$$

The equation $\mathbf{M}$ replaces the inverse Fisher matrix in the quadratic power estimator to create a weighted quadratic power estimator:

$$\delta\hat{a}_p \propto \frac{1}{2}\mathbf{M}^{pp'}\left(\mathbf{\Delta}^T\mathbf{C}^{-1}\mathbf{C}_{T,p'}\mathbf{C}^{-1}\mathbf{\Delta} - b_{p'}\right). \tag{A.18}$$

A.2 Implementation

We attempt to apply this power estimation to our pipeline for the first time. As such, we will be very explicit about our chosen starting point in the power estimation iterator. We start by constructing our covariance matrices. Due to spectral smoothness of the foregrounds and ease in its calculation, the foreground covariance matrix is constructed in η-space. We estimate the value of the foregrounds as the maximum of the calculated σ^2:

$$\mathbf{C}_T(\eta,\eta) = \begin{bmatrix} \max(\sigma^2) & \dots & \dots & \dots & 0 \\ \vdots & 0 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 \end{bmatrix}. \tag{A.19}$$

This matrix is the initial guess in our iterator; $\mathbf{C}_T(\eta,\eta)$ will be modified by $\delta\hat{a}_p$ to generate better estimates for each iteration.

In order to take $\mathbf{C}_T(\eta,\eta)$ into a space that can be applied to data, we will need Fourier

transform matrices. An FFT is taken along every column of an identity matrix,

$$\mathbf{F}(k_z, z) = N_f \Delta z \begin{bmatrix} \mathrm{FFT}(1 & 0 & \dots & \dots & 0) \\ \mathrm{FFT}(0 & 1 & \dots & \dots & 0) \\ \mathrm{FFT}(0 & \vdots & \ddots & & \vdots \\ \mathrm{FFT}(0 & \vdots & & \ddots & \vdots \\ \mathrm{FFT}(0 & 0 & \dots & \dots & 1) \end{bmatrix}^T , \tag{A.20}$$

where N_f is the number of frequencies (to normalize an IDL[3] forward Fourier transform) and Δz is the redshift channel width. The same is done for the inverse Fourier transform matrix,

$$\mathbf{F}^{-1}(z, k_z) = \Delta k_z \begin{bmatrix} \mathrm{FFT}^{-1}(1 & 0 & \dots & \dots & 0) \\ \mathrm{FFT}^{-1}(0 & 1 & \dots & \dots & 0) \\ \mathrm{FFT}^{-1}(0 & \vdots & \ddots & & \vdots \\ \mathrm{FFT}^{-1}(0 & \vdots & & \ddots & \vdots \\ \mathrm{FFT}^{-1}(0 & 0 & \dots & \dots & 1) \end{bmatrix}^T , \tag{A.21}$$

where Δk_z is the k_z channel width. The foreground covariance matrix in *redshift* space can now be constructed with the Fourier transform matrices,

$$\mathbf{C}_T(z, z) = \mathbf{F}^{-1}(z, \eta)\mathbf{C}_T(\eta, \eta)\mathbf{F}^{-1\dagger}(\eta, z). \tag{A.22}$$

The Fourier transform is unitary: $\mathbf{F}^\dagger = \mathbf{F}^{-1}$, thus the use of $\mathbf{F}^{-1\dagger}$ appears redundant. While they are in theory unitary, in practice the units break the unitary nature by a factor. Therefore, it is important to be explicit and pass around units in the Fourier transform matrices.

In contrast to the foreground covariance matrix, we construct the instrument covariance matrix in frequency space. A frequency vector of calculated σ^2 is formed into a diagonal

[3]Interactive Data Language, the native code for εppsilon.

matrix:

$$\mathbf{C}_I(z,z) = \begin{bmatrix} \sigma_0^2 & \dots & \dots & \dots & 0 \\ \vdots & \sigma_1^2 & & & \vdots \\ \vdots & & \sigma_2^2 & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \sigma_n^2 \end{bmatrix}. \tag{A.23}$$

Any missing data is filled with a large σ^2 in order to minimize its contribution. Each pixel has its own associated $\mathbf{C}_I(z,z)$ matrix. The full covariance matrix in redshift space is

$$\mathbf{C}(z,z) = \mathbf{C}_T(z,z) + \mathbf{C}_I(z,z), \tag{A.24}$$

where each pixel has its own associated covariance matrix.

The inverse of the covariance matrix is made using pre-defined inverse functions. The inverse covariance matrix for $\{k_z, k_z\}$-space is made through Fourier transforms:

$$\mathbf{C}^{-1}(k_z,k_z) = \mathbf{F}(k_z,z)\mathbf{C}^{-1}(z,z)\mathbf{F}^{\dagger}(z,k_z). \tag{A.25}$$

Now with estimates for the covariance matrix and inverse covariance matrix in k_z-space, we have all the tools necessary to start calculating various components of the power estimator.

A.2.1 Fisher and $\mathbf{M}$ *matrices*

In Equation A.9, we describe the general Fisher matrix, $\hat{\mathbf{F}}^{(a)}_{pp'} = \frac{1}{2}\mathrm{Tr}[\mathbf{C}^{-1}\mathbf{C}_{T,p}\mathbf{C}^{-1}\mathbf{C}_{T,p'}]$. If we choose to construct the Fisher matrix in η-space, it is a straightforward calculation.

