Improving 21-cm Epoch of Reionization Power Spectrum Limits by Characterizing and Mitigating Radio Frequency Interference

Michael Wilensky

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

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Miguel Morales, Chair

Matthew McQuinn

Paul Wiggins

Program Authorized to Offer Degree:
Physics

University of Washington

Abstract

Improving 21-cm Epoch of Reionization Power Spectrum Limits by Characterizing and Mitigating Radio Frequency Interference

Michael Wilensky

Chair of the Supervisory Committee:
Professor Miguel Morales
Physics

The Epoch of Reionization (EoR) is the period in which the universe's hydrogen transformed from being predominantly neutral to being predominantly ionized. An extremely promising way to explore the EoR is by making use of the 21-cm line from the hyperfine transition in neutral Hydrogen. Ultimately we seek 3D maps of the intergalactic medium throughout this transition. However, preliminary efforts will focus on constraining the spatial power spectra of these maps at various cosmological redshifts.

Redshifted 21-cm radiation can be observed using modern radio telescopes. The expected signal from the EoR is extremely faint. There are numerous systematic effects that, left unchecked, will totally dominate the EoR signal in the power spectrum. One of these effects, which is the primary subject of this thesis, is radio frequency interference (RFI). RFI is essentially any unwanted non-astrophysical signal observed by the instrument. We develop a new method for finding RFI in radio interferometers, called SSINS, which finds significantly fainter RFI than previous methods. Based on semi-analytic calculations and simulations, we place limits on total allowable apparent RFI flux density in the final power spectrum integration at roughly 1 mJy for relatively optimistic EoR models. We also explore the effects of RFI flagging on power spectrum measurements, finding that it can produce deleterious excess power in much the same way that observed RFI does. These results suggest that

current RFI flagging strategies must be totally reconsidered to preserve the dynamic range of the power spectrum measurement. By examining a season of data from the Murchison Widefield Array, we expose residual RFI undetected by SSINS that contaminates power spectra at a meaningful level. Contaminated power spectra usually contain observations that already contain flags from SSINS, demonstrating the capacity for SSINS to improve power spectrum measurements. Finally, we draw an upper limit on the EoR power spectrum. Our lowest upper limit is $\Delta^2 \leq 1.61 \cdot 10^4$ mK^2 at $k = 0.258$ h Mpc^{-1} at a redshift of 7.1.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

AOFLAGGER: : Default RFI flagger in the MWA pre-processing pipeline, developed by André Offringa.

BANDPASS: : The frequency response of a receiving element.

BASELINE: : Refers to a pair of antennas that jointly report visibilities, or the vector displacement between them.

CALIBRATION: : The act of adjusting the raw data to account for various instrumental effects in the data acquisition pipeline, such as cable reflection modes.

CHROMATIC: : Frequency-dependent.

DTV: : Digital television.

DAB: : Digital audio broadcasting.

MWA: : Murchison Widefield Array, a radio interferometer in Western Australia.

HERA: : Hydrogen Epoch of Reionization Array, a radio interferometer in South Africa.

EOR: : Epoch of Reionization. A cosmological phase transition spanning hundreds of millions of years in which the Hydrogen content of the universe changed from being almost entirely neutral to being almost entirely ionized.

EOR WINDOW: : A region of the cylindrically averaged power spectrum that is free of foreground contamination.

εPPSILON: : Error Propagated Power Spectrum with Interleaved Observed Noise. Power spectrum estimation software employed in this work.

FHD: : Fast Holographic Deconvolution. Calibration and imaging software employed as a part of the power spectrum estimation pipeline in this work.

FLAGS: : A bit mask for the data determining which samples are to be used and which are to be excluded in the analysis.

FOREGROUNDS: : Astrophysical sources between us and the EoR.

FOREGROUND WEDGE: : A region of the cylindrically averaged power spectrum to which foreground contamination is isolated.

FOURIER TRANSFORM: : The act of decomposing a function into a superposition of plane wave basis functions. Named after Joseph Fourier.

GLEAM SURVEY: : Galactic and Extragalactic All-sky MWA Survey. A point-source catalog employed as a sky model throughout this work.

IMAGING: : Synthesizing an image of the sky from the visibilities.

INTERFEROMETER: : A class of instruments whose reported measurements are an intentional combination of two or more signals.

JY: : Jansky. Unit of spectral flux density. Equivalent to 10^{-26} W m^{-2} Hz^{-1}.

REDSHIFT: : A factor by which radiation is lengthened due to some effect. Cosmological redshift refers to that redshift arising from the expansion of spacetime.

RFI: : Radio Frequency Interference. Any undesired non-astrophysical radio signal measured by the instrument. Usually generated by people.

PSF: : Point-spread Function. Determines how point-sources appear in a synthesized image.

SSINS: : Sky-Subtracted Incoherent Noise Spectra. A faint RFI examination and excision pipeline developed and presented in this work.

VISIBILITY: : A cross-correlation of two electric field measurements from the receiving elements. Typically reported as a function of frequency, rather than lag.

ACKNOWLEDGMENTS

First, while I am the sole author of this manuscript, I want it to be understood that the work presented here would have been impossible without the help of others. All those in my life have played a role, whether it be big, small, kind, or mean. It would be impossible and useless to list every one of my personal encounters, so I will select a few to mention. If you are a friend or family member of mine reading this and do not find your name here, I apologize.[1] It is not that you are unimportant to me; my finitude has failed you.

My foremost acknowledgements go to my mother, Sara Napoli, who bore me into this world and patiently nurtured me, cheered for me, and remained my guardian despite my greatly misguided behavior. I also would like to acknowledge her husband, Anthony Napoli, who cares for my mother while she cares for everyone else. I deeply acknowledge my sister, Lissette Wilensky, who has been my greatest mentor and who rescued me from insanity by reminding me of the probability of simpler explanations. I also acknowledge my extended family, including the Wilenskys, the Bogens, and the Napolis, for never being unwelcoming.

Next, I would like to acknowledge my advisor, Professor Miguel Morales, for patiently endowing me with the knowledge and guidance I needed to succeed in this endeavor, despite that it might have been an incredibly repetitive exercise. I would also like to acknowledge my unofficial secondary advisor, Dr. Bryna Hazelton, for her vigilance and constant willingness to help. I acknwoledge and thank Dr. Nichole Barry, for clearing some paths before me, and for regular, helpful conversations. I especially acknowledge Dr. Ruby Byrne for her excellence as a colleague and as a friend, and who originally convinced me to join this wonderful research group. I offer my best wishes to the rest of the graduate students in

[1]Rodrick always said to say, "I apologize," and not, "I'm sorry." Sometimes I remember.

our group: Pyxie Star, Dara Storer, and Nicel Mohamed-Hinds, with a special thanks to Pyxie for work directly related to the success of this thesis. I also want to acknowledge the students that I directly mentored, for their patience and tireless work ethic: Jordan Brown, Alex Foster, Imani Ware, and Yu Hou.

There are a number of personal friends who have been instrumental to my development. I begin by acknowledging Dr. Matthew Poulos, for suggesting I might switch my major over to Physics. I would also like to acknowledge the D-Squad (from D-Quad), AKA the 2-Cube Crew, AKA The Town Mayors. You know who you are. I would like to especially thank Travis Fife, Matthew Krawczyk, Steven Kezian, and Elizabeth Aaron for keeping in contact over the long years that have separated us. I thank Dr. Jesse Stryker, Dr. Ian Guinn, and Dr. Jonathan Craig,[2] for serving as mentors, colleagues, and friends.

Last of the personal acknowledgements, but certainly not least, goes to Claire Anne West, who is my romantic partner, who is my best friend, and who I love more than I can say. You are intelligent, beautiful, kind, diligent, and amazing. I cannot wait to experience the rest of our journey together. I also love your family, and want to acknowledge specific members here: Dr. William West, Deborah West, and Matthew West.

Finally, this scientific work makes use of the Murchison Radio-astronomy Observatory, operated by CSIRO. I acknowledge the Wajarri Yamatji people as the traditional owners of the Observatory site. Support for the operation of the MWA is provided by the Australian Government (NCRIS), under a contract to Curtin University administered by Astronomy Australia Limited. I acknowledge the Pawsey Supercomputing Centre which is supported by the Western Australian and Australian Governments.

[2]Rip and Tear.

DEDICATION

To my father, David Edward Wilensky, who could be both tough and tender, though usually not both at the same time.

Chapter 1
INTRODUCTION

This thesis presents one small part of the scientific effort to learn about a cosmological period known as the Epoch of Reionization (EoR). This is the period when the hydrogen content of the universe transitioned from being predominantly neutral to being predominantly ionized. Precisely when this occurred is still a topic of active research, though some constraints do exist. See [27], [46], and [59] for comprehensive reviews of the subject.

1.1 The Expanding Universe, Reionization, and the Hyperfine Transition

Cosmology is probed by observing astrophysical radiation such as the light of stars and galaxies, or in our case, distant radio emission. A basic, yet striking prediction of Einstein's theory of General Relativity is that space is expanding as time progresses [21]. This was first confirmed experimentally by Edwin Hubble, whose original estimation of the eponymous constant describing the rate of this expansion has since been refined, though not without some contention [74, 80]. The expansion of space over time gives rise to a subtle apparent motion between distant sources, oft compared to the relative motion of raisins in a rising raisin cake. This subtle apparent motion essentially yields a Doppler shift that causes blue light to redden, or, if that blue light has been traveling for a significant amount of time, to lengthen beyond optical wavelengths. This phenomenon is thusly referred to as "cosmological redshift." The speed of this apparent motion is larger for sources further away according to Hubble's law, making radiation from ancient sources severely redshifted. It is typical to use the redshift of some observed emission as a proxy for its age, since this unit is more directly relatable to the picture of spacetime presented by General Relativity on scales as large as are considered in cosmology. Observed emission is said to have a redshift, z, if by comparing

the observed wavelength, λ_o, to its wavelength when it was emitted, λ_e one finds that

$$\lambda_o = (z+1)\lambda_e, \tag{1.1}$$

with $z > 0$. Figure 1.1 depicts the expanding universe along with several historical landmarks that we describe below.

The absence of Gunn-Peterson troughs for quasars at redshift ~ 6 or less [26] suggests that reionization was complete by this time, when the universe was 7 times smaller[1] than it is now. This means any signals from the EoR will be redshifted by at least this amount, which directly determines the frequencies at which such signals can be measured. It turns out this has a rather profound impact on the types of challenges faced by EoR search efforts.

To introduce how we will probe the EoR, we discuss why it is called the Epoch of *Re*ionization. Winding the clock back on an ever-expanding universe, we find that, early on, its residents must have been in an incredibly hot and dense state. Persistent bound states between particles, such as atoms or atomic nuclei, require that the constituent particles are close enough to bind, but also not so energetic that they spontaneously unbind. In the very early universe, it is too hot for even atomic nuclei to form. However, the universe cools as it expands, eventually allowing for their formation. At this point, the matter content of the universe is essentially an opaque plasma consisting mainly of free protons and electrons. Before long, it is cool enough that bound states between protons and electrons can form, i.e. neutral Hydrogen atoms. This phase is known as Recombination. The neutralization of matter allows for photons, once trapped by Compton scattering within the plasma, to escape. We call these ancient photons the Cosmic Microwave Background, and our observations of them have allowed for tremendous advances in cosmology over the last few decades.

Eventually, subtle overdensities in the matter distribution started to form clumps, leading to the first stars. Stars emit a continuum of radiation frequencies depending on their temper-

[1]Here we mean to refer to the relative distances between objects, and not the total size of the universe, which may be infinite.

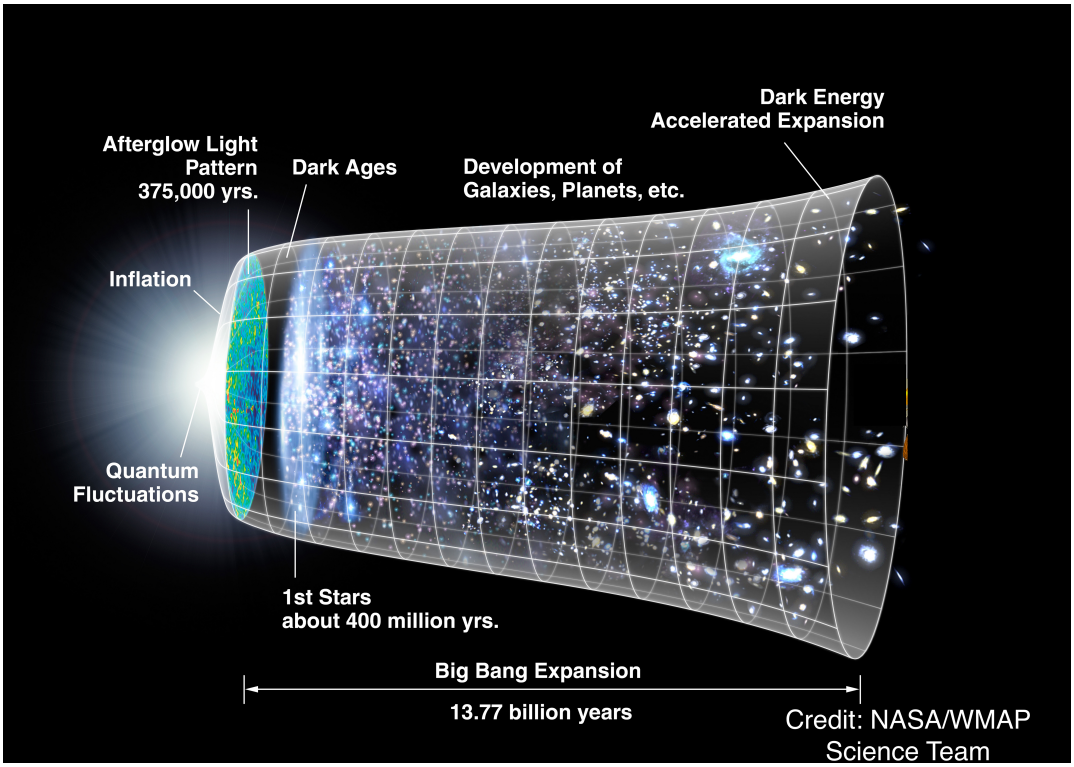


Figure 1.1: An illustration (not drawn to linear scale) of the expanding universe with several cosmological eras indicated as time progresses from left to right. The Epoch of Reionization begins approximately when the first stars form, and continues for several hundred million years until almost all the Hydrogen is ionized and too rarefied to recombine with electrons.

ature, and some frequencies in this continuum are appropriate for ionizing neutral Hydrogen via photon absorption. Today, we observe that most Hydrogen is once-more ionized. We infer that there must have been a cosmological-scale phase transition in the intervening time, and we call that transition period the Epoch of Reionization.

Ultimately, we would like to understand very specific details about the process of reionization, such as the relative contributions of physical processes, distinct morphological features, a detailed timeline, etc. In order to probe some of these details, we will essentially observe the evolution of neutral Hydrogen during the EoR. An excellent, unique marker for large quantities of neutral Hydrogen is the 21-cm line that arises from the hyperfine transition in neutral Hydrogen. This transition is often erroneously described as arising from a relative flip in the spins of the proton and electron in neutral Hydrogen. Since the two anti-aligned states are not themselves eigenstates of the Hamiltonian, a transition between an aligned state and an anti-aligned state cannot be accounted for by an absorption or emission of a photon. The relevant transition is actually between the exchange-symmetrical and exchange-antisymmetrical entangled spin states.[2] This is a so-called "forbidden transition," meaning that it is extremely rare[3] and so a large concentration of 21-cm radiation is usually indicative of a proportionately tremendous amount of neutral Hydrogen.

A promising way to study the EoR is by measuring the spatial power spectrum of the 21-cm brightness temperature. This quantity represents the brightness of the Fourier modes of the 21-cm signal on some patch of the sky, and thus provides direct information about which spatial scales are relevant during reionization. We discuss this quantity in technical detail in chapter 2. Since the EoR ended by $z \sim 6$, we must actually measure significantly longer wavelengths to probe the EoR. Typical frequencies to study the EoR are between 70 and 230 MHz. Thus, low frequency radio telescopes are the required instrument for using the 21-cm line to probe the EoR.

---

[2]See Feynman Lectures on Physics, Volume III, chapter 12, available freely at `https://www.feynmanlectures.caltech.edu`

[3]The mean lifetime is $\sim 3 \cdot 10^7$ years [27].

1.2 Radio Telescopes

Radio telescopes minimally consist of highly sensitive radio antennas, or receiving elements, and various hardware components to acquire electric field measurements from them. In practice, radio telescopes are highly sophisticated precision instruments that require interdisciplinary collaboration between multiple fields of science and engineering to construct, maintain, and study them. By cross-correlating electric field measurements from multiple receiving elements in radio arrays (interferometers), one can use synthesis techniques to reconstruct images of the radio emission from the sky. As will be discussed later in mathematical detail, radio interferometers naturally measure Fourier amplitudes of sky brightness as a function of frequency. This makes them well-suited to a power spectrum measurement, which inherently requires information about the Fourier amplitudes of the sky.

The vast majority of this work is performed using a radio interferometer known as the Murchison Widefield Array (MWA; [92]). We have also applied some of this work and ensuing lessons on the Hydrogen Epoch of Reionization Array (HERA; [24]). Both of these telescopes are in extremely remote locations in order to limit reception of anthropogenic radio signals. The MWA was constructed and commissioned about one decade ago, and has several seasons of high-quality EoR data archived offsite. One of the prevailing lessons that we have learned from studying this data is that our ability to detect the EoR depends critically on our ability to understand and model the behavior of our instrument in exquisite detail. HERA, which is currently being commissioned, is designed with such lessons in mind.

The Phase I configuration of the MWA consists of 128 pseudorandomly distributed receiving elements, where each receiving element is a 4x4 tile of crossed dipole antennas [13, 92] (Figure 1.2). The antennas within a tile are all connected to one analog beamformer so that their individual signals can be combined into a single signal per polarization for passage to the correlator, where they will be channelized, cross-multiplied with signals from other tiles, and then time-averaged to form measurements called visibilities. The channelization process employed by the MWA Phase I involves a 2-stage polyphase filter bank (PFB) and multi-

Figure 1.2: A 4x4 tile of crossed dipoles antennas that form a single receiving element of the Murchison Widefield Array, a radio interferometer located at the Murchison Radio Observatory in Western Australia. Radio interferometers cross-correlate electric field measurements from multiple receiving elements to form visibilities. These visibilities can then be used to construct images of the radio emission on the sky. Photo credit: Dr. Natasha Hurley-Walker.

ple stages of bit selection. The 2-stage PFB causes a non-smooth instrumental frequency response [79] that, when coupled with subtle nonlinearities introduced by the multiple quantization stages, makes for difficult calibration problems [11]. As will be made clear in the main body of this work, unacknowledged frequency structure in the data tends to occlude EoR measurement efforts. Sensitivity calculations suggest that the MWA will need upwards of 1000 hours of integration to make a detection of typical EoR signals [12, 24].

HERA will eventually consist of 350 14-meter parabolic dishes whose feeds will be suspended at the dish foci by ropes attached to utility poles [24]. Each dish will constitute an individual receiving element of the array. The channelization procedure also involves a PFB, however it is only a single-stage PFB, and so the bandpass is significantly smoother than that of the MWA. HERA will also possess substantially more collecting area than the MWA, which gives it more sensitivity. Calculations performed using 21cmSense[4] [76, 77] suggest that, all other analysis choices (integration time, analysis bandwidth, etc.) being equal, HERA will make a substantially more significant detection than the MWA for a given fiducial model [24].

The reality of such sensitivity calculations is that they rarely incorporate contamination from systematic effects. Equivalently, they assume that systematic effects have been solved or corrected for in the final power spectrum measurement. In practice, 21-cm EoR experiments are fraught with systematic contamination [46], largely due to the extreme dynamic range required for such a measurement.

1.3 Systematic Effects

Most radiation that we observe in these instruments is not from neutral Hydrogen during the EoR. Astrophysical sources such as bright radio galaxies and diffuse galactic synchotron emission are usually referred to as foregrounds, since they lie between us and the EoR. In terms of brightness, these foregrounds are 4-5 orders of magnitude brighter than the

[4] `https://github.com/jpober/21cmSense`

expected EoR signal, even in dimmer parts of the sky. These sources are spectrally smooth. In theory, due to this spectral smoothness, the power spectrum of these sources should be mostly contained in lower-order line-of-sight modes. However, due to the chromaticity of the instrument point-spread-function (PSF), power from foregrounds reaches into higher-order line-of-sight modes. In a cylindrical power spectrum, this makes a wedge-like shape; the depth to which the mode-mixing reaches into line-of-sight modes is proportional to the magnitude of the perpendicular mode in question. This "foreground wedge" is an extremely well-studied phenomenon inherent to any interferometric measurement of the power spectrum [23, 33, 44, 45, 55, 56, 71, 91, 95].

Since these foregrounds are so bright, a detection in the wedge is impossible without some amount of foreground subtraction. Foreground subtraction is perilous. Modeling errors can result in unintentional subtraction of the EoR signal along with the foregrounds, known as signal loss. At best, signal loss results in subsequent corrections in the literature that explain the mistake so that others can avoid it [4, 22]. At worst, it generates artificially low upper limits on the EoR signal power, leading to false scientific inferences that need to be retracted [3, 75]. Fortunately, there exists a region of the power spectrum outside the wedge, uncontaminated by foregrounds, which is called the EoR window. Assuming other systematic effects can be handled, a detection of the EoR signal can be made in the window without foreground subtraction.

Systematic effects that introduce chromaticity into the data and failure to remove chromatic effects can contaminate the EoR window. Many instances of this are found in calibration since one of the primary goals of calibration is estimation of the instrument's frequency response. Examples include incomplete sky models resulting in chromatic calibration errors [9, 19, 25, 38, 72], failure to remove cable reflections [10, 14, 40, 41, 43, 52, 64], and other general bandpass calibration errors [38, 94]. This list is by no means exhaustive; further description of such effects and more references can be found in [46, 55].

More direct sources of chromaticity in the data are radiating sources that are not spectrally smooth, unlike the foregrounds. Such sources are typically anthropogenic, such as

digital television (DTV) or frequency modulated (FM) radio broadcasts. Non-astrophysical radio frequency sources are termed radio frequency interference (RFI), and their role in 21-cm EoR science is the main topic of this work.

1.4 Radio Frequency Interference

Unchecked, RFI poses a significant threat to 21-cm EoR power spectrum measurements [101]. Unlike the foregrounds, which are spectrally smooth, RFI has sharp spectral structure. See Figure 1.3. This means that even extremely faint RFI sources provide substantial contamination in the EoR window.

RFI is a problem for any radio astronomy experiment, and so numerous mitigation schemes have been created. Reviews of RFI mitigation schemes are presented in [5], and [7]. For radio interferometers, these can generally be categorized as pre-correlation and post-correlation methods. Since radiated intensity in the far-field falls at least as fast[5] as r^{-2}, where r is the distance between the emitter and receiver, the first line of defense against RFI is to place the telescope in a remote location where anthropogenic signals will appear more faintly than they would otherwise. Both the MWA and HERA are hundreds of kilometers from large population centers. However, owing to the sensitivity of EoR experiments and the proclivity of RFI to propagate long distances, remoteness alone will not sufficiently protect these experiments. Additional examples of pre-correlation techniques that are implemented in radio telescopes are using an external monitoring system to shut off the correlator during high RFI events, blanking correlator inputs that are statistically suspicious, and signal cancellation techniques.

Given the vast data rates of modern radio telescopes, it is common to perform RFI detection offline in the post-correlation stage of the pipeline. Typically such approaches produce flags: a bit mask for the visibilities describing whether certain samples were determined to be corrupted by interference or other noxious effects. Flagged samples are then excluded

[5]Propagation effects can make this sharper, e.g. reflections. See [63] and citations therein.

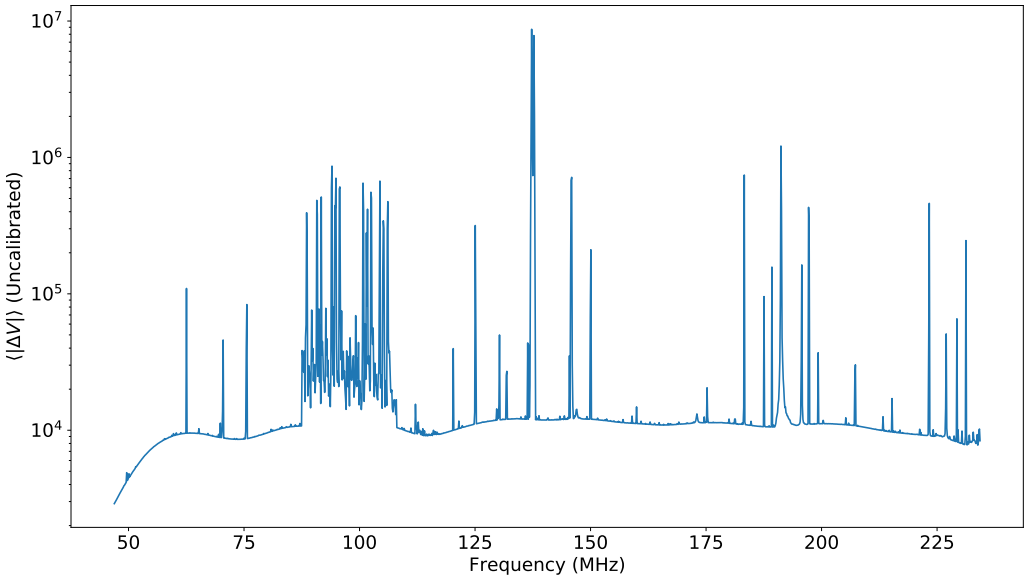


Figure 1.3: A time-averaged incoherent noise spectrum (chapter 3) of one night of HERA data. The sharp variations as a function of frequency are largely due to RFI, while the smooth floor is a combination of the foreground brightness temperature and instrument response. Much of the RFI displayed here is extremely bright and easily detectable both by eye and by RFI excision algorithms. However, there are faint transmissions throughout the entire displayed band that are impossible to discern by eye and require careful algorithms to detect them. HERA also occasionally observes much broader transmissions than are observed in this plot, such as digital television signals. Some known visible transmitters: FM radio (88 to 108 MHz), ORBCOMM (137 MHz), analog television carriers (above 174 MHz).

from the ensuing analysis. Post-correlation mitigation strategies include neural network approaches [18, 42, 96, 103], filters in the uv-plane [82], variants of fringe filters [6, 62], as well as methods that take advantage of the cyclostationarity of common RFI signals [17, 34]. Several post-correlation detection methods are compared in [61], including surface fitting, singular value decomposition, and combinatorial thresholding algorithms such as SumThreshold. Successful implementations of the SumThreshold method are presented in [68] and [73] using the aoflagger[6] and SERPent[7] software packages, respectively. We present an RFI detection algorithm in this work similar to SumThreshold. It is part of an RFI excision pipeline package named SSINS[8] [102]. The major improvements that SSINS makes over previous algorithms is that it performs excision in a much more sensitive space, and specifically targets frequency allocations that can be programmed by the user.

While there is a strong consensus that RFI flagging is a highly important step in preprocessing EoR data [10, 43, 52], and much effort has been devoted to removing RFI contaminated data from EoR studies, there has been relatively little study to specifically characterize the manner in which undetected RFI systematically affects EoR detection efforts. Interestingly, RFI excision itself can induce chromatic structure in radio data that can systematically bias power spectrum estimates, which is specifically studied in [65] using the LOw Frequency ARray (LOFAR; [97]): a telescope located in the Netherlands. We also demonstrate the capacity for flagging to induce spectral structure on otherwise achromatic data in chapter 5 using the MWA. This effect is different than the one studied in [65] in that it appears at coarser data resolution, which invalidates much of the mitigation strategies put forth in [65]. We provide alternate solutions. This has important consequences for overall measurement sensitivity, and can also substantially increase foreground subtraction demands if our solutions are not implemented.

In general, any given mitigation method is bound to leave undetected RFI at some level.

---

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

[7] `https://github.com/daniellefenech/SERPent`

[8] `https://github.com/mwilensky768/SSINS`

For instance, [63] presents a statistical estimate for the apparent brightness of undetected RFI in LOFAR data after AOFlagger, along with an empirical limit on the contamination levels judged from imaging RFI-flagged data. Another example is [10], in which two rounds of RFI flagging were performed on MWA EoR data: first with AOFlagger, and then with SSINS. In that study, SSINS found a substantial number of faint DTV interference events remaining after the first round of flagging. Furthermore, removal of any 2-minute snapshots containing any DTV remnants helped improve EoR limits, where the improvement was quite substantial in a certain subset of the data. A natural line of inquiry that follows from this is to investigate the level to which undetected RFI occludes EoR detection. In chapter 4, we provide a quantitative, theoretical expectation for the systematic bias in the power spectrum that results from undetected RFI, and thereby quantify a tolerance for RFI contamination in EoR data based on the expected brightness of the EoR signal.

We find that these calculations place strict limits on the amount of RFI allowed in a given analysis. This demands a quantitative appraisal of current RFI mitigation techniques. We present such an appraisal of SSINS in chapter 7. We find that a number of undetected RFI sources exist in the data, which leave noticeable contamination in the EoR window. Contaminated power spectra usually contain observations within which SSINS already detected and excised RFI. We therefore use SSINS to identify and dispose of contaminated observations, despite that it cannot identify which samples in those observations still contain RFI. We use this strategy in order to produce an upper limit on the EoR power spectrum signal using a season of MWA data.

Chapter 2

INTRODUCTION TO RADIO INTERFEROMETRY AND SOME FORMALISM REGARDING 21-CM POWER SPECTRA

[1] and [90] are standard references for radio interferometry fundamentals. We discuss some important results and apply them to 21-cm power spectrum measurement. Bold-faced font represents vector quantities whose domain and dimensionality must be inferred from context, while regular font represents scalar quantities.

2.1 Visibilities

The fundamental measurement of a radio interferometer is known as a visibility, which is a time-averaged cross correlation of the electric field measurements from two receiving elements, usually reported in the frequency domain. Visibilities sample the spatial coherence function on the ground, $\Gamma(u, v)$, which is related to the intensity distribution on the sky, $I(l, m)$, by the Van Cittert-Zernike theorem:

$$\Gamma(u, v) = \int\int_{\text{FoV}} dl\, dm\, I(l, m) e^{-2\pi i(ul+vm)}. \tag{2.1}$$

Here, u and v represent distances between points on the ground in a rectilinear coordinate system, measured in wavelengths of the relevant emission, λ. Supposing the vector displacement between two points is $\Delta\mathbf{x}$, then $\mathbf{u} = \Delta\mathbf{x}/\lambda$. l and m represent direction cosines for a vector pointing towards sources in the field of view (FoV) from the ground. This expression is approximate in the sense that it assumes sources on the sky are spatially incoherent, that the sources are extremely far away, that we are not integrating over a wide swath of the sky (i.e. the FoV is small), and that the points for which we are measuring coherence lie in a plane that is parallel to the approximately flat region of sky we are integrating over.

Implicitly, only one frequency is considered, and it also assumes that the waves propagated through a non-refractive medium.

Most of these assumptions are more or less obeyed in many radio astronomy contexts. Non-planar antenna arrays and wide-field instruments give rise to so-called w-terms.[1] Failing to correct for widefield effects causes smearing of sources in regions away from phase center. The MWA is a wide-field instrument, and so imaging MWA data requires understanding of these effects. To first order, the ionosphere is refractive at radio frequencies, and sometimes much worse when it is active [37, 58]. For this reason, we do not use observations with poor ionospheric quality in our EoR analyses.

From equation 2.1, we see that if the spatial coherence function on the ground is known perfectly and continuously, then synthesizing an image of the sky reduces to a Fourier inversion problem. This is impossible in practice since receiving elements take up some nonzero area on the ground and finite resources dictate that only so many samples can be generated. The incomplete sampling of the uv-plane means that Fourier inversion of the visibilities results in a point-spread function (PSF). The spatial distribution of these samples in the uv-plane determines the shape of the PSF. Since the distance between pairs of antennas (baselines) is measured in wavelengths, this PSF is naturally chromatic. This is ultimately what gives rise to the foreground wedge [33, 56, 60]. In addition to these constraints, each receiving element performs an implicit sum of the electric field that it collects. This means that the correlations reported in the visibilities contain information from a region in the uv-plane, rather than sampling it at points. The influence or weight that each visibility receives from the points in the uv-plane that contribute to it is related to the combined receiving element beam pattern. The beam pattern is usually also chromatic. Finally, there is an additional bandpass response that can result from analog and digital filters in the signal pipeline.

Assuming a primary beam response, $A(l, m, \nu)$, which acts as a complex-valued, chro-

[1] As in (u, v, w).

matic window function overtop the frequency-dependent sky brightness distribution, and a bandpass response, $\Psi(\nu)$, the visibility from antennas j and k, V_{jk}, can be written

$$V_{jk}(u_{jk}, v_{jk}, \nu) = \int\int_{\text{FoV}} dl\, dm\, I(l, m, \nu) A(l, m, \nu) \Psi(\nu) e^{-2\pi i (u_{jk} l + v_{jk} m)}. \tag{2.2}$$

This measurement process is illustrated in Figure 2.1. In practice, the raw visibilities need to be calibrated due to variations in electronic amplification and other such effects that result in different antenna gains. Formally, this would amount to indexing quantities like A and Ψ to represent idiosyncrasies between antennas. Instrument calibration in 21-cm cosmology is extremely important and the requirements for EoR measurements are strict [9, 94]. However, it is not the main subject of this work, and we will ignore it in the analytic formalism for the sake of simplicity.

Through careful calibration and superposition of the visibilities in the uv-plane and reconstruction of the sampling function of the array, one may still do Fourier inversion to develop a quantity on the sky closely related to the sky brightness. If the antenna response is well-known, then one can attempt to reconstruct the apparent coherence function from the visibility samples by smearing each visibility according to the beam pattern of each baseline in an effort to undo the implicit average performed by the antennas. This is referred to as "gridding," since the process usually involves taking the sampled visibilities to a finely meshed grid of the uv-plane for digital reconstruction. One can also grid with a kernel other than the instrument response, which can have a variety of effects on the analysis, some of which can be beneficial for a given scientific goal [10, 66]. This introduces an extra factor of the beam (or other gridding kernel) in the estimator that can be taken care of with appropriate reweighting to return an estimate of the apparent coherence function. The gridded, reweighted visibilities can then be Fourier transformed to create a "dirty" map of the apparent sky; "dirty" refers to the fact that this estimate of the apparent sky will still be corrupted by the instrument PSF.

The importance of reducing the problem to one of Fourier transforms cannot be understated. The Fourier transform is an extremely well-studied mathematical tool that is ubiqui-

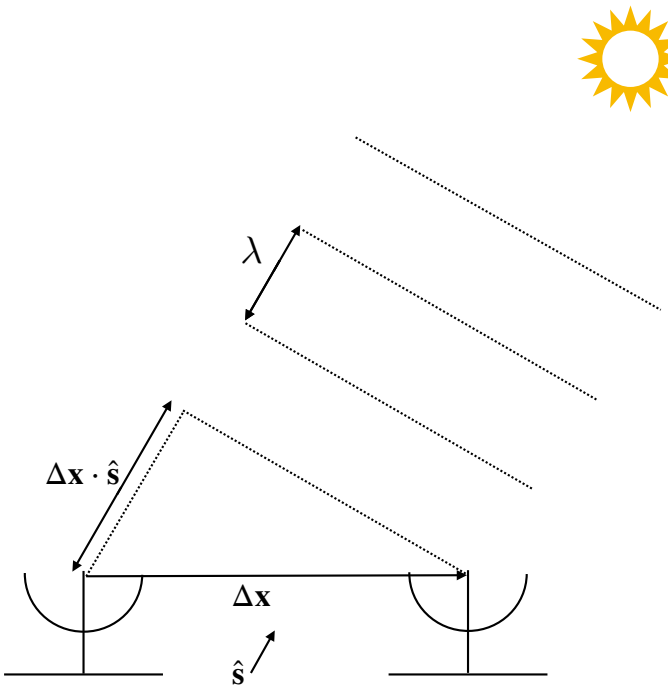


Figure 2.1: Cartoon of two dish antennas pointed at zenith, with an off-zenith source. The source's contribution to the phase of the visibility at frequency ν with corresponding wavelength λ is given by the relative phase of the wavefronts arriving at each of the antennas. For an extremely distant point-source, this is determined exactly by the dot product between the antenna displacement vector and a vector pointing to the source, $\hat{\mathbf{s}}$, whose horizontal components are the direction cosines l and m. The amplitude for a point-source will be a product of the source intensity and the instrument response, i.e. its apparent brightness. An extended source's amplitude and phase, or the total visibility from all sources visible by the antennas, would be determined by the integral in equation 2.2.

tous in physics and other sciences. There are algorithms that allow for rapid computation of the Fourier transform, called Fast Fourier Transforms (FFT), that are not necessarily available for more complicated integral transforms. For an excellent, mathematically rigorous approach to Fourier analysis, see [85]. More application-focused references are widespread. [98] presents a summary of some useful properties of Fourier transforms and introduces the Lomb-Scargle periodogram, which we employ in 21-cm power spectrum estimations with the MWA.

2.2 Cosmological Coordinates, Brightness Temperatures, and Power Spectra

In a more general context, the power spectrum of a signal usually refers to the complex modulus square of that signal's Fourier transform. In 21-cm power spectrum measurements, we are trying to measure the power spectrum of the 21-cm brightness temperature fluctuations. Since the EoR is ancient, a cosmologically accurate coordinate system is required in order to construct the power spectrum. A useful general resource for distance measures in cosmology is [35]. In addition, [57] highlights some useful coordinate conversions specifically for 21-cm cosmology. [44] and [57] both discuss differences in Fourier conventions and the relationship between power spectra constructed in cosmological and interferometric coordinates. We repeat a few results here for ease of reference.

2.2.1 Cosmological Coordinates

First, for a narrow field of view, direction cosines, (l, m), and angles on the sky, $\boldsymbol{\theta} = (\theta_x, \theta_y)$, are nearly equal. To convert an angle to a transverse cosmological position, $\mathbf{r}_\perp = (r_x, r_y)$ we need the transverse comoving distance, $D_M(z)$:

$$\mathbf{r}_\perp = D_M(z)\boldsymbol{\theta}. \tag{2.3}$$

For a flat universe, which we assume in this work, $D_M(z)$ and the line-of-sight comoving distance,[2] $r_\parallel(z)$ are equal. $r_\parallel(z)$ is given exactly by

$$r_\parallel(z) = \frac{c}{H_0}\int_0^z \frac{dz'}{E(z)}, \tag{2.4}$$

and for a flat universe with negligible photon density,

$$E(z) = \sqrt{\Omega_M(1+z)^3 + \Omega_\Lambda}. \tag{2.5}$$

H_0 is the Hubble parameter, while Ω_M and Ω_Λ are the matter density and contribution to the density due to a hypothetical cosmological constant, respectively. For the relatively short cosmological time spans considered in this work, the right-hand-side of equation 2.4 is implicitly a nearly linear function of frequency. Rewriting equation 1.1 in terms of frequency and differentiating equation 2.4 gives the slope of this relationship, which we will denote with β:

$$\beta \equiv -\frac{c\nu_e}{H_0 E(z)\nu_o^2}. \tag{2.6}$$

ν_e is the emitted frequency corresponding to the 21-cm line, while ν_o is the observed frequency. It is typical to use the frequency corresponding to the central redshift of the observation, z_0, in the evaluation of this slope. For our frequencies, the maximum fractional difference between the linear and exact relationship is $\sim 10^{-4}$. This allows for the frequency axis of the interferometric data to be genuinely thought of as a distance, and for a straightforward Fourier conjugate, denoted by η, to be defined. The vector $\mathbf{r} = (r_x, r_y, r_\parallel)$ forms the cosmological coordinate vector.