As shown in Equation A.15, $\mathbf{C}_{T,p}$ is a matrix of one non-zero value. Thus, when we calculate $\mathbf{C}^{-1}\mathbf{C}_{T,p}\mathbf{C}^{-1}\mathbf{C}_{T,p'}$, we get a matrix which is squared along the diagonal. Non-diagonal elements for indices pp' are equal to elements in $p'p$. Using the symmetric nature of the covariance matrix, this can be simplified to the square of each element.

We condense the Fisher matrix to

$$\hat{\mathbf{F}}^{(a)}(k_z,k_z) = \frac{1}{2}|\mathbf{C}^{-1}(k_z,k_z)|^2. \tag{A.26}$$

In Equation A.17, we described the weighted Fisher matrix, or the $\mathbf{M}$ matrix, that we would like to use. It involves both the matrix square root and the inverse of the Fisher matrix, which we use pre-defined code libraries to obtain.

A.2.2 Condensing the power estimator

We would like to calculate the power estimator for the cross power spectrum, or the power generated from the even set crossed with the odd set (see §3.3.4). Equation A.18 thus changes to

$$\delta\hat{\mathbf{a}}(k_z) \propto \frac{1}{2}\mathbf{M}(k_z,k_z)\left(\mathbf{e}^T(k_z)\mathbf{C}^{-1}(k_z,k_z)\mathbf{C}_{T,p'}(k_z,k_z)\mathbf{C}^{-1}(k_z,k_z)\mathbf{o}(k_z)-\mathbf{b}(k_z)\right), \quad \text{(A.27)}$$

where $\mathbf{o}$ is the vector of the odd data set and $\mathbf{e}$ is the vector of the even data set. We have removed mention of pp' indices wherever possible and have replaced them with explicit function arguments. We can further reduce this equation if we construct weighted data

$$\hat{\mathbf{e}}(k_z)=\mathbf{C}^{-1}(k_z,k_z)\mathbf{e}(k_z) \quad \hat{\mathbf{o}}(k_z)=\mathbf{C}^{-1}(k_z,k_z)\mathbf{o}(k_z). \quad \text{(A.28)}$$

We can then describe $\mathbf{C}_{T,p'}\mathbf{C}^{-1}\mathbf{o}$ as $\hat{\mathbf{o}}_{p'}$. Further, if we use the fact that $\mathbf{C}^{-1}$ is symmetric (equal to its transpose) and that $\mathbf{A}^T\mathbf{B}^T=(\mathbf{BA})^T$, we can reduce even more:

$$\mathbf{e}^T\mathbf{C}^{-1}\hat{\mathbf{o}}_{p'}=\hat{\mathbf{o}}_{p'}^T\hat{\mathbf{e}}. \quad \text{(A.29)}$$

This notation is understandably confusing. We have two vectors in a multiplication, where in one we have chosen a particular element. However, this does come naturally out of our reductions. Fortunately, we can rewrite Equation A.29 as the particular element of the Hadamard product: $\hat{\mathbf{o}}_{p'}^T\hat{\mathbf{e}}=(\hat{\mathbf{o}}\circ\hat{\mathbf{e}})_{p'}$. Our final notation is thus

$$\delta\hat{\mathbf{a}}(k_z) \propto \frac{1}{2}\mathbf{M}(k_z,k_z)\big(\hat{\mathbf{o}}(k_z)\circ\hat{\mathbf{e}}(k_z)-\mathbf{b}(k_z)\big). \quad \text{(A.30)}$$

In practice, we replace the direct even–odd cross power with the even–odd sums and differences, as done in §3.3.4.

The quadratic power estimator can be a daunting task to implement. However, if the theory is simple enough, the final estimator can actually be relatively concise. We plan to test this in full with simulation and real data in the future.

VITA

Nichole Barry was born in Roseburg, Oregon to Lisa and David Barry. She lived in Roseburg, OR; Vale, OR; Cottonwood, CA; and Burney, CA with her parents and brother, Justin. She went to University of California, Davis in the Integrated Studies Honors Program where she majored in Physics. In her time at UC Davis, she worked at Crocker Nuclear Laboratory, DELTA Group, and the Safeway deli in Burney, in addition to performing research for the LUX dark matter experiment with Professor Mani Tripathi. She then married Marshall Styczinski, and they both attended graduate school at the University of Washington. She joined Professor Miguel Morales' radio cosmology group her second year, where the topic of calibration was supposed to be a two month project. In addition to Physics, she is active within the Astrobiology community. She will continue working with MWA data as a postdoctoral research at the University of Melbourne in Australia with Professor Rachel Webster.

In her free time, Nichole likes to kayak, rock climb, run, and hike. She and Marshall have hiked in the Canadian Rockies, Sierra Nevadas, Cascades, Olympic Pennisula, Coromandal Range, Arthur Range, and the Dolomites. She has had three fuzzy children: Peanut, Sydney, and Nutmeg.