We denote the Fourier dual to the vector, $\mathbf{r}$, with the vector, $\mathbf{k} = (k_x, k_y, k_\parallel)$. Occasionally we will single out just the transverse components, $\mathbf{k}_\perp = (k_x, k_y)$. We denote the magnitude of $\mathbf{k}$ with the italic k rather than bold. To transform between interferometric Fourier modes, $\mathbf{u}$, and cosmological modes, $\mathbf{k}_\perp$, one must use [57]:

$$\mathbf{k}_\perp = \frac{2\pi}{D_M(z_0)}\mathbf{u}. \tag{2.7}$$

[2]This is usally denoted as $D_C(z)$ in the cosmology literature, but we deviate so that our notation is a bit more self-consistent.

In principle one could use a redshift-dependent expression instead of the central redshift for D_M, which leads to effects essentially similar to redshift-space distortions. We will discuss this very briefly in Appendix D. The relationship between η and $k_\parallel$ is [57]

$$k_\parallel \approx \frac{2\pi H_0 E(z_0)\nu_e}{c(1+z_0)^2}\eta. \tag{2.8}$$

2.2.2 The 21-cm Power Spectrum

Part of the purpose of this subsection is to provide yet another accounting of the relationship between power spectra as described in theoretical discussions and those that are produced by experimental measurements. The interested reader should also read Appendix A and Table I in [44], as well as section 3.1 of [57].

The 21-cm power spectrum is specifically the power spectrum of the brightness temperature fluctuations of 21-cm radiation about the CMB temperature. At our frequencies, the conversion between observed intensity and brightness temperature is well-approximated by the Rayleigh-Jeans law:

$$I = \frac{2k_B T\nu^2}{c^2}. \tag{2.9}$$

k_B is Boltzmann's constant, and T is the brightness temperature. I and T should implicitly be thought of as scalar fields on the sky at this point.

The power spectrum, $P(\mathbf{k})$, is defined by the equation [44]

$$\langle\tilde{T}(\mathbf{k})\tilde{T}^*(\mathbf{k}')\rangle = P(\mathbf{k})\delta^3(\mathbf{k}-\mathbf{k}'), \tag{2.10}$$

where the tildes indicate a quantity that has been Fourier transformed (also indicated by the values of its arguments in this expression), δ^3 indicates a 3-dimensional Dirac delta function, and the brackets indicate an ensemble average. The Dirac delta function represents the notion that different $\mathbf{k}$-modes are uncorrelated with one another.

One can define a similar quantity in interferometric coordinates [44]:

$$\langle\tilde{T}'(\mathbf{u},\eta)\tilde{T}'^*(\mathbf{u}',\eta')\rangle = P(\mathbf{u},\eta)\delta^2(\mathbf{u}-\mathbf{u}')\delta(\eta-\eta'), \tag{2.11}$$

Here, $\tilde{T}'$ has been obtained by measuring the sky intensity, converting to a brightness temperature, and Fourier transforming, all in the interferometric coordinates. By explicitly substituting the Fourier transform expressions and taking very special care of coordinate conversions and Fourier conventions, one can arrive at a conversion between these two quantities, expressed by equation (A10) in [44]. We reproduce it here in our notation:

$$P(\mathbf{k}) = \frac{c(1+z_0)^2 D_M(z_0)}{H_0 \nu_e E(z_0)} P(\mathbf{u}, \eta) \tag{2.12}$$

Equations 2.10 and 2.11 correspond with the "Fourier transform and modulus square" power spectrum definition in the sense that one can imagine integrating both sides of either equation and finding that the delta functions effectively set $\mathbf{k} = \mathbf{k}'$, etc. The integration necessitates a small $\mathbf{k}$-space volume factor on the left-hand-side, which would be infinitesimal in the continuous limit. In an observational setting, the finite survey volume sets a limit on the resolution in $\mathbf{k}$-space, which in turn sets this volume factor. This is exactly one over the survey volume, denoted by $V_{\mathcal{M}}$. We will introduce such an estimator, with some additional instrumental complications from equation 2.2, in chapter 4.

Returning to the sky from the visibilities produces an estimate of the sky that was built from incomplete information. This is an important point to consider when interpreting the results of a 21-cm EoR upper limit analysis. Any analysis choice will affect the ensuing power spectrum estimate, and it is important to determine whether estimation choices preserve the properties of the EoR signal, rather than enhancing or diminishing certain aspects. In the imaging and power spectrum analyses this work, we estimate the apparent sky, and therefore a apparent power spectrum. We then normalize with respect to the effective survey volume provided by the choice of estimator, which in this case is the volume of the beam.

We have introduced much of the theoretical formalism describing interferometry and the relevant cosmological measurements that we will use in this work. Now we turn to the matter of addressing the complication of radio frequency interference in the measurement.

Chapter 3

ABSOLVING THE SSINS OF PRECISION INTERFEROMETRIC RADIO DATA: A NEW TECHNIQUE FOR MITIGATING FAINT RADIO FREQUENCY INTERFERENCE.

This chapter is largely based on work in [102].

3.1 Introduction

As mentioned in Chapter 1, most RFI is anthropogenic. Many RFI signals are narrow compared to the frequency resolution of the MWA or HERA. This causes contaminated visibilities to appear spined, similar to the spectrum in Figure 1.3. Some signals, such as DTV or digital audio broadcasting (DAB), appear relatively broad even for an EoR experiment (Figure 3.1). Regardless of the signal type, RFI tends to imprint sharp frequency structure overtop smooth radio emission from the foregrounds. In chapter 4, we quantify the level of power spectrum contamination that this structure provides. In this chapter, we discuss RFI flagging methods.

RFI contamination of EoR measurement was foreseen in its conception, and many clever excision algorithms have been developed to automatically flag RFI in large volumes of EoR data. The standard flagger for MWA data is called AOFlagger. It is effective at locating signals that are narrow in time and/or frequency, but is less effective with signals that are broad in both. In particular, DTV signals in the MWA EoR Highband above 174 MHz are ill-flagged by AOFlagger [68].

Like many current state-of-the-art methods, aoflagger flags a single baseline at a time. Clearly, single baseline flaggers are fundamentally limited to the sensitivity attainable by a single baseline. As we will show, there exists RFI observed by the MWA that evades detection

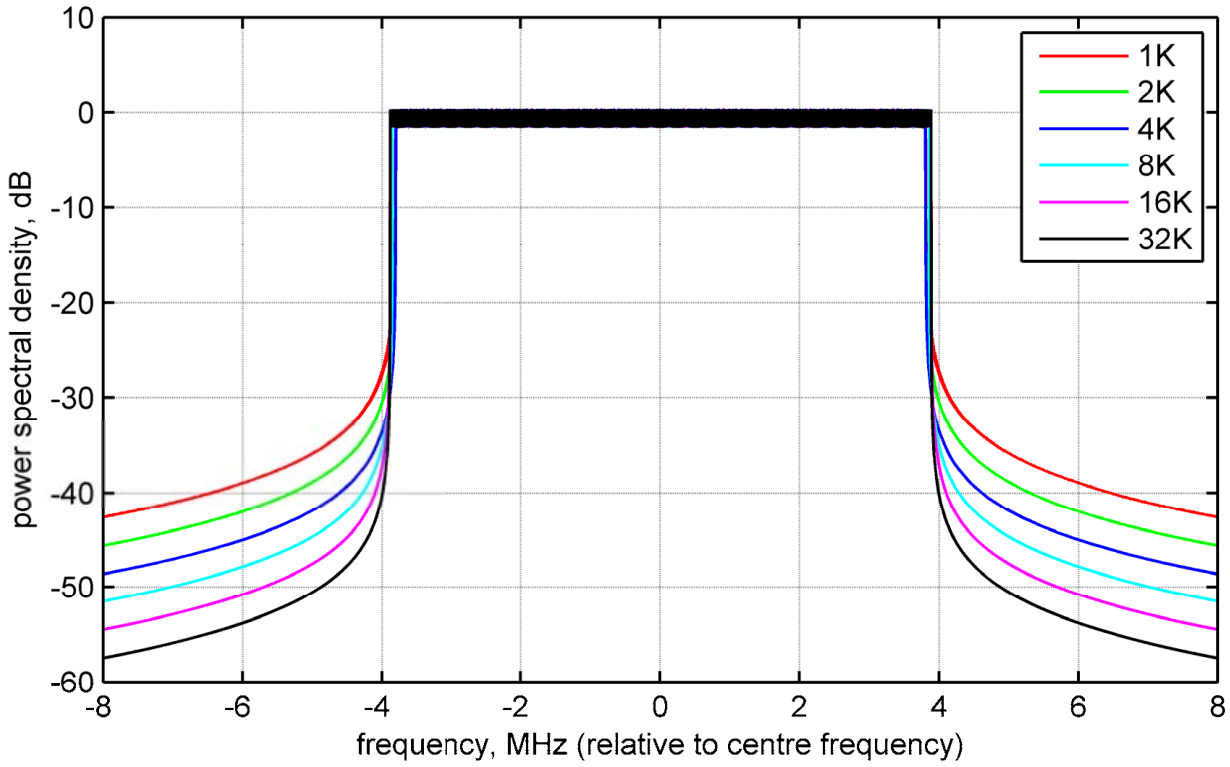


Figure 3.1: Adapted from European Telecommunication Standards Institute, ETSI EN 302 755 v1.4.1. Shows theoretical DTV power spectral density of a T2 signal for an 8 MHz allocation. It is largely flat over the allocation, occupying almost all of the allocation at full power. A few different guard band conventions are shown. While the signal is theoretically smooth over most of the allocation, the sharp cutoff over the guard interval presents difficulties in EoR power spectrum measurements. Different regions can have different allocations. While DTV in South Africa is allocated 8 MHz per channel, in Western Australia it is only allocated 7 MHz. This figure does not show potential boosting of the pilot carriers, which can add some spectral structure.

by AOFLAGGER because it is fainter than the single baseline thermal noise.

We offer Sky-Subtracted Incoherent Noise Spectra (SSINS) as a flexible, sensitive, and statistically simple RFI analysis pipeline to detect, classify, and excise such faint RFI. SSINS was developed for the MWA, and it has also undergone limited testing on HERA and the Long Wavelength Array (LWA; [88]). It is highly effective at excising DTV signals in MWA data that are missed by AOFLAGGER

After explaining the SSINS method in detail (§3.2), we show several common cases of RFI that are successfully flagged by SSINS in the MWA EoR highband, which extends from 167.1 Mhz to 197.7 Mhz (§3.3). We also image RFI and hypothesize about the source (§3.4). We then discuss the implementation of SSINS, explain how it can be customized for different RFI environments, and summarize RFI mitigation efforts for data used in an EoR power spectrum limit in [10] (§3.5). Finally, we discuss possibilities for future work (§3.6).

3.2 The Method

The measurements from an interferometer can be broken down into three components: the astrophysical sky signals, thermal noise, and RFI. We will first separate the slowly varying sky signal from the modulated RFI (§3.2.1), then increase our sensitivity to the faint RFI signal through successive integrations and frequency-matched flaggers (§3.2.2–3.2.4). In Figures 3.2–3.4 we compare an observation free of RFI (left) with an observation contaminated by faint RFI (right).

3.2.1 Sky Subtraction

The first step, sky subtraction, relies on the time variation of the sky signal being slow relative to the visibility cadence. We denote the visibility belonging to the ij antenna pair, time integration, t_n, frequency, ν, and polarization, p, as $V_{ij}(t_n, \nu, p)$. By subtracting data from subsequent time integrations, the majority of the astrophysical sky signal will be removed while leaving the thermal noise and much of the modulated RFI. We write the

sky-subtracted visibilities as

$$\Delta V_{ij}(t_n, \nu, p) = V_{ij}(t_{n+1}, \nu, p) - V_{ij}(t_n, \nu, p). \tag{3.1}$$

What remains has a noise-like component and potentially an RFI-like component.

To separate the thermal noise from the RFI it is helpful to understand the statistical properties of the thermal noise. For visibility data, the noise is a circular complex Gaussian process. In other words, for each visibility, the real and imaginary components of the noise are independently and identically distributed Gaussian random variables with mean equal to zero. Following standard derivations such as in [104], the amplitudes, X, of a circular Gaussian process are Rayleigh distributed:

$$f_X(x;\sigma) = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}, \tag{3.2}$$

where σ^2 is the variance of the Gaussian which describes the real (or imaginary) component of the process. The mean of the Rayleigh distribution is

$$E[X] = \sqrt{\frac{\pi}{2}}\sigma, \tag{3.3}$$

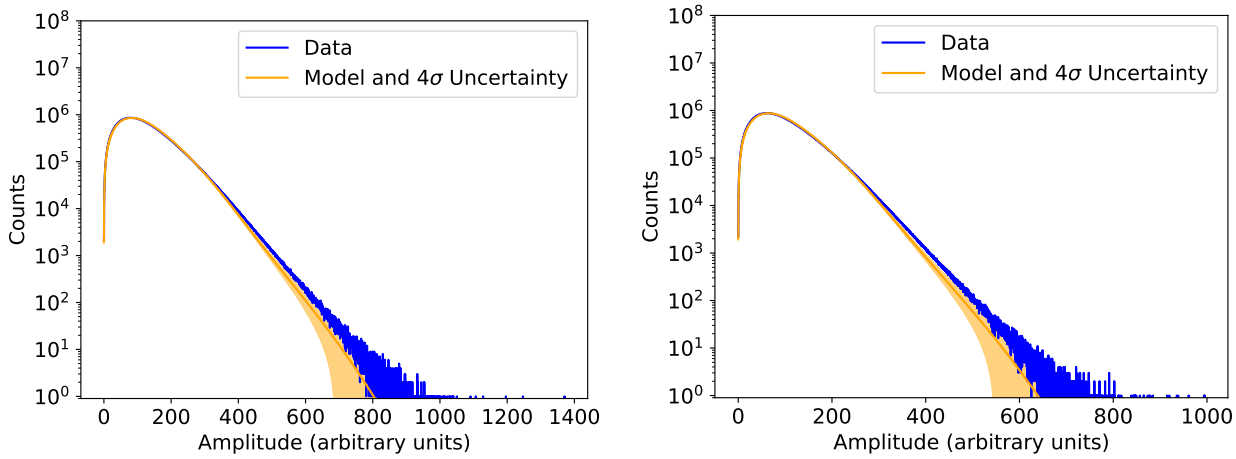
and the variance is

$$E[X^2] - E[X]^2 = \frac{4-\pi}{2}\sigma^2. \tag{3.4}$$

Figure 3.2 shows the Rayleigh-like amplitude distributions of the sky-subtracted visibilities for our two example observations along with a Rayleigh-mixture fit. Because the observed noise has a frequency dependence, each frequency channel was fit independently by maximum likelihood estimation to form the final model fit and 4σ errors. The similarity of Figures 3.2(a) and (b) shows that the sensitivity of a single visibility is not sufficient to mitigate the faint RFI that we are interested in.

3.2.2 The Incoherent Average

We boost our RFI sensitivity by averaging the amplitudes of the sky-subtracted visibilities over all of the baselines in the array, leaving a single dynamic spectrum per polarization.



(a) (b)

Figure 3.2: Histograms of the amplitudes of time-differenced visibilities for two different two-minute MWA observations, similar to Figure 8.2 of [11]. (a) belongs to an observation deemed clean by the methods shown in this paper, while (b) belongs to an observation that is shown to have some digital television contamination (§3.2.4). The measurements are shown in blue, while an accompanying Rayleigh-mixture fit is shown in orange with 4σ error bars. These figures are nearly indistinguishable, despite differences in contamination.

This is not an entirely novel idea. We refer the reader to the last two paragraphs of §4.2 in [68]. Formally, for an array with N_A antennas, the incoherent average, Y, is

$$Y(t_n, \nu, p) = \frac{2}{N_A(N_A - 1)} \sum_{i=1}^{N_A} \sum_{j>i}^{N_A} |\Delta V_{ij}(t_n, \nu, p)|. \tag{3.5}$$

Since we discard the phase of the sky-subtracted visibilities in the average, we call this an incoherent average. Note that we do not include autocorrelations in the average. We refer to the remaining spectra as sky-subtracted incoherent noise spectra (SSINS), incoherent noise spectra, or just noise spectra.

Examples of incoherent noise spectra for the same observations as in Figure 3.2 are shown in Figure 3.3. The spectra in Figure 3.3(b) actually show some RFI from digital television broadcasting, although it is extremely difficult to discern from the surrounding noise even in this very sensitive space. In order to determine the nature of such features, we boost the contrast in a way that allows for the rigorous application of statistics.

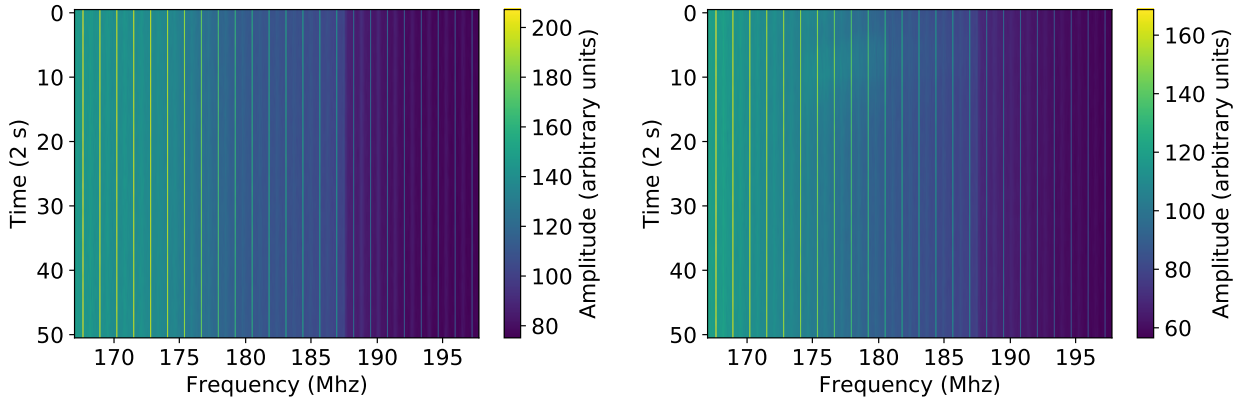
3.2.3 Mean Subtraction

Mean subtraction transforms the data so that the data of a clean observation will be standardized: it will appear as if it were sampled from a zero-mean, unit width Gaussian probability distribution. We describe this process formally using the central limit theorem.

In our context, we take the central limit theorem to say the following (see [16]). Let $(X_1, X_2, ..., X_N)$ be independent and identically distributed random variables with finite mean, μ, and variance, Σ^2. Then, as N grows large, the sample means given by

$$S_N = \frac{1}{N} \sum_{k=1}^{N} X_k \tag{3.6}$$

converge in distribution to a normally distributed random variable of mean, μ, and variance, Σ^2/N. Let us assume the thermal noise to be independent between baselines and ignore



(a) (b)

Figure 3.3: The incoherent noise spectrum for the E-W polarization corresponding to the histograms in Figure 3.2, on a two second cadence. The MWA employs a two-stage polyphase filter bank, involving a coarse channelization and then a fine channelization of each coarse channel. The periodic banding in frequency that is seen in these spectra is a result of that filter. The spectrum on the left is clean. However, if one examines the spectrum on the right extremely closely, they may notice a smudge between 174 and 181 Mhz in the first twenty seconds of the observation. This smudge is made much more obvious after the mean-subtraction transformation, shown in Figure 3.4, and it is DTV interference.

baseline-to-baseline noise variation[1] Having averaged over so many baselines[2] in Equation 3.5, the central limit theorem states that the thermal background in the incoherent noise spectrum will be very nearly normally distributed at each frequency with a mean described by Equation 3.3 and variance described by Equation 3.4, divided by the number of baselines in the array, which we denote N_{bl}:

$$N_{bl} = \frac{N_A(N_A - 1)}{2}. \tag{3.7}$$

Note that the Rayleigh distribution has only a single parameter, so we can actually relate these two quantities. Using Equations 3.3 and 3.4, and writing the underlying mean for each frequency as μ_ν, we write the underlying thermal background distribution for each frequency as

$$f_Y(y; \mu_\nu, N_{bl}) = \sqrt{\frac{N_{bl}}{2\pi C \mu_\nu^2}} \exp\left[-\frac{N_{bl}}{2C\mu_\nu^2}(y - \mu_\nu)^2\right], \tag{3.8}$$

where

$$C = \frac{4}{\pi} - 1 \tag{3.9}$$

is the ratio of the Rayleigh variance to the square of its mean.

At this stage, the background distribution for the entire spectrum is a mixture distribution of all the frequencies. If for each frequency we subtract the mean and normalize with respect to the standard deviation, then the background for the entire spectrum will be described by a single distribution (the standard normal distribution), thus simplifying the problem.

If there are sufficiently many time samples in an observation, then an estimation of the means per frequency can be obtained simply by taking the mean of the spectrum in time:

$$\hat{\mu}(\nu, p) = \frac{1}{N_t - 1} \sum_{n=1}^{N_t - 1} Y(t_n, \nu, p). \tag{3.10}$$

Writing the mean only as a function of frequency and polarization of course assumes that the thermal process is at least wide-sense (weakly) stationary in time. In the event that there is

[1]Statements of the central limit theorem exist for non-identically distributed sequences, which would be important here if baseline-to-baseline variation in noise levels were a dominant effect. See [16].

[2]Over 8000 for the MWA

some drift, then a trend line or trend polynomial can be calculated for each frequency and polarization. For now, we will work with the simplest, most prevalent case.

Now we consider the quantity

$$Z(t_n, \nu, p) = \frac{Y - \hat{\mu}}{\sqrt{C\hat{\mu}^2/N_{bl}}} \tag{3.11}$$

We call this the mean-subtracted incoherent noise spectrum. Here, for each data point, we have subtracted the estimated mean and normalized with respect to the estimated standard deviation in the corresponding frequency channels and polarizations. Should the observation be totally clean of RFI and if sky-subtraction indeed fully removed the sky from the data, then the remaining data, which will be purely thermal, ought to be very nearly distributed according to

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \tag{3.12}$$

otherwise known as the standard normal distribution. In other words, the quantity in Equation 3.11 is a z-score for each sample in the spectrum.

An example of mean-subtraction results following from Figures 3.2 and 3.3 is shown in Figure 3.4. The accompanying histograms are plotted along with a standard normal distribution. The clean observation in Figure 3.4(c) conforms to the standard normal distribution exceedingly well, while the contaminated observation in Figure 3.4(d) has outliers well beyond what is expected, indicating highly non-thermal behavior. Recall that in Figure 3.2, it was nearly impossible to discern any difference between these two observations.

The mean-subtracted spectra are very useful for highlighting RFI that is only marginally brighter than the surrounding thermal noise after incoherently averaging over all baselines. In order to programmatically identify and flag RFI, we deploy a frequency-matched flagger in the mean-subtracted spectrum.

3.2.4 The Frequency-Matched Flagger

We extend the probability theory from the previous section to develop a frequency-matched flagger. We introduce this development by showing the results of thresholding on a single-

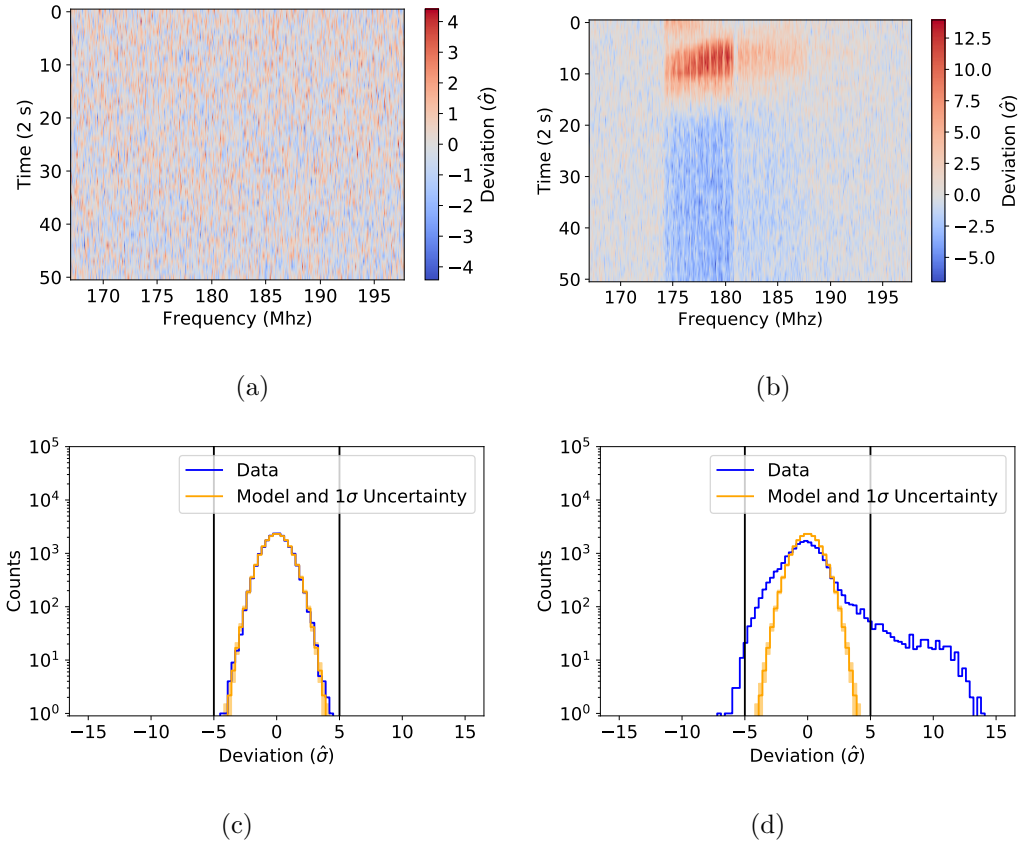


Figure 3.4: Top Row: the mean-subtracted incoherent noise spectra for a single polarization, corresponding to the data in Figure 3.3. Bottom Row: Histograms for the mean-subtracted incoherent noise spectra in the top plots, shown with vertical lines that demarcate a point beyond which very few outliers are expected at this data volume. The spectrum in (a) is clean. There are no discernible features such as clustering in time or frequency and the extent of the data is within the range of outliers expected for a data volume of this size. The spectrum in (b), however, features a noticeable cluster of positive outliers that extends from 174-181 Mhz in the beginning of the observation, and a slightly less noticeable one from 181-188 Mhz. Not only is this clustering antithetical to stationary noise, but the data in the brightest feature has outliers as strong as $14\hat{\sigma}$, which is not expected at this data volume. In correspondence with this reasoning, the purportedly clean observation in (c) looks highly Gaussian, while the contaminated observation in (d) is clearly deviating from the thermal model.

sample basis, and then we show how information from multiple samples can be incorporated.

If the total data volume is M, then the expected number of outliers in a clean mean subtracted spectrum, N_{out}, beyond a certain threshold, τ, is

$$N_{out} = M\text{erfc}\left(\frac{\tau}{\sqrt{2}}\right), \tag{3.13}$$

where erfc is the complementary error function:

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty dt\, e^{-t^2}. \tag{3.14}$$

This includes both positive and negative outliers. For a typical two-minute MWA EoR highband observation, the number of expected outliers when $\tau = 5$ is about 0.05. In other words, about one in twenty observations ought to have just a single sample of that strength or greater. Statistically speaking, samples of that strength are exceedingly unlikely to be thermal and so we can be confident that those samples are contaminated.

It is tempting to flag all data beyond the significance threshold initially, but this can be problematic when there is RFI contamination. Recall that the estimated mean and standard deviation are determined by a time-average in each frequency channel, as described in Equation 3.11. If RFI that is brighter than the thermal noise contaminates some times, then the mean estimate for the contaminated channels will be skewed upward relative to an estimate drawn from only the clean data. As a result, some clean data may appear to be outlying in the negative direction beyond the significance threshold. For example, consider the bright red cluster followed by the blue trough in Figure 3.4(b). As shown in Figure 3.4(d), some data in this observation lies beyond the threshold in the negative direction when the z-scores are initially calculated. Flagging everything beyond the threshold initially will inevitably flag some clean data in this case. We can decrease this type of overflagging by taking an iterative approach to flagging, wherein the RFI contamination of the mean estimate is progressively removed in each iteration. This also allows us to probe deeper into the observation for fainter RFI in each iteration.

We describe the basic iterative flagging procedure below and present a flow chart in Figure 3.5. First, we calculate the z-scores of each noise spectrum sample using mean-subtraction.

Next, we identify the time and frequency of the strongest outlier in the mean-subtracted spectrum that is beyond the threshold in either the positive or negative direction. Then, we flag this time and frequency in all polarizations regardless of the polarization in which the outlier was found. This is done in case that RFI is present in other polarizations at extremely faint levels; it is often the case that RFI is not exactly polarized along the directions measured by the antennas. We then repeat this process until no more outliers exist beyond the threshold. The results of this process using $\tau = 5$ on the contaminated observation shown in previous figures is shown in Figure 3.6(a). In this figure, we deviate from the previous convention in Figures 3.2-3.4. Now, the same observation is in both panels. We are comparing differences in flagging between the single-sample outlier iteration detailed above and a frequency-matched flagger, detailed in the main text below. We can see that there appears to be incomplete flagging in the feature noticed in Figure 3.4(b), and that another feature immediately adjacent to it of similar width persists in the observation at fainter levels. Any individual remaining sample in these features lies beneath the significance threshold, and so cannot be caught by the method above at the threshold set. To boost our sensitivity to faint features such as these, we combine samples of the mean-subtracted spectrum across frequencies.

So far, we have only relied on knowledge of the thermal background without including any specific information about the particular source of the RFI. In the examples shown, data are taken in Western Australia over the frequencies 167-197 Mhz. A subset of the Western Australian digital television channels are broadcast in this observing band and are broadcast 7 Mhz wide, adjacent to one another[3]. The broad features shown in the previous figures correspond to DTV channels 6 and 7. The MWA also sometimes observes DTV channel 8. We can hunt for these particular contaminants by summing the mean-subtracted spectrum over the frequencies belonging to a particular type of contaminant. A similar targeted sub-band summing method was employed in §4 of [68] exactly for DTV interference.

---

[3] `https://www.acma.gov.au/~/media/Licence-Issue-and-Allocation/Publication/pdf/TVRadio_Handbook_Electronic_edition-pdf.pdf?la=en`

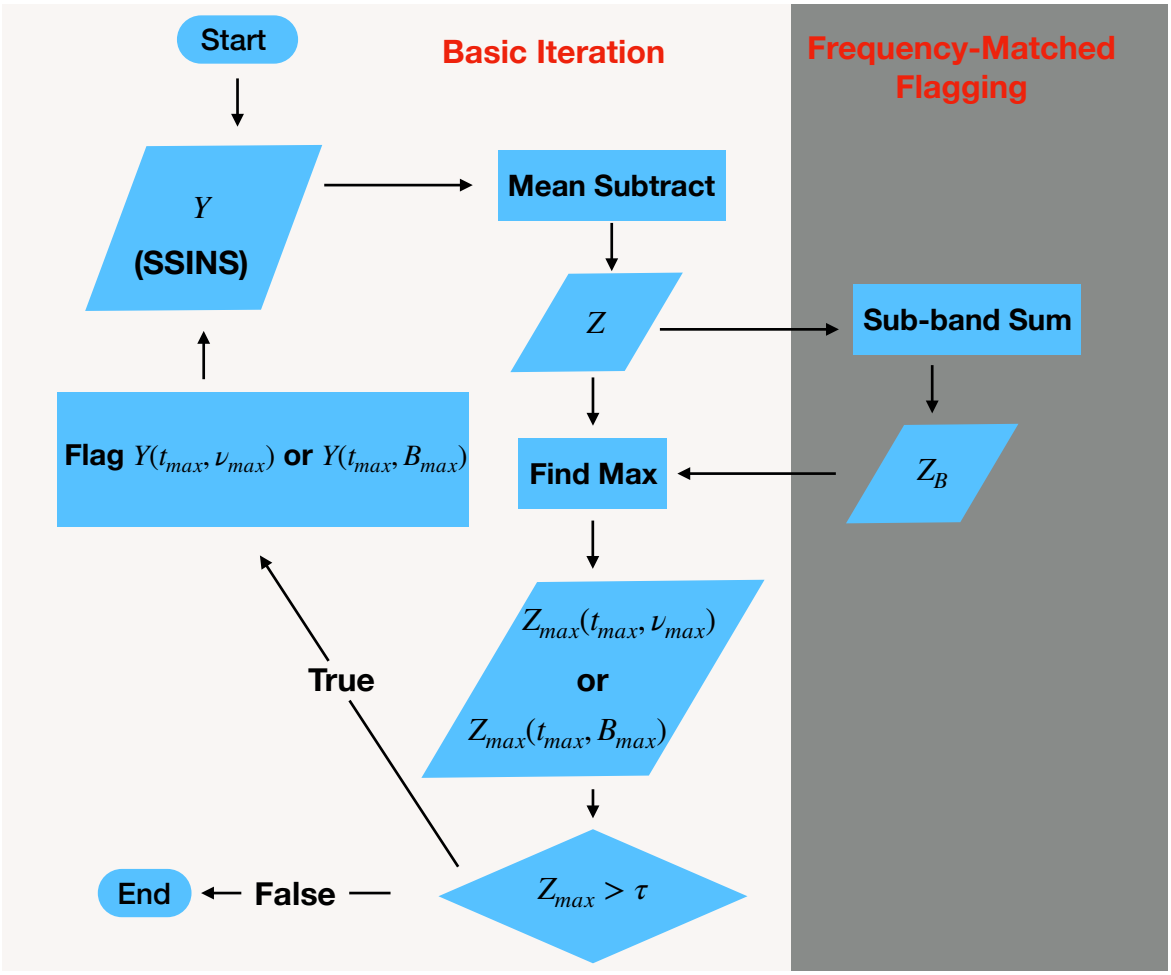
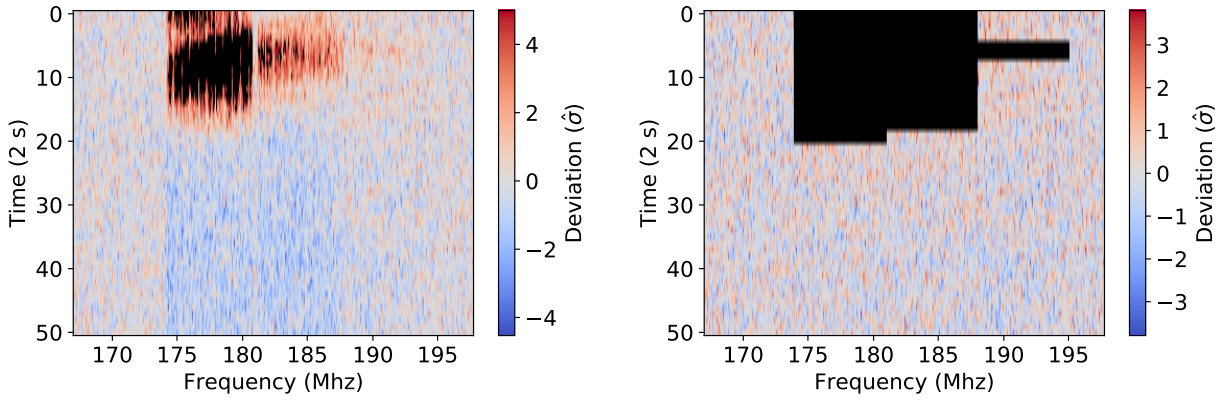


Figure 3.5: A flowchart of the iterative flagging procedure detailed in §3.2.4. The white area shows the basic iteration performed on the z-scores calculated in Equation 3.11, while the grey shows two additional steps that are included when frequency-matched flagging is implemented with the z-scores calculated in Equation 3.15.



(a) (b)

Figure 3.6: A comparison of single-sample thresholding and frequency-matched flagging on the contaminated observation of the previous figures. Black data is flagged. In (a), we show the results of iteratively flagging single-sample outliers beyond the $\tau = 5$ threshold. Leftover RFI persists, since some RFI is beneath this threshold on a single-sample basis. In (b), we show the results of frequency-matched flagging, where we have used bandwidth information about Western Australian DTV to inform the flagger. No remaining interference can be seen in the spectrum. A movie of each flagging iteration is available on reasonable request to the author.

Formally, if an RFI signal occupies a sub-band, B, that spans N_B frequency channels of the instrument between the signal's lower frequency, ν_L, and upper frequency, ν_U, we take

$$Z_B(t_n, p) = \frac{1}{\sqrt{N_B}} \sum_{\nu=\nu_L}^{\nu_U} Z(t_n, \nu, p). \tag{3.15}$$

This sum is precisely constructed so that a clean Z_B will be standard normal distributed just like a clean Z, so that the same significance threshold can be applied to any sub-band as is applied to a single sample. When RFI in a sub-band is observed for only part of the observation, even a sequence of very weak positive (or negative) outliers at a given time can sum to be greater in absolute value than the desired significance threshold, thereby increasing the sensitivity to RFI that occupies that sub-band. We can then iterate similarly as in the single-sample case to flag the observation for sub-band outliers in addition to single-sample outliers, where now we identify the strongest outlier among all suspected RFI sub-bands and times in the observation and then proceed as before. We show an example of the frequency-matched flagger in Figure 3.6(b), where it successfully excised DTV RFI belonging to multiple broadcasting channels. A movie showing the status of the observation in each iteration is available on reasonable request to the author.

The SSINS frequency-matched flagger can be adapted to search for any RFI contaminant within the observing band, including RFI that occupies the entire band. In §3.3, we illustrate these adaptive capabilities using specific examples of common RFI signals that we observe with the MWA.

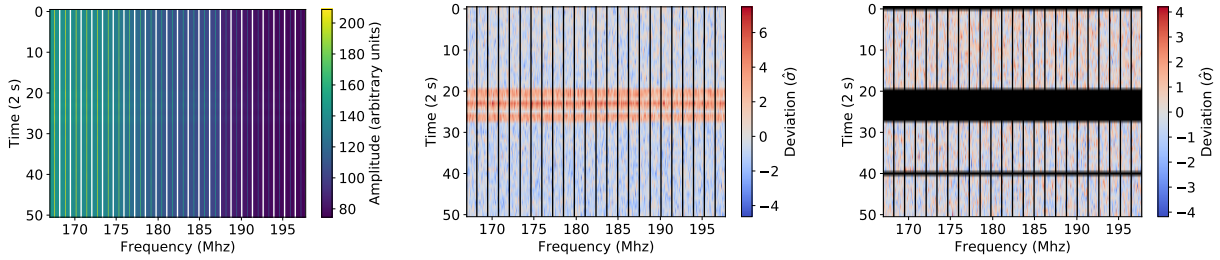
3.3 Some Common RFI Occupants in MWA Data

The sensitivity boost afforded by the baseline averaging allows SSINS to identify a wealth of faint RFI that goes undetected by single baseline algorithms. However, RFI that persists throughout the entire observing time can still evade SSINS due to the fact that such RFI will always contaminate the mean estimation no matter how deeply it is flagged. We have successfully implemented methods in SSINS to compensate for this behavior. Below, we explore the behavior of SSINS through examples for three different RFI classifications:

broadband streaks, narrowband RFI, and DTV. In each case, the data was passed through AOFLAGGER before making the incoherent noise spectra, except where indicated in §3.3.2.

3.3.1 Faint Broadband Streaks

First, we show an observation with faint streaks that are not band limited. See Figure 3.7. These streaks are too faint to be detected by the standard single-baseline AOFLAGGER implementation and appear to be quite common in our brief survey of MWA EoR Highband data. The exact nature of these streaks is so far unknown. Since they are not band-limited, we cannot directly appeal to our knowledge of the RFI environment in the same way that we can with the DTV interference shown earlier. It is possible that they are instrumental in origin. Despite not knowing the physical origin of such events, these streaks still occupy a programmable sub-band for the frequency-matched flagger, that is, the whole observing band, and so we can still adequately flag them using the algorithm in this paper.



(a) (b) (c)

Figure 3.7: The SSINS of an observation which show faint broadband streaks. The pre-existing flags in (a) and (b) are over the coarse band edges, which are routinely flagged in pre-processing due to systematic difficulties. The streaks are barely noticeable to the eye in (a), but the mean-subtracted spectrum in (b) shows them prominently. The frequency-matched flagger excises these streaks neatly, as shown in (c). These exceedingly faint events are often missed by AOFLAGGER and are not uncommon. They can appear isolated within an observation, or in series like this.

3.3.2 Narrowband Interference

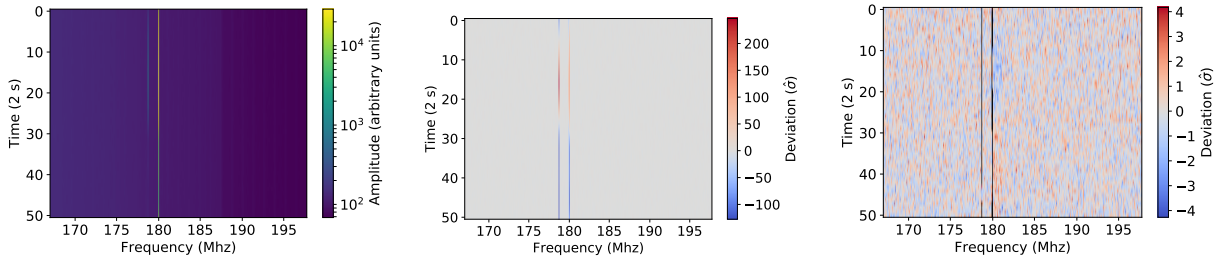
Next we show an example of narrowband interference observed by the MWA that only occupies one or two fine frequency channels (Figure 3.8). Though this RFI is present in the observation at dramatically fainter levels after AOFLAGGER, we show the pre-AOFLAGGER spectrum in this case because of the interference's proximity to a coarse band edge, which makes viewing the event difficult after routine flagging of the edges.

In this case, there are two interference events that are present through the entirety of the observation. The higher frequency line is a full order of magnitude brighter than the lower frequency line, and they are separated by exactly one coarse channel. We hypothesize that the dimmer line is caused by the brighter line's proximity to a coarse band edge, which are prone to aliasing due to the cascaded Fourier transform in the digital signal path[4].

Narrowband RFI such as this that persists throughout the entire observation poses a detection problem for the SSINS algorithm, since the estimation of the thermal parameters will always be contaminated by the RFI. The statistical properties of the RFI usually do not perfectly resemble the thermal noise, but even so, such RFI is almost always incompletely flagged by the SSINS frequency-matched flagger. Rarely, there is even RFI whose brightness varies so little over the course of the observation that it is totally camouflaged by the mean-subtraction step. Clearly, narrowband RFI that only occupies some fraction of the observation does not have this problem. In Figure 3.8(c), we show the mean-subtracted spectrum after frequency-matched flagging. We can see that the fainter event (alias) was flagged almost completely. If we examine the brighter event at the higher coarse channel, we notice the flag mask is thicker at some times compared to others. The RFI seems to occupy two adjacent fine frequency channels. The times indexed 10 to 20 are missing flags in the higher fine frequency channel, while the times indexed 30 to 48 are missing flags in the lower fine frequency channel.

We handle incomplete flagging in the following way. First, note that as we flag more times

[4]See "Aliases" in `https://wiki.mwatelescope.org/display/MP/Memos?preview=/14156367/18481185/MEMO_CascadedFT_2012_05_25.pdf`



(a) (b) (c)

Figure 3.8: Example of narrowband RFI that lasts the entirety of an observation. Note the logarithmic color scale in (a). This extremely bright event is a full two orders of magnitude brighter than the typical clean spectrum, and so the finer structure typically seen in an MWA incoherent noise spectrum is washed out in this colormap. We also note a disturbance precisely one coarse channel wide to the right of the main interference event in (c), characterized by a blue trough followed by a red excess. This is the only known instance of this feature in MWA data so far. Its exact nature is unknown. It may be related to the sheer brightness of this RFI event.

in a channel, fewer samples enter into the estimation of the mean. Eventually, the uncertainty in the estimator will be so high that the estimate will be untrustworthy. Furthermore, as more and more data within a channel are flagged, the chance that clean data remain in that channel diminishes. Combining these ideas, we set a threshold where once the amount of remaining unflagged data falls below the threshold, the entire remainder of the channel is flagged. The exact value that this threshold should take depends on the RFI environment of the telescope. While many narrowband events in the MWA survey shown in §3.5.2 would demand a threshold as aggressive as 0.7 (flag if less than 70% remains), an extremely brief survey of HERA data shows that a threshold of 0.25 successfully flags many of HERA's persistent narrowband occupants.

3.3.3 A DTV Signal

Finally, we show an observation with DTV interference that was partially caught by AOFLAGGER. In this case, the DTV was caught on some baselines, but not all of them, so we can see a leftover DTV footprint in the noise spectrum. See Figure 3.9.

We perform a similar demonstration as with the fainter DTV events shown in §3.2.4 by first flagging only bright single-sample outliers and then checking to see if fainter occupants lie beneath the single-sample threshold (Figure 3.9(c)). A notable feature that emerges from single-sample flagging is that the cluster of significant outliers around the DTV event seem to span more than 7 Mhz. The excess width of this event is not seen in the pre-AOFLAGGER spectrum (not shown), indicating that this is a pre-processing artifact. As described in [67], AOFLAGGER deploys a morphological detection algorithm that can overflag broad contaminants in a baseline by an amount proportional to the algorithm's user-set aggression threshold. Indeed, summing the AOFLAGGER flags for this observation over the set of baselines does show overflagging of this event in frequency by an amount similar to the feature we see in the incoherent noise spectrum. A feature of this size manifests in the incoherent noise spectrum due to the fact that visibilities have been averaged in time and frequency relative to the operating time-frequency resolution of AOFLAGGER. Flags are applied before averaging, so time-frequency bins with fewer samples entering them will have noise that has not been averaged down as much compared to those bins in which all possible contributing samples were averaged together. The overflagged bins contributing to the incoherent noise spectrum are then brighter than the surrounding uncontaminated ones, thus appearing like RFI to our statistical test. We can adapt the frequency-matched flagger with a custom sub-band to identify this pre-processing artifact. The results of frequency-matched flagging are shown in Figure 3.9(d).

This type of DTV interference is extremely common in MWA EoR Highband Observations. Roughly one third of the observations included in the EoR limit in [14] had some trace of DTV RFI according to the frequency-matched flagger. These observations were

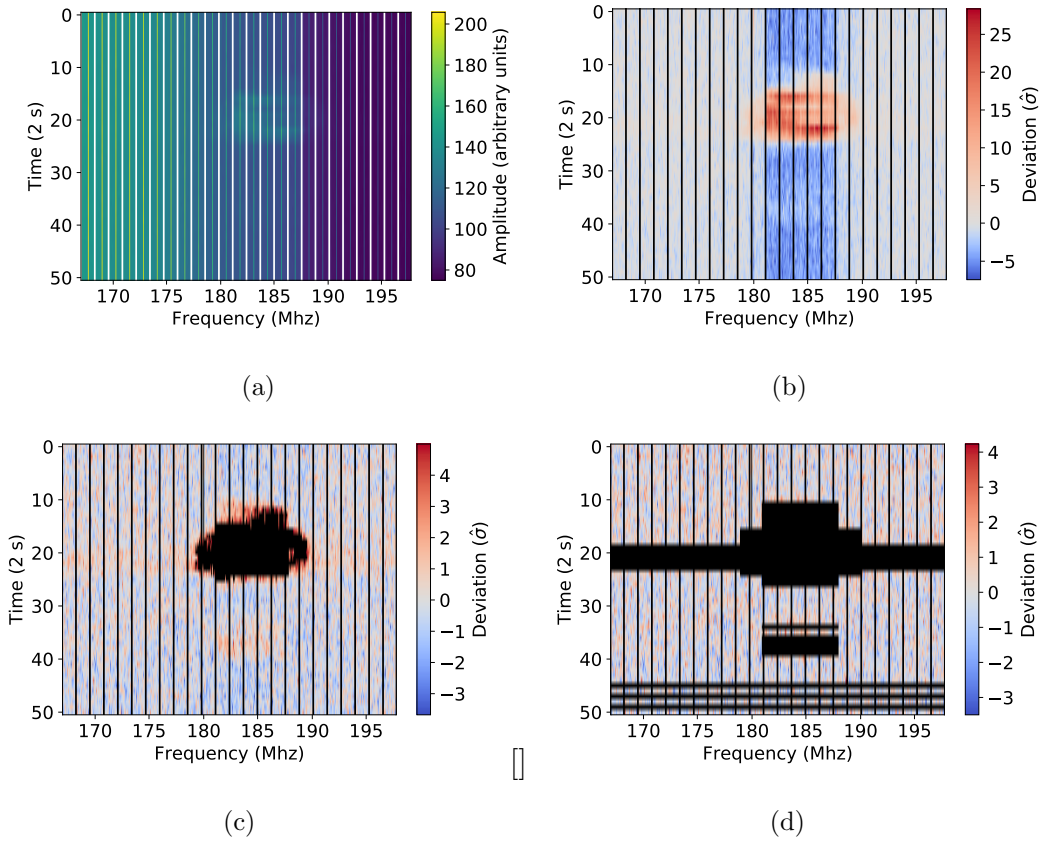


Figure 3.9: The SSINS of a DTV example after applying AOFLAGGER (a), alongside the mean-subtracted spectrm (b), as well as the results of single-sample iterative flagging (c), and frequency-matched flagging (d). From the flags reported by AOFLAGGER, we know it was caught on some baselines, however there is clearly leftover DTV corresponding to DTV channel 7. Several features emerge after single-sample flagging in (c). First, it appears that there is a broadband streak (§3.3.1) simultaneous with the DTV interference. Second, it appears there may have been a second DTV interference event later in the observation and much fainter. Third, the DTV interference seems to have associated outliers that span more than the advertised 7 Mhz, which is a pre-processing artifact described in the main text. With the frequency-matched flagger we can search for DTV, broadband streaks, single-sample outliers, and the pre-processing artifact simultaneously. The results are shown in (d), where all notable features are excised.

subsequently removed in a reduction of the same data set in [10], as discussed in §3.5.2. While in this case AOFLAGGER was able to remove some of the interference, it is often the case that the contamination is beneath the level where AOFLAGGER can make an appreciable difference.

3.4 Imaging DTV Interference

The MWA's extremely remote location makes it unlikely that it would directly observe a DTV transmission. However, it is clear from noise spectra such as those in the previous sections that the MWA is observing DTV. There are several hypotheses to explain this, including tropospheric ducting and reflection off of aircraft or satellites. We explore these possibilities by imaging DTV events found using SSINS.

The images were made using the Fast Holographic Deconvolution software[5] (FHD; [8, 86]). So that RFI could not drive the calibration solution, we calibrated on a clean part of each observation and transferred that calibration to the contaminated part of the respective observation. In order to see these faint DTV events in the images, it was necessary to subtract out 52000 sources from the GaLactic and Extragalactic All-sky MWA (GLEAM; [36]) catalog using FHD. Example images using the observations from Figures 3.4(b) and 3.9 are shown in Figures 3.10 and 3.11, respectively. We separated Figure 3.11 into 2s-snapshots and made a movie (available on reasonable request), where it is clear that some sort of flying object moving nearly due North-South is reflecting DTV into the array.

Due to the nonzero extent of the moving source, we can rule out the possibility of satellites and near-Earth asteroids, which would appear point-like. Aircraft are reflective and fly low enough to appear extended in the image. The speed (extent) of the object is atypically slow (large) for a commercial jet such as a Boeing 737 flying at a typical cruising altitude of 10 km. A parallactic estimation of the object's altitude[6] ultimately proved inconclusive due to the faintness of the RFI (Xiang Zhang, personal communication). As an alternative

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

[6] See [47] and citations therein for other uses of this type of altitude estimation

candidate, we suggest this could be a slow, low-flying bush plane whose apparent angular size may be due to its low height and near-field effects.

3.5 Frequency-Matched Flagger Customization

In this section, we describe the process of developing a customized SSINS frequency-matched flagger for a radio telescope. This is an exploratory process wherein the user becomes familiar with the incoherent noise spectra of their telescope. Some of the spectra features will be due to RFI, while other features may point to subtleties of the instrument. Once equipped with a thorough catalog of sub-bands to search, the user may process large amounts of data in an automated way. To exemplify what can be attained from this process, we summarize the RFI occupancy analysis for the data used in the EoR limit presented in [10].

3.5.1 Exploring the Data and Developing a Frequency-Matched Flagger

The first step to applying SSINS to a dataset is acquiring the software. It is implemented in PYTHON along with a comprehensive set of unit tests. It is publicly available on GitHub[7], where the user will also find installation instructions including the list of dependencies: PYUVDATA[8], NUMPY, SCIPY, SIX, H5PY, PYYAML, ASTROPY, and optionally MATPLOTLIB. There are tutorials and other documentation available[9] with simple usage examples for getting started.

After becoming familiar with the software, the next step is to examine the data, typically by hand, without attempting to use the frequency-matched flagger. While going through a survey of observations baseline by baseline in a modern radio telescope would be untenable, SSINS compresses the data of each observation into two dynamic spectra per polarization: a raw incoherent noise spectrum and its mean-subtracted form. It is useful to look at both

[7] `https://github.com/mwilensky768/SSINS`

[8] `https://github.com/RadioAstronomySoftwareGroup/pyuvdata`

[9] `https://ssins.readthedocs.io/en/latest`

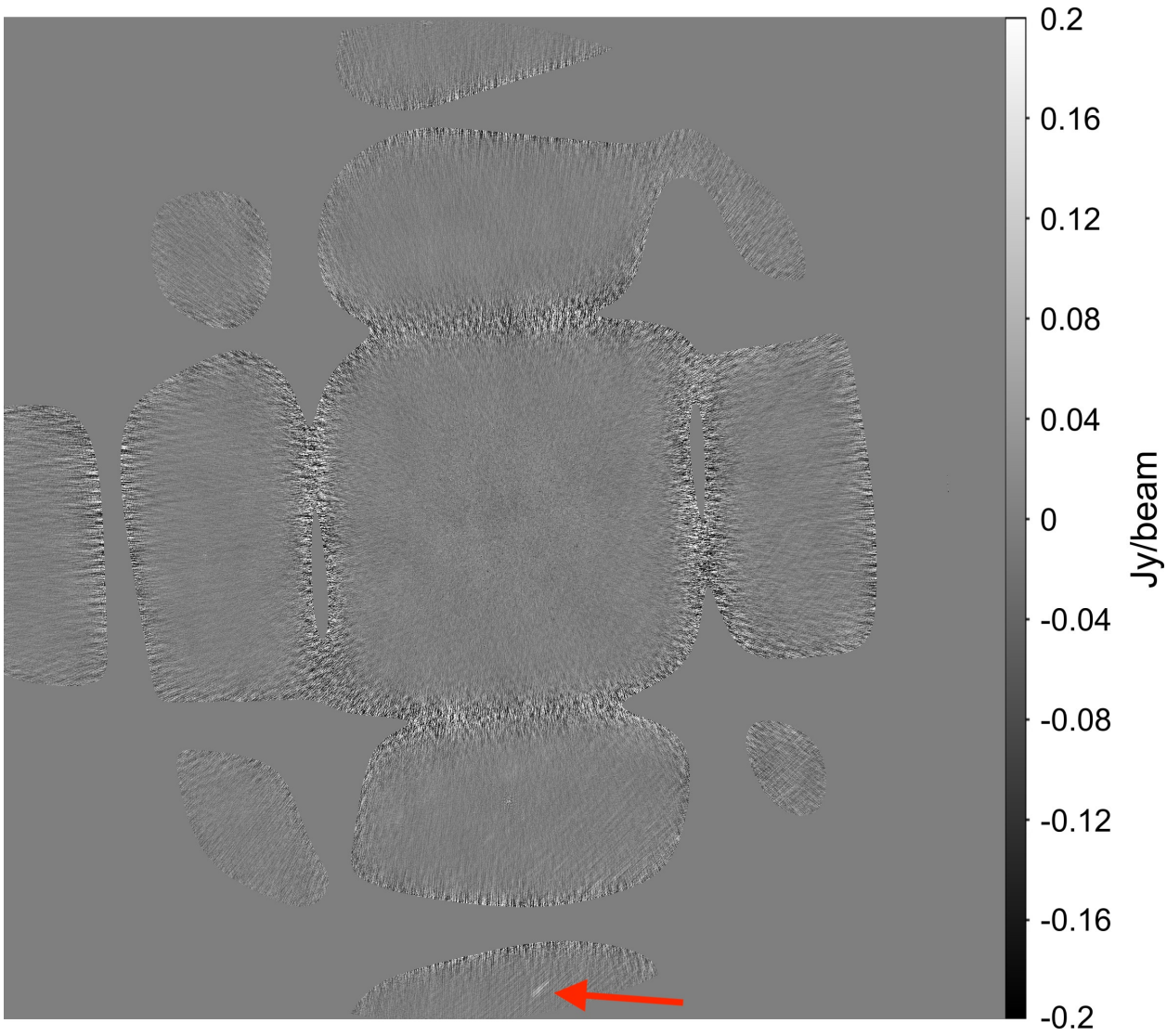


Figure 3.10: A full-sky horizon-to-horizon image of the DTV channel 6 event in Figure 3.4(b), with 52000 GLEAM sources removed using FHD. Other than the shape of the MWA beam and some diffuse structure not included in the calibration/subtraction model, the image is largely featureless. However, all the way in the second southern sidelobe lies a faint streak belonging to the DTV6 event (annotated with a bright red arrow). We hypothesize that the DTV signal is reflecting off of an aircraft on the southern horizon into the array. This is but one example of a collection of observations that show a similar feature in the second southern sidelobe, some of which are at the same time of night but on different dates, as one might expect from scheduled flights to and from Perth.

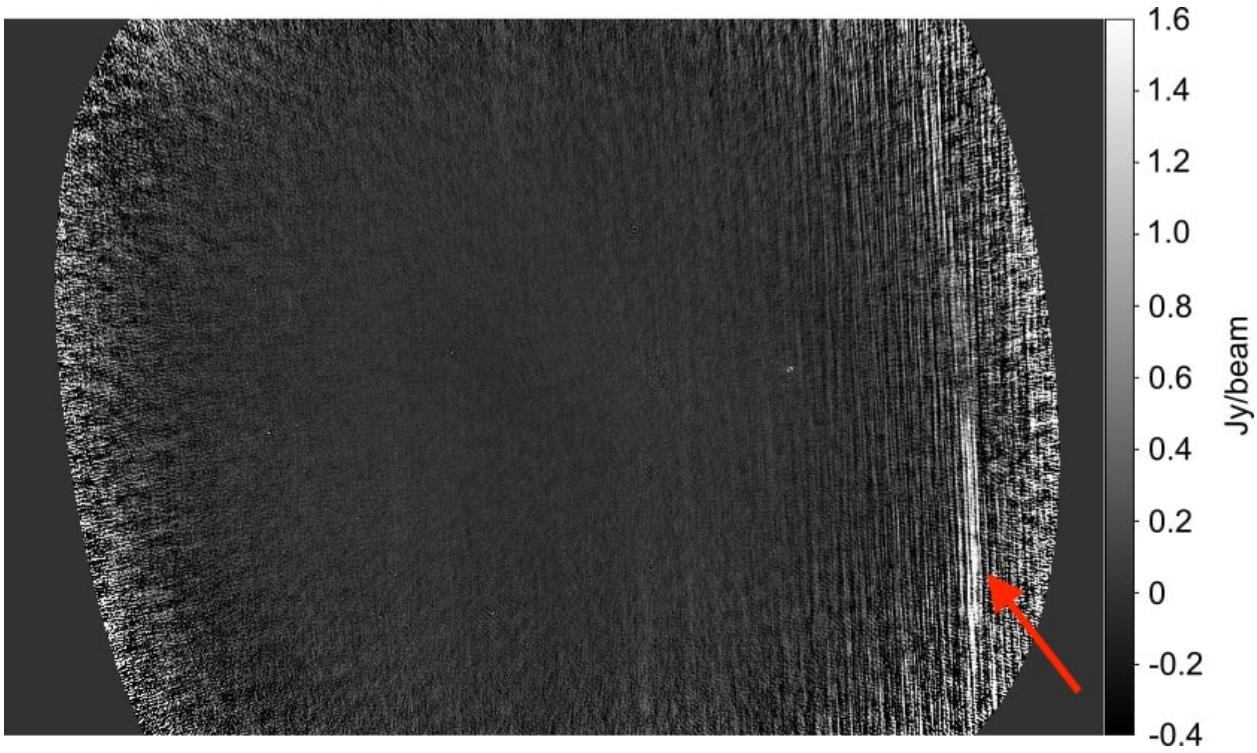


Figure 3.11: The Stokes I residual image of the DTV event in Figure 3.9, showing just the primary beam (center lobe in Figure 3.10). This is 34 seconds of data. The broad North-South streak in the Eastern edge of the primary beam (annotated with a bright red arrow) suggests possible motion of a source - likely a reflective aircraft. A snapshot-by-snapshot movie is available upon reasonable request to the author.

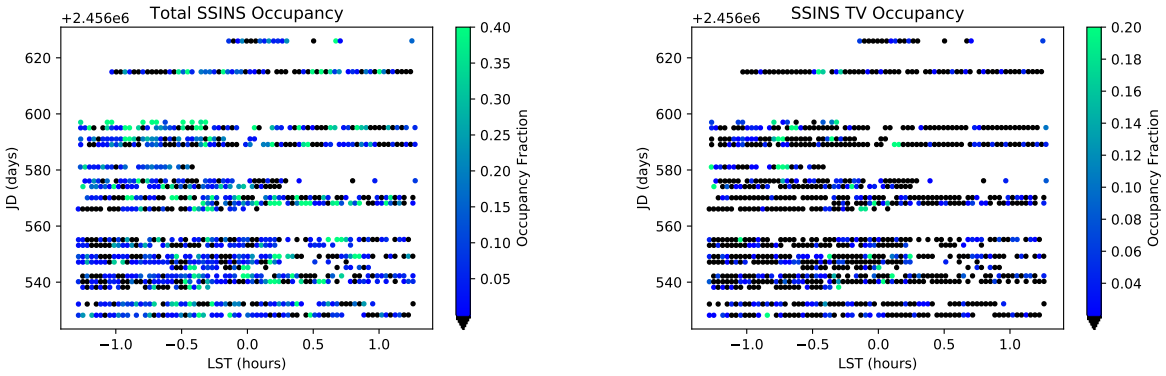
forms of the spectra side-by-side, particularly if narrow RFI is expected. In this way, an entire season of observations can be examined in short order.

While features in the incoherent noise spectra may correspond well with known factors in the RFI environment, it is also likely that SSINS will reveal RFI and even other instrumental features that had not been previously observed or considered. As the user catalogs the various characteristics of the data, they can build a dictionary of different sub-bands or single frequencies that appear commonly occupied. These occupations, regardless of their physical cause, can be input into the frequency-matched flagger settings. For example, some faint features in HERA incoherent noise spectra were later found to correspond with correlator malfunctions, some of which had been previously identified in other cases through other means. While technically this was not RFI per se, it was a feature that could be identified and removed with SSINS.

Once the set of identifiable contaminants has been collected, the frequency-matched flagger can be deployed for a first round of flagging. The frequency-matched flagger is very quick. A 2-minute, 30 Mhz MWA spectrum can be flagged in less than a second, while a 10-minute, 100 Mhz HERA spectrum that is significantly more contaminated typically takes between 20 and 60 seconds. As with the initial inspection, the results of the frequency-matched flagger can be quickly ascertained by eye. Oftentimes, fainter RFI than could be initially seen in the noise spectra is unearthed as the frequency-matched flagger iterates. Of course, the user will wish to add these to the dictionary of occupants. If no more identifiable sub-band occupancies are revealed, one may wish to adjust the parameters of the flagger, such as the significance threshold. This iterative process of data examination and deep cleaning is often illuminating.

3.5.2 Results with a Season of MWA Data

The SSINS package itself was developed and tested in the manner described above using a season's worth of data from 2013. This same dataset was used for the EoR power spectrum limit featured in [10] and [14]. Below, we describe the filter settings for an RFI analysis of



(a) (b)

Figure 3.12: (a) A scatter plot of the total RFI occupation in a season of data as seen by SSINS. (b) A scatter plot of the DTV occupancy as seen by SSINS. In each panel, each circle represents a single two-minute observation. Each line of circles is a night of observations on the Julian Date (JD) shown on the vertical axis, while the horizontal axis gives that observation's local sidereal time (LST) in hours. Black circles were not found to have additional RFI after AOFLAGGER. All observations outside of the plotted LST range and all missing circles in a night are observations that were removed by a previous jackknife test, detailed in [14]. In (a), the color shows the fraction of samples in the noise spectra found to be contaminated, disregarding coarse band edges, while in (b), the color shows the fraction of times contaminated by DTV interference, regardless of which broadcasting channels were identified. These plots provide an occupancy overview of the season, letting one pick out particularly bad days or times of night by eye, such as the line of high occupancy observations nearest JD 2456600, which was a day that also had many observations removed by the previous jackknife test.

this dataset and also summarize occupancy levels.

The data in this analysis was pre-processed with COTTER, which uses AOFLAGGER to identify and flag bright RFI. Then, SSINS was run on the data in order to identify and catalog leftover faint RFI. We used a significance threshold of 5 and sought single sample outliers, broadband streaks, DTV interference, as well as the broader DTV sub-band discussed in §3.3.3, all of which are occupants detailed in §3.3. We also completely flagged a fine frequency channel if it ever reached an occupancy fraction of 0.7 during frequency-matched flagging. We present RFI occupancy for the season in Figure 3.12, summarizing total RFI occupancy as seen by SSINS after AOFLAGGER as well as DTV occupancy. Such figures can be used to pick out particularly bad days from the season and search for patterns in time.

The occupancy data from SSINS was used to make data cuts for the sake of improving the EoR limit in [10]. We did not reflag the data using SSINS, but instead just used its outputs to develop quality metrics for observations. Ultimately, we cut all observations with any trace of DTV as well as all observations with greater than 40% occupancy. We found that this improved the limit despite removing roughly 1/3 of the observations originally included.

3.6 Discussions and Conclusions

We have described the SSINS RFI detection algorithm in detail. The substantial sensitivity boost afforded by the incoherent average over the baselines allows for detection of faint RFI that escapes other high-performing single-baseline algorithms. We demonstrated its effectiveness on several different types of RFI found in the MWA EoR highband, including DTV interference. We have implemented countermeasures for persistent RFI with some success, but we plan to make improvements to these detection efforts. Overall, we observe that the increased sensitivity afforded by SSINS helps us understand the general pervasiveness of faint RFI.

Developing a custom SSINS frequency-matched flagger for a new RFI environment is an iterative process that is often quite instructive. We confirmed the quality of our own

EoR highband frequency-matched flagger by successfully imaging faint DTV reflected from aircraft as well as successfully improving the EoR limit in [10]. To ease the development process, we have written tutorials for basic software usage and provided an overview of how SSINS can be used in practice.

Though SSINS is capable of detecting fainter RFI than current state-of-the-art algorithms and is a general improvement to the field, we plan to further improve the methodology to exploit its full capabilities. We propose the following items as future work:

- **Improved Narrowband RFI Detection**: The current implementation tends to underflag narrowband RFI that persists through the entire observation. An algorithm that preliminarily identifies entirely contaminated channels and interpolates over these channels during mean estimation will allow for more complete flagging.

- **Blind Sub-band Hunting**: Presently, the user must input specific sub-bands into the frequency-matched flagger that are established from knowledge of the RFI environment or a hand grading of the data. An algorithm that searches for likely occupied sub-band candidates would ease analysis of extremely large data sets.

As SSINS is applied to more radio projects in the future, more possibilities for enhancements will be found. In this light, we look forward to exploring the full capabilities of the SSINS framework. In the next chapter, we theoretically investigate the effect of residual RFI contamination on power spectrum measurements.

Chapter 4

QUANTIFYING EXCESS POWER FROM RADIO FREQUENCY INTERFERENCE IN 21-CM POWER SPECTRA

This chapter is largely based on work in [101].

4.1 Introduction

In this chapter, we quantify the effect of RFI on measurements of the 21-cm power spectrum during the EoR. Specifically, we investigate how the frequency structure of RFI source emission generates contamination in higher-order wave modes that is much more problematic than smooth-spectrum foreground sources. Using a relatively optimistic EoR model, we find that even a single relatively dim RFI source can overwhelm the EoR power spectrum signal of ~ 10 mK^2 for modes $0.1\ h\ \mathrm{Mpc}^{-1} < k < 2\ h\ \mathrm{Mpc}^{-1}$. If total apparent RFI flux density in the final power spectrum integration is kept below 1 mJy, an EoR signal resembling this optimistic model should be detectable for modes $k < 0.9\ h\ \mathrm{Mpc}^{-1}$, given no other systematic contaminants and an error tolerance as high as 10%. More pessimistic models will be more restrictive. These results emphasize the need for highly effective RFI mitigation strategies for telescopes used to search for the EoR.

In section §4.2, we lay the foundation for the theoretical formalism used in this paper and state assumptions that are used. In §4.3, we first calculate the expected power spectrum of several common types of RFI source using this formalism. Using a fully functional imaging and power spectrum pipeline, we then simulate RFI power spectra and compare them to simulated power spectra of sources from the GLEAM sky catalog [36] as well as theoretical expectations for the EoR signal. In §4.4, we draw conclusions from this work.

4.2 Theoretical Formalism

In this section we establish the theoretical formalism that is used for the power spectrum calculations in this work. This includes choices of notation, definitions of useful quantities, and assumptions about quantities that affect the calculation.

4.2.1 The Power Spectrum Estimator

Since the goal of this work is to extract the effect that an RFI source has on the power spectrum, we will make several simplifying assumptions in order to ease theoretical calculations. We will later compare these theoretical calculations to realistic in-situ simulations wherein qualities of the instrumental measurement process are included, so that we can be sure that the gross effects predicted from the theoretical calculations are still present along with the nuances that arise from measurement. For analytic descriptions of EoR power spectra that include effects of the instrument and analysis choices, see [44, 45].

Astrophysical foreground contamination is largely contained within a wedge-shaped region of the cylindrically averaged Fourier domain known as the foreground wedge, e.g. [23, 56, 71, 91, 95]. Without extremely accurate foreground removal, this contamination excludes lower-order line-of-sight modes in power spectrum measurements. The depth of exclusion depends linearly on the length of the perpendicular mode in question, hence the wedge shape. The higher-order line-of-sight modes with strongly diminished foreground contamination form the EoR window, e.g. [44, 45]. As is shown in §4.3, the most problematic feature of RFI seen in this analysis is that it provides very strong excess power in the EoR window. This contribution arises strictly from the intrinsic frequency structure of RFI. With this in mind, we will ignore the chromatic point-spread function that leads to the foreground wedge in order to simplify the theoretical calculations. Moreover, concentrating on the spectral characteristics of the RFI makes the theoretical derivation equally applicable for all EoR power spectrum analyses [55].

We also choose to ignore the fine details of the bandpass response of the instrument,

chromaticity of the primary beam, as well as chromatic errors in calibration. These effects are all deeply important in power spectrum estimation [9, 14, 19, 25, 72]. While generally not negligible, they are beyond the scope of the theoretical calculations in this work. We will include the fact that an instrument possesses a limited field of view and is only sensitive to some range of radiation frequencies. Since it is highly relevant to this analysis and convenient to implement theoretically, we will also include a frequency tapering function for this range of frequencies. A frequency tapering function is commonly used in power spectrum estimation that gracefully limit foreground spillover into the EoR window [91] and ensure that a sufficiently narrow range of frequencies is used to prevent the influence of cosmological evolution [57].

Bearing these assumptions in mind, we define the power spectrum estimator using the following expression:

$$P(\mathbf{k}) = \frac{1}{V_{\mathcal{M}}} \left| \int_{\mathcal{M}} \mathrm{d}^3\mathbf{r} A\big(\boldsymbol{\theta}(\mathbf{r})\big) \Psi\big(\nu(r_\parallel)\big) I\big(\boldsymbol{\theta}(\mathbf{r}), \nu(r_\parallel)\big) \mathrm{e}^{-\mathbf{i}\mathbf{k}\cdot\mathbf{r}} \right|^2, \tag{4.1}$$

where $\mathcal{M}$ is the cosmological region of space for which the power spectrum is being constructed, $V_{\mathcal{M}}$ is the cosmological volume of that region. In this analysis, the primary beam, A, and the frequency tapering function, Ψ, are more readily expressed and understood in terms of angular coordinates and the observed frequency of emission, ν, hence the listed arguments of these functions above. For the relatively narrow frequency bands considered in typical EoR analyses, the relation between comoving line-of-sight distance to a narrow emitter and its observed frequency of emission is very nearly linear i.e. for the purposes of calculation, the following relationship is justified:

$$r_\parallel(\nu) \approx \beta\nu + c_0. \tag{4.2}$$

The slope of this linear relation, β, is given by equation 2.6. Physically, c_0 is the extrapolated comoving distance at infinite redshift, which is not meaningful by itself since the relationship is not genuinely linear. In this case, the constant merely serves to make the linear relation consistent with the universe's geometry for the relevant redshift. A linear

relationship allows the Fourier transform over the line of sight coordinate to be directly written as a Fourier transform over frequency instead, which is both helpful when considering observational quantities and also what is done in the practice of power spectrum estimation.

4.2.2 Calculating a Power Spectrum for Point Sources

First, we calculate the power spectrum for a single point source. For this case, equation 4.1 can be written as

$$P(\mathbf{k}) = \frac{1}{V_{\mathcal{M}}}\left|\beta \int_{\nu_L}^{\nu_U} \mathrm{d}\nu\, \mathrm{e}^{-\mathrm{i}k_\parallel r_\parallel(\nu)} \Psi(\nu) D_M(\nu)^2 \phi(\nu)\kappa(\nu) \int_{FoV} \mathrm{d}^2\boldsymbol{\theta}\, \mathrm{e}^{-\mathrm{i}D_M(\nu)\mathbf{k}_\perp\cdot\boldsymbol{\theta}} I_0 \delta^2(\boldsymbol{\theta}-\boldsymbol{\theta_0}) A(\boldsymbol{\theta})\right|^2, \tag{4.3}$$

where we have chosen to write the transverse spatial integrals as angular integrals since that is a more natural coordinate system in which to express the flux of a point-source as well as the frequency tapering function and primary beam (whose achromaticty is now explicitly manifested). Here, ν_L and ν_U represent the lower and upper bounds of the frequencies observed by the instrument, $D_M(\nu)$ is the transverse comoving distance [35] introduced in the coordinate conversion from $\mathbf{r}_\perp$ to $\boldsymbol{\theta}$, $I_0\phi(\nu)$ is the flux density of the source measured in Jy (I_0 is its total flux), $\kappa(\nu)$ converts this flux density into brightness temperature, and the Dirac delta function represents the point-like nature of the source at angular position $\boldsymbol{\theta_0}$. Evaluating the angular integral and assuming a flat cosmology[1] leaves

$$P(\mathbf{k}) = \frac{1}{V_{\mathcal{M}}}\left|\beta \int_{\nu_L}^{\nu_U} \mathrm{d}\nu\, \mathrm{e}^{-\mathrm{i}(k_\parallel + \mathbf{k}_\perp\cdot\boldsymbol{\theta_0}) r_\parallel(\nu)} I_0^{\mathrm{app}} \phi(\nu)\kappa(\nu)\Psi(\nu) r_\parallel(\nu)^2\right|^2, \tag{4.4}$$

where $I_0^{\mathrm{app}} = A(\theta_0)I_0$ is the apparent flux of the source. At this point, even with the simplification afforded by equation 4.2, this integral is only analytically calculable for special cases. To go further will require a specification of the frequency dependence of the source and typically some numerical methods.

Placing the source at phase tracking center eliminates the $\mathbf{k}_\perp$ dependence in the power spectrum, thus the cylindrically averaged power spectrum of a source at phase tracking

[1] This lets one equate the transverse and line-of-sight comoving distances. See [35].

center will be constant along lines of constant $k_\parallel$. Simulation results do not show a strong difference between power spectra of centered and off-center RFI sources, particularly within the EoR window. For this reason, we will only show theoretical power spectra for sources at phase center. A brief analytic discussion of the effects of placing a source off of phase-center is presented in Appendix D.

4.2.3 Imposing a Frequency Dependence

In general, RFI sources may take on a great variety of frequency profiles. Many transmissions only occupy a single channel in an EoR analysis, although some can be relatively wide compared to the observing band.

Two types of common and relatively broad RFI signals observed in EoR analyses include DTV and digital audio broadcasting (DAB). These types of signals are well-approximated by top-hats in frequency over their allocation.[2] The widths and locations of these profiles are usually subject to matters of protocol in their country of origin. For this work we use the Western Australian 7 MHz wide allocations, as is frequently observed by the MWA. For example, see [68], [84], and [102].

We would write a top-hat dependence of width $\Delta\nu$ and center frequency ν_0 like so:[3]

$$\phi(\nu) = \frac{1}{\Delta\nu}\Pi\left(\frac{\nu - \nu_0}{\Delta\nu}\right), \tag{4.5}$$

where $\Pi(x)$ is the rectangle function, given by

$$\Pi(x) = \begin{cases} 1, & \text{if } |x| < \frac{1}{2} \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

Given the slight complication in frequency dependence from converting to brightness temperature and applying a tapering function, we choose to evaluate the integral in equation 4.4 numerically for the comparisons made in this work.

---

[2]Examples of technical standards for such signals can be searched for at `http://www.etsi.org/standards-search`. DTV broadcasts may be referred to as digital video broadcasting (DVB) in such manuals.

[3]The normalization in equation 4.5 ensures consistent units for I_0 in equation 4.4.

A sufficiently narrow signal (the resolution of the analysis or finer) can be formally handled using a Dirac delta function frequency dependence. This renders the Fourier transform over frequency analytically straight-forward. The result is a constant power on all k-modes.

Due to the relative ease of evaluating power spectra for narrowband point sources, one may glean information about ensembles of narrowband point-sources using this analytic machinery. We relegate the precise mathematical details to Appendix C and summarize the results here. In principle, any two sources could destructively interfere for some mode in the power spectrum, however, any two sources will interfere constructively for modes perpendicular to the separation vector of the sources. The power of any ensemble for some given mode is bounded from above by the case of total constructive interference; the power is proportional to the square of the sum of the fluxes. The modes for which this can be achieved are intimately related to the angular distribution of the emitters. For a random ensemble drawn from a uniform distribution over the sky and whose true flux distribution is independent of its uniform angular distribution, the expected power (average over realizations) is bounded from below by an incoherent sum of powers; it is proportional to the quadrature sum of the fluxes i.e. a linear sum of individual powers. For this particular distribution, baselines longer than the inverse width of the primary beam should observe relatively little coherence. Since this is the minimum length of a physical baseline, an angularly uniform distribution of RFI emitters are expected to add power incoherently for any mode in an estimated power spectrum, although specific realizations can defy this expectation. In conclusion, it is expected that adding many uniformly distributed RFI sources should tend to increase contamination relative to a single source in a predictable way. Therefore, if we can understand the strength of a single RFI source on a power spectrum measurement relative to the EoR, we can quantify the degree to which RFI needs to be mitigated to make an EoR detection feasible.

4.3 Results

In this section, we apply the theoretical formalism in §4.2 to make rough predictions about the general behavior and contamination levels of RFI in the power spectrum. In order

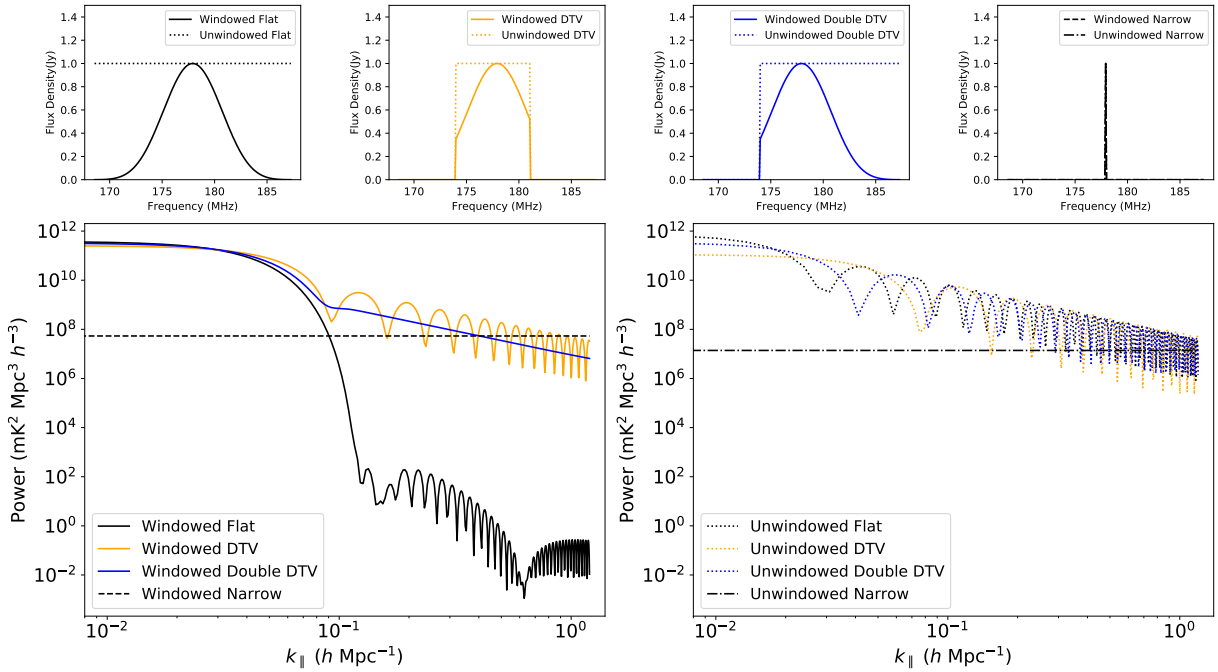


Figure 4.1: Theoretical frequency structures and power spectra along the line of sight for various 1-Jy point-sources. The top panels show the windowed (solid, narowband is dashed) and unwindowed (dotted, narrowband is dashed-dotted) frequency structure of each source in Jy, while the bottom panels show the corresponding power spectra of the windowed (bottom left) and unwindowed (bottom right) sources. The sources are placed at the phase center of a highly idealized instrument detailed in §4.2, so there is no $\mathbf{k}_\perp$ dependence. For the broader sources, the unwindowed spectra are essentially identical to one another except in the lowest-ordered k-modes. In the windowed case, we see that the flat-spectrum (foreground) source contamination has a sharp falloff starting at modes greater than ~ 0.05 h Mpc^{-1}. This is exactly the desired effect of the Blackman-Harris window function. However, the window function does not have nearly the same effect on the RFI sources, whose contamination remains many orders of magnitude greater than that of the foreground source. The narrowband source is in the middle of the observing band, so the window function has no effect on the frequency structure. However, the power spectra differ because the effective cosmological volume probed in the frequency-tapered case is smaller than in the non-tapered case.

to compare these predictions to realistic power spectra, we also simulate power spectra of RFI sources and sources from the GLEAM sky catalog using the FHD[4]/εPPSILON[5] power spectrum pipeline [8, 86]. Ultimately we find that, in both simulation and theory, RFI provides orders-of-magnitude more contamination than typical point-like foreground sources within the EoR window. We do not estimate power spectra for diffuse foregrounds, which are substantial. Including these may make foregrounds more competitive with RFI on some modes.

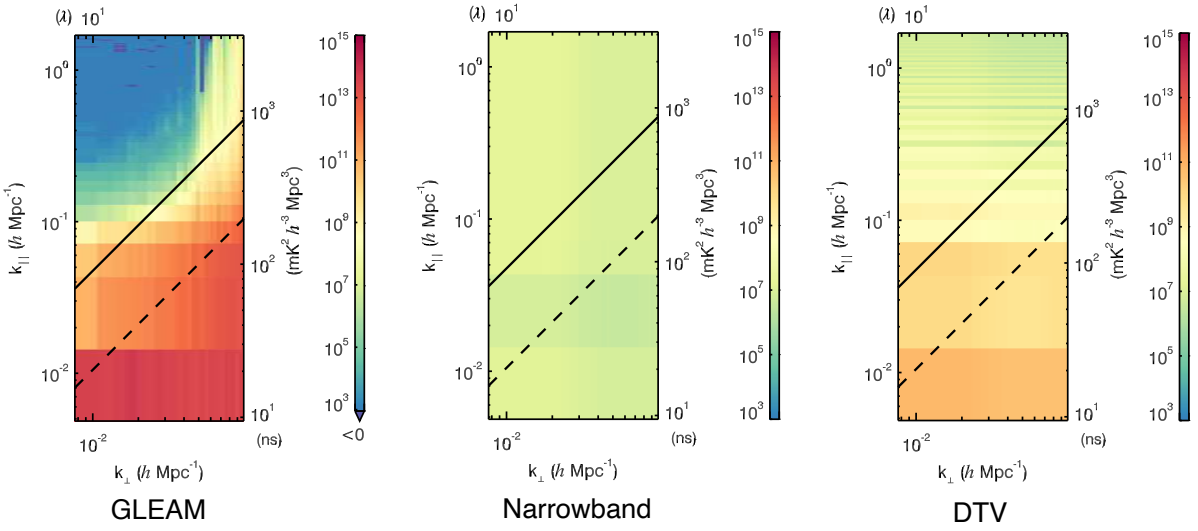
4.3.1 Theoretical Results for RFI Power Spectra

We specifically consider power spectra from sources with the following frequency dependences: flat broadband, a single DTV channel, two simultaneous frequency-adjacent DTV channels (both received from the same position e.g. from an aircraft reflection), and a narrowband source. A flat, broadband source emulates an astrophysical foreground source so that there is a baseline contamination level to which we can compare that afforded by band-limited RFI. For all sources, we numerically evaluate the integral in equation 4.4, with and without a Blackman-Harris tapering function. The results for a collection of 1 Jy sources at phase tracking center are shown in Figure 4.1. We used the same observing band as was used in the final limit in [10], centered on redshift 7 for the 21-cm line. Of striking significance is that the RFI sources universally provide dramatic contamination in the higher order line-of-sight modes in the power spectrum compared to the flat-spectrum (foreground) source due to sharp spectral cutoffs within the observing band.

As mentioned in §4.2, a frequency tapering function is typically employed to reduce power leakage of foregrounds into the EoR window, thus boosting the dynamic range of the measurement. The desired effect is shown in Figure 4.1, where the flat-spectrum source, which resembles an astrophysical foreground, displays a sharp falloff at higher $k_{\parallel}$ relative to its unwindowed power. This effect occurs because the tapering function smoothly approaches

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

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



(a) (b) (c)

Figure 4.2: Cylindrical power spectra for simulations with GLEAM (left), narrowband (middle), and DTV (right) sources. The solid line marks the scales corresponding to the horizon, while the dashed line demarcates the scale of the primary field of view. Smooth-spectrum source contamination is predominantly contained within the foreground wedge, which extends everywhere below the solid line. Above the solid line is the EoR window. The structure seen in these plots generally match our theoretical expectations. The EoR window contamination of a single 1-Jy RFI source is substantially higher than what is included from GLEAM.

zero at the boundaries of the observing band, eliminating sharp spectral cutoffs for this source [31].

On the other hand, the RFI sources are band-limited, and so the tapering function fails to eliminate the sharp spectral cutoffs of the RFI sources. Comparing the power spectra of windowed and unwindowed RFI sources, we observe that the tapering function has little effect on overall contamination levels, in great contrast to the flat-spectrum source. Generically, as the bandwidth of a source approaches the bandwidth of the observing window, and thus more closely resembles a foreground source, the EoR window contamination is more effectively reduced by the frequency tapering function. However, we see that even in the case of the two-channel DTV event, which occupies nearly three-quarters of the band, the window function makes hardly any difference relative to the single-channel DTV event. Additional calculations not shown indicate that an emitter must occupy greater than 95% of the observing band to appreciably close the several order-of-magnitude gap between the RFI and foreground contamination seen in Figure 4.1. Some experimentation shows that different tapering functions may close this gap more effectively, however this typically comes at the expense of the effective and crucial dynamic range boost offered by the Blackman-Harris window.

This suggests an interesting and very simple mitigation strategy for appropriately sized sources such as DTV RFI. If the observing band is set equal to the band allocated for the RFI source, then the tapering function will affect the RFI power spectrum in the same way the flat-spectrum source was affected in Figure 4.1. Of course, any narrower RFI source within the same allocation will be relatively unaffected by this choice. While regional radio allocations are set so that particular types of signals are broadcast over certain frequencies, RFI other than the regionally allocated signals can be observed through several mechanisms e.g. intermodulation products arising from nonlinearities in the instrument signal path. Furthermore, the size of the observing band is subject to important constraints, including sensitivity requirements for EoR detection and the need to preserve the assumption of isotropy over the probed cosmological volume. Therefore, this particular strategy is extremely limited in

scope, and should not be the primary mitigation strategy. The enormous contamination of RFI in the EoR window suggests that specialized post-excision mitigation strategies are needed if significant RFI sources are missed by excision algorithms. We specifically discuss the degree to which RFI sources need be mitigated in terms of allowable apparent flux density in §4.3.3.

The effectiveness of the tapering function on a narrowband source is strictly a function of the location of that source within the observing band. The window function multiplicatively adjusts the constant power level provided by the source by the square of the value of the window function at the frequency of the source. Here, we have chosen to keep the narrowband source at the center of the window, so that we could see the full effect of an unmitigated narrowband source.

There is also an interesting difference between the two-channel DTV event and the single-channel DTV event, which is that the single-channel event exhibits a lobed structure as a function of $k_{\parallel}$, while the two-channel event does not. This is an interaction between the lobes of the Fourier transforms of the DTV signal and tapering function that depends on the location of the event within the observing band [31]. This effect is of little practical consequence, since the overall contamination levels are roughly identical regardless of the presence of lobes.

As mentioned in §4.2, these theoretical calculations ignore many important aspects of measuring EoR power spectra with radio interferometers. In order to check these predictions against realistic power spectra made from instrumental visibilities and a fully functioning analysis pipeline, we employed the simulation capabilities of the FHD/εPPSILON pipeline. We also simulate the power spectrum of sources from the GLEAM catalog, so that we can compare simulated RFI power spectra to a power spectrum of a sky full of foregrounds.

4.3.2 Simulated RFI Power Spectra

We show cylindrical power spectra for sources from the GLEAM catalog and two different RFI sources in Figure 4.2. In these simulations, we used the MWA Phase I [92] as our

example instrument, thus capturing a realistic baseline distribution and beam pattern for an interformeter. The beam is simulated using an average embedded element and mutual coupling model [87], chosen at the central frequency of observation (about 178 MHz). This allows us to accurately depict some typical chromatic effects of an instrument and analysis pipeline. To simulate each RFI source, we modeled a single fictitious, phase-centered, flat-spectrum, 1 Jy source, and simulated visibilities using FHD. Using PYUVDATA[6] [32], we then modified the frequency structure of the fictitious source to match our desired RFI sources. We then passed these visibilities back through FHD to create the input HEALPix cubes [29] for εPPSILON. The simulated GLEAM sources lie in the same observing field as was used in [10], which is a region of the sky with minimal bright sources and low sky temperature.

As expected from our theoretical work, the narrowband source has a nearly constant power spectrum, while the DTV source has a lobed structure along $k_\parallel$ and relatively little structure along $k_\perp$. The power spectrum of the simulated GLEAM catalog demonstrates how foreground power is largely confined to the foreground wedge, with very little power escaping into the EoR window. Because the RFI sources are at phase center, they do not exhibit any sort of foreground wedge. While an off-center RFI source would not be immune to the effects that cause the foreground wedge, the amount of power leakage into higher line-of-sight modes from the wedge (a $\sim 0.01 - 0.1\%$ effect in the foregrounds of Figure 4.2(a)) is subdominant to the spillover that occurs from the band-limited nature of the RFI (a $\sim 10\%$ effect for the DTV power spectrum in Figure 4.2(c)). We can see that this spillover drastically contaminates the EoR window.

The cylindrical power spectra from Figure 4.2, averaged over the $k_\perp$ axis, are shown in Figure 4.3, along with some additional simulated power spectra for comparison to Figure 4.1. We can see that in a realistically simulated power spectrum, there is still a several order of magnitude discrepancy between the contamination of astrophysical foregrounds and that of RFI within the EoR window, thus confirming the earlier theoretical predictions. A natural

[6]https://github.com/RadioAstronomySoftwareGroup/pyuvdata

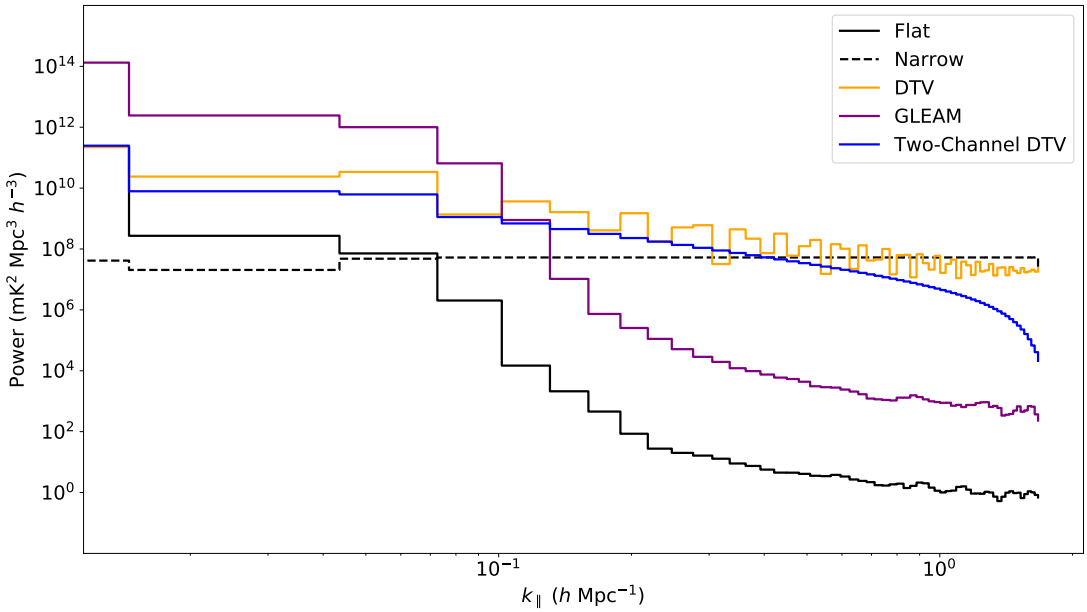


Figure 4.3: Simulated power spectra (averaged over $\mathbf{k}_\perp$) analogous to the theoretical spectra in Figure 4.1, with an additional line for a simulated power spectrum of sources from the GLEAM catalog. The RFI spectra show a strong correspondence with the theoretical prediction. The two-channel DTV power spectrum is seemingly smooth compared to the lobed single-channel power spectrum, although the difference in power is more pronounced at higher modes than theoretically predicted. The falloff for the 1-Jy foreground source is similar, although the theoretical lobed structure is not visible. There is still a large difference in contamination between the foreground source and the RFI sources. Moreover, even the contamination from the simulated GLEAM catalog has several orders of magnitude less contamination than a single 1-Jy DTV or narrowband source at high $k_\parallel$.

line of investigation is to see what level of RFI is sustainable given the high dynamic range required to measure the EoR signal in the power spectrum.

4.3.3 How to Develop an RFI Budget for EoR Detection

The final RFI budget for a given analysis depends on several qualities of the experiment. The type of RFI signal observed by the instrument as well as details about the RFI environment of the telescope, such as the number of emitters and their angular distribution, all affect the amount of RFI contamination. Furthermore, the cosmological modes of interest and actual strength of the EoR signal in those modes determines the amount of allowable contamination. To design a budget, the analyst must also decide on a model for the EoR. The final RFI budget designed by the analyst will of course be affected by the choice of model. An RFI power spectrum that is beneath the expected EoR signal by a significant amount for the modes of interest gives a budget for those modes.

In this work, we design some example budgets for different RFI realizations, adopting the fiducial model used in [10]. This is a relatively optimistic model made using 21CMFAST [53] and an astrophysical parameterization from [70]. We focus on modes between 0.1 and 2 h Mpc^{-1}, which represents the foreground avoidance strategy employed in MWA EoR analyses. For these modes the model predicts a signal strength of order $\Delta^2 \sim 10$ mK^2, where $\Delta^2 = (k^3/2\pi^2)P(k)$. A foreground removal strategy might endeavour to measure the EoR on lower modes where some models expect a stronger EoR signal, which would affect the overall RFI budget compared to these examples.

Designing a Budget for a Single Source

One could parameterize a budget for a single RFI source of given emission profile, a given EoR model, and certain modes by defining a signal-to-interference ratio (SIR) as a function of k:

$$\mathrm{SIR}(k) = \frac{\Delta^2_{\mathrm{EoR}}(k)}{\Delta^2_{\mathrm{RFI}}(k)}, \tag{4.7}$$

and demanding that this ratio be greater than a prescribed value. The power spectrum of a single RFI source is proportional to the square of its apparent flux, but will scale differently as a function of k depending on the frequency dependence of the source, meaning that the final budget for a source will be a function of k. For a narrowband source, this relation is given exactly by

$$\Delta^2_{\text{narrow}}(k) = \zeta(\nu_0)(I_0^{\text{app}})^2 \frac{k^3}{2\pi^2} \tag{4.8}$$

where

$$\zeta(\nu_0) = \frac{1}{V_{\mathcal{M}}} \Psi(\nu_0)^2 \kappa(\nu_0)^2 r_{\parallel}(\nu_0)^4 \beta^2. \tag{4.9}$$

We did not find an exact analytic formula for a DTV power spectrum in this work, particularly after applying a Blackman-Harris window. However, examining Figures 4.1 and 4.3 closely, one can see that the envelope for the DTV power spectrum goes roughly[7] as k^{-2}, i.e.

$$P_{DTV}(k) \propto \frac{(I_0^{\text{app}})^2}{k^2} \tag{4.10}$$

indicating that, to reasonable approximation,

$$\Delta^2_{\text{DTV}}(k) \propto (I_0^{\text{app}})^2 k. \tag{4.11}$$

These theoretical figures yield two facts: (1) the SIR scales inverse-quadratically with the brightness of the RFI source in any case, and (2) a lower SIR can generally be expected at higher k.

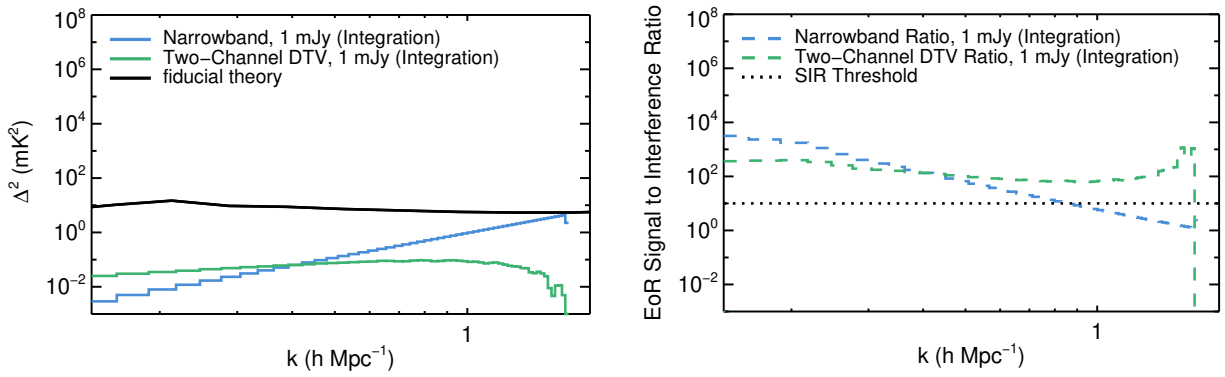
While these theoretical scaling relations are conceptually helpful, power spectrum measurements involve complicated pipelines that can alter the RFI power spectrum compared to these theoretical predictions. As shown in §4.3.2, there is not a strong discrepancy between the theoretical RFI power spectra and their simulation counterparts. However, for the purposes of drawing an accurate budget, it is better to compare the expected EoR signal to injected RFI signals processed by a 21-cm power spectrum pipeline. To this end, Figure 4.4(a) shows simulated power spectra for two different 1 mJy RFI sources alongside the

[7]This is nearly exact in the theoretical unwindowed case, where a suitable approximation using the sinc function and its derivative describes the Fourier transform of the DTV brightness, implying the leading order term in the power spectrum goes as k^{-2}.

fiducial EoR model in this work. The simulation pipeline emulates the analysis used in [10], so we show similar modes as were used in that work. Figure 4.4(b) shows the SIR for these two different cases as a function of k, along with a flat line at a hypothetical SIR threshold of 10, exemplifying a hypothetical maximum error tolerance of 10%. Due to the different scaling of the RFI power spectra, one will arrive at different budgets depending on which types of RFI sources contaminate the measurement. For example, assuming errors greater than 10% of the fiducial EoR model are unacceptable, a 1 mJy narrowband source would be acceptable for modes $k \lesssim 0.9\ h\ \mathrm{Mpc}^{-1}$, and a two-channel DTV source would pose no problems for the modes shown. On the other hand, if only 1% errors or less are considered acceptable, a narrowband source would exclude $k \lesssim 0.5\ h\ \mathrm{Mpc}^{-1}$, and a 1 mJy two-channel DTV source would nearly saturate this error tolerance on all modes shown.

Including Integration and Ensemble Effects

An observed RFI source may not be present in every snapshot used within a power spectrum integration. Since snapshots are averaged together in an integration, the apparent flux of a source may be diluted in the full integration relative to the original contaminated snapshots. Additionally, even if a source appears in a very consistent location relative to the telescope, over enough time this will not be a consistent location in celestial coordinates. In this way an RFI source may be smeared over the sky and fail to perfectly cohere with itself. This means that the per-snapshot fluxes of RFI sources can in many cases be higher than the mJy level and yet still resemble Figure 4.4 as long as there is sufficient dilution. For a highly consistent source, the dilution factor is equal to the fraction of contaminated snapshots in the integration. Any realistic source is unlikely to be consistently located in celestial coordinates over the course of a single season, but may reappear at old locations if multiple seasons are combined. Sources that appear in different locations due to seasonal variation or some other reason may instead be considered as many individual sources each diluted by a factor equal to the number of snapshots, which can then be analysed under the dynamics of a source ensemble. Summarizing, when integrating many snapshots together, one may dilute



(a) (b)

Figure 4.4: (a) 1 mJy narrowband (blue) and two-channel DTV (green) RFI power spectra, along with the fiducial EoR model from [10] (black). We can see that even at an integrated flux density of 1 mJy, a narrowband RFI source can compete with the EoR signal at higher k-modes. The DTV source has flatter scaling as a function of k, and so is substantially lower then the EoR model on all modes shown. (b) The EoR signal to RFI power ratios (dashed) for the spectra shown in (a), along with a level at 10, indicating the minimum allowable signal-to-interference ratio (dotted) for a hypothetical budget. The SIR for the narrowband source is below the threshold for modes $k \gtrsim 0.9\ h\ \mathrm{Mpc}^{-1}$. This means that modes beyond this value are excluded by a 1 mJy narrowband source in a budget drawn from this EoR model. In other words, a budget of a single narrowband source of integration flux density equal to 1 mJy is sufficient if this EoR model is accurate, 10% errors are considered acceptable, and desired measurements are on modes $k \lesssim 0.9\ h\ \mathrm{Mpc}^{-1}$. On the other hand, the SIR for the DTV source is greater than 100 for most modes shown, indicating that a budget of a single 1 mJy two-channel DTV source in the final power spectrum integration is sufficient for the modes shown if only 1% errors are acceptable and the EoR model is accurate. Alternatively, a 3 mJy DTV source will be acceptable if the error budget is 10%, but this error budget will be closer to saturation on more modes than in the narrowband case. We relate the quantities shown in this figure to what can be expected from ensembles of RFI sources and sources that appear intermittently in §4.3.3, highlighting the connection between snapshot and integration flux density of sources.

the apparent flux by a factor, χ, that depends on the consistency of the source relative to the observing strategy:

$$I_0^{\text{app, integration}} = \chi I_0^{\text{app, snapshot}} \tag{4.12}$$

Another important consideration when developing an RFI budget is the nature of ensembles. Given the occupancy study in [102], it is relatively likely that any given snapshot contains some extremely faint RFI of some variety. Therefore, a power spectrum integration consisting of many averaged snapshots probably contains an ensemble of RFI emitters. The results in Appendix C allow us to quantitatively relate the single-source and ensemble power spectra. We repeat some important results here.

First, for modes measurable by a physical instrument, the expected power of a random ensemble narrow RFI point-sources uniformly distributed over the sky is the incoherent sum of their individual powers. This can be phrased in terms of the average squared flux of the sources:

$$\langle P_{\text{ensemble}} \rangle \propto N \langle (I_s^{\text{app}})^2 \rangle \tag{4.13}$$

N is the number of emitters in the ensemble. Without knowing the exact apparent flux distribution and number of emitters, we cannot exactly know the average allowable apparent flux for an incoherent ensemble. However, a many-source ensemble is likely to probe this expectation to within some statistical fluctuation, and so we can infer possible allowances from the above relation. On the other hand, a very particular distribution of RFI emitters can add power coherently for particular modes in the power spectrum. Perfectly coherent and constructive interference is summarized with the relation

$$P_{\text{ensemble}} \propto (\sum_{s=1}^{N} I_s^{\text{app}})^2 = (I_{\text{total}}^{\text{app}})^2. \tag{4.14}$$

We note that this is equivalent to the power of a single source whose apparent flux is equal to the total apparent flux of the ensemble. Of course, for certain modes and combinations of sources, destructive interference is possible as well. Combining this relation with Figure 4.4, we see that if the total RFI flux density after dilution can be kept at roughly the mJy level or

lower, the EoR model is accurate, and there are not other systematic effects, EoR detection will certainly be feasible even in the worst case scenario of total constructive interference.

With these effects in mind, we develop RFI budgets for some hypothetical ensembles, given the fiducial model shown in Figure 4.4. We design the budgets so that they resemble what is shown in Figure 4.4. Since RFI flagging is commonly performed on a per-snapshot basis, a per-snapshot budget may be more readily applicable than a full integration budget. To this end, we describe RFI budgets for a given ensemble in terms of an allowed snapshot flux density per source and an allowed integration flux density per source. The example RFI budgets and parameters that determine them are summarized in Table 4.1.

As a simple example, consider a single source that appears in a consistent celestial location in 10 snapshots out of a 10000-snapshot integration (0.1% of the data). Since the source appears in a consistent celestial location in every snapshot, there is a single source in the final integrated image. To yield the same budget as indicated illustrated in Figure 4.4, we demand that the SIR be the same as the single mJy RFI source. This reduces to equation 4.12, with $\chi = 0.001$. Therefore, its power spectrum will resemble Figure 4.4 if its per-snapshot flux density is 1 Jy. Alternatively, suppose this source appears in uniformly distributed locations on the sky at the same apparent brightness for the sake of simplicity, but occupies ten times as many snapshots. The final integrated image then contains 100 incoherent copies of the RFI source, each diluted by the number of snapshots ($N = 100$, $\chi = 0.0001$). Invoking equation 4.13, which does not need ensemble brackets when the source fluxes are all identical, we arrive at the relation

$$P_{\text{ensemble}} \propto N \frac{(I_0^{\text{app, snapshot}})^2}{N_{\text{snapshot}}^2}. \tag{4.15}$$

This works out to be exactly the same power as a single 1 mJy RFI source, thus yielding the same SIR. In this alternative example, the RFI source is afforded more occupancy within the budget since its spatial inconsistency makes it less problematic in the final integration. Summarizing, a single bright, consistently missed RFI source can be sustainable so long as it is sufficiently diluted by the observing strategy.

Now consider the data set in the [10] limit. Out of the original 1029 snapshots proposed

Table 4.1: EoR RFI budgets under different circumstances. For narrowband sources, these give 10% or less fractional excess power relative to the fiducial EoR model for $k \lesssim 0.9\ h$ Mpc^{-1}. For two-channel DTV sources, these give 1% or less fractional excess power on modes considered in this work. All occupancies and flux densities are expressed per individual RFI source. The integration flux density refers to the flux density of a source in the final integrated spectrum, while the snapshot flux density reflects how bright the source was in its original snapshot. Note that for the final row, 100 incoherent copies of the source appear in the final integration.

Number of Sources	Number of Snapshots	Coherence	Occupancy	Snapshot Flux Density	Integration Flux Density
311	1029	Coherent	0.1%	3.3 mJy	3.2 μJy
311	1029	Incoherent	0.1%	58 mJy	57 μJy
1	10000	Coherent	0.1%	1 Jy	1 mJy
1	10000	Incoherent	1%	1 Jy	100 μJy

for the integration, 311 snapshots were found to contain residual DTV interference. Assuming one source per contaminated snapshot, we analyse two example realizations of this ensemble below.

For the first example, let us assume this ensemble sums power incoherently. Following similar logic as in the preceding paragraphs (applying equation 4.15 and demanding an identical SIR), we find that if each emitter in this ensemble has a snapshot flux density as low as 58 mJy, corresponding to an integration flux density of 57 μJy, the ensemble SIR will resemble Figure 4.4(b). We note that despite the ensemble power spectrum resembling that of a single 1 mJy source, the total RFI integration flux density in this incoherent ensemble is substantially higher at 17.6 mJy. This demonstrates how incoherent ensembles can allow more total integrated RFI flux density for the same SIR.

Alternatively, consider a situation where sources appear in consistent locations on the sky, such as reflections from regularly scheduled aircraft flying due North-South as in [102]. In this particular example, the sources will exhibit coherence for East-West modes. If East-West baselines are highly favored in the analysis, then there will be strong constructive interference in the power spectrum. In the worst case scenario of perfect constructive interference, the total integration flux density budget of 1 mJy can be split over the 311 coherent sources, as in equation 4.14.[8] This ensemble of sources will resemble Figure 4.4(b) if the snapshot flux densities of each source are as low as 3.3 mJy, corresponding to an integration flux density of 3.2 μJy per source.[9] In these examples, the total number of contaminated snapshots is quite high, and so the per-source allowable RFI brightness is quite low. Given the relative remoteness of the MWA, we expect comparable overall occupancy or worse in most radio telescopes. This low tolerance emphasizes the need to effectively filter RFI and mitigate its effects within EoR data sets.

The exact dilution factor is strongly dependent on the RFI environment, length of in-

[8]This is mathematically equivalent to considering it as a single source with dilution equal to 311/1029, similar to the very first example ensemble in the preceding paragraphs.

[9]Given that the effect of RFI removal was noticeable at the 10^4 mK^2 level, the brightest sources found by SSINS in this data set must have been substantially brighter than these hypothetical figures.

tegration, and telescope operations (pointing schedule, etc.). Without knowing the exact source of RFI and its mechanism of arrival, there is a great deal of uncertainty in this dilution factor. Setting an accurate per-snapshot budget requires a detailed study of the RFI environment of an EoR telescope. Imaging of contaminated snapshots to understand RFI propagation mechanisms can help determine the consistency of different RFI emitters. This can help inform modifications to telescope operations and data cuts in order to prevent observation of consistent RFI in long integrations.

4.4 Conclusions

We used theoretical calculations and end-to-end simulations to investigate the level to which RFI occludes EoR detection. For simulation, we used the FHD/εPPSILON pipeline, with the MWA as our simulated instrument. There was a strong correspondence between the theoretical contamination estimate and the simulated results, verifying the conceptual framework used in this analysis. We conclude that relatively low levels of RFI contamination are sufficient to overwhelm the EoR signal in the 21-cm power spectrum.

Specifically, simulations show that a single narrowband source of 1 mJy apparent flux density in the final power spectrum integration offers excess Δ^2 that is cubic in k, with greater than 10% fractional errors relative to an optimistic EoR model for $k \gtrsim 0.9\ h\ \mathrm{Mpc}^{-1}$. The contamination of this single source scales quadratically as its flux density, and so a single narrowband source of flux density 10 mJy can begin to overwhelm the EoR signal on modes $k \gtrsim 0.3\ h\ \mathrm{Mpc}^{-1}$. This flux density reflects the brightness of the source after forming a power spectrum integration, rather than its brightness in a single snapshot. Depending on the observing strategy and consistency of the source, this may be a strong dilution relative to its snapshot flux density. However, for long power spectrum integrations, it is unlikely that only a single source will be present in the final integration, and so the average allowed apparent flux density of any one RFI source in the final integration may be substantially lower than this number depending on how many sources are actually present in the measurement set. Averaged over realizations, a random angularly uniform ensemble of narrowband sources

is expected to add power incoherently. Furthermore, certain ensembles may add power coherently for certain modes. The power spectrum of a single source of a given flux density, e.g. 1 mJy, represents the maximum power spectrum of a coherent ensemble whose total flux density is the given flux density of that single source. Ultimately, this means that if the assumed optimistic EoR model used here is accurate, a total RFI apparent flux density budget of $\lesssim 1$ mJy in the final power spectrum integration will be sufficient for attempting to measure modes $k \lesssim 0.9\ h\ \mathrm{Mpc}^{-1}$ assuming a 1 - 10% error budget depending on the type of RFI that dominates the ensemble.

This work assumes a foreground avoidance strategy, and so does not focus on lower-order modes where the EoR signal may be stronger. If foregrounds can be successfully mitigated on these modes, it is possible that a less stringent RFI budget can be deduced. On the other hand, the EoR model used is relatively optimistic, and a more pessimistic EoR model will produce a stricter budget.

The size of contamination relative to the faintness of the offending RFI source helps explain the noticeable general improvement of the 21-cm power spectra made by removing RFI-contaminated observations in [10]. Furthermore, we expect the results of this analysis to be fairly generic between experiments, particularly predictions for narrowband RFI contamination and any contamination levels within the EoR window. This strongly motivates quantifying the effectiveness of RFI excision implementations, increasing their sensitivity if necessary, and even possibly implementing post-excision mitigation strategies.

Chapter 5

CHARACTERIZING EXCESS POWER FROM FLAGGING IN 21-CM POWER SPECTRA

In this chapter, we explore how RFI flags affect power spectrum measurements. We find that RFI flags produce excess power in the EoR window in much the same way as residual RFI. Without modifying current flagging strategies or implementing extremely accurate foreground subtraction, 21-cm EoR experiments will fail to make a significant detection.

5.1 Introduction

[65] explores some effects of excision methods in 21-cm EoR power spectrum analyses. This study is performed on LOFAR data, where RFI flagging is done at much finer spectral resolution than is used in the EoR analysis. The flagged visibilities are then downsampled to the analysis time and frequency resolution. Conventionally, flagged data are not included when downsampling, except when all contributing visibilities are flagged at high resolution. If there are some unflagged contributors, the resulting downsampled visibility is left unflagged at low resolution (referred to as partially flagged). Otherwise, the resulting visibility is flagged (referred to as fully flagged). Due to chromatic variations in flagging, conventional downsampling causes spectral fluctuations in the averaged visibilities that produces enough excess power in the EoR window to prevent detection of the expected EoR signal. This excess power can be effectively mitigated by improved downsampling schemes, weighting the downsampled visibilities identically regardless of how many samples contributed, subtraction of a low-resolution forward model, and Gaussian Process Regression (GPR: [39, 51]). However, many of these techniques only address the excess power resulting from partially flagged visibilities. In this chapter we primarily focus on fully flagged visibilities, which also produce

spectral variations that result in excess EoR window power. Fully flagged and nearly fully flagged data is more common in the MWA EoR highband from 167-198 MHz than in the LOFAR band considered in [65]. This is due to the presence of broad DTV interference in the MWA band as well as a lower correlator resolution. We propose an alternate mitigation strategy that applies to fully flagged visibilities where the strategies in [65] do not.

In [33], the foreground wedge is understood in terms of the chromatic sampling function of the array, specifically by examining multi-baseline effects. The excess power in the window from fully flagged samples can be understood similarly. In §5.2, we describe the theoretical concepts and methods used in this paper. In §5.3 we use simulations to demonstrate and explain this effect, as well as provide a solution. In §5.4, we further demonstrate the effect using flags inherited from running SSINS and AOFlagger on MWA data. In §5.5, we discuss mitigation strategies and consequences for overall experimental sensitivity.

5.2 Theoretical Principles and Methods

The essential theoretical principle that explains excess power from fully flagged visibilities is that they cause sharp spectral variations in the sampling function of the array. This causes power from spectrally smooth foregrounds, which would otherwise be contained in the wedge, to leak up into the EoR window. The shape of the leaked power is directly related to the shape of the spectral variations via a Fourier transform. The overall amplitude is proportional to the total foreground power. Since the foregrounds are tremendously bright compared to the expected EoR signal, even small variations caused by a minute fraction of fully flagged samples cause enough power in the EoR window to prevent detection.

Due to the rotation of the Earth, the uv-modes sampled by a terrestrially bound interferometer depend not only on the array layout and the observing frequencies, but also on the sidereal times present in the data. For instance, if the phase center is held constant, a projected baseline center for a given pair of antennas and frequency tracks an ellipse in the uv-plane. This effect is typically utilized to sample more modes than would otherwise be included by a given antenna pair in a technique called rotation synthesis. If LST sampling is

even, the average location of a given baseline will be halfway along the arc of the ellipse that is swept out during the range of LSTs. If LST sampling is uneven, then the average location of the baseline shifts away from this center. Extending this logic, if the set of sampled LSTs for one frequency is different than that of another frequency, the average locations of the baseline centers will possess spectral variations that are potentially very sharp. We estimate the size of the baseline discrepancy for a particular example in §5.3. This is the primary effect on the sampling function that causes the excess power in the EoR window. This means that any analysis that coherently averages uv-planes from multiple LSTs can have excess power as a result of flagging, regardless of the choice of power spectrum approach (e.g. delay vs. gridding).

The spectral shape of the sampling function variation is directly influenced by the shape of the flags. This shape subsequently determines the form of the power spectrum contamination. There are five basic flagging shapes in the MWA EoR highband, resulting from different sources:

1. Narrowband (single-frequency transmitters)

2. Broad, band-limited (DTV)

3. Frequency Comb (routine coarse band edge flagging)

4. Uniform Random (false positives from the flagging algorithm)

5. Broad, not band-limited (various, e.g. extremely bright ORBCOMM spillover)

The spectral variation introduced by any given flagging shape is approximately identical to the spectral variations introduced by leaving an unflagged RFI source that occupies the same frequencies as those that are flagged, but with a strength given by the foreground brightness and flagging occupancy. Therefore, we can understand the form of the excess window power in terms of the analysis presented in chapter 4. Narrowband flags produce constant power

in the line-of-sight modes, while DTV flags produce lobed power-law contamination. A frequency comb of flags produces a line-of-sight comb in the power spectrum, while uniformly random flags produce uniformly random contamination. Finally, totally achromatic flags, equivalent to not using contaminated integrations, produces no excess power in the window.

Flags tend to be irregular in terms of sidereal times since most RFI is not locked to a sidereal schedule. Consequently, various modes experience chromatic disruptions in sampling, even if the physical baselines that are flagged is consistent. To demonstrate the resulting excess power, we simulate MWA visibilities with various flags applied in a number of circumstances and calculate the resulting power spectra using the FHD/εPPSILON pipeline. While this is a gridded power spectrum pipeline, we expect that the effect will still appear in delay spectrum pipelines as well. Unless a given baseline is flagged perfectly consistently as a function of frequency and LST, that baseline will sample its rotation ellipse irregularly as a function of frequency, and this will throw power into high delay modes.

5.3 Conceptual Demonstration

In this section, we simulate two minutes of MWA visibilities from two different source catalogs and various flags using FHD. First, we examine the gridded, flagged visibilities of a single source displaced from phase center, which demonstrate the sharp spectral fluctuations in the reconstructed uv-plane that create the excess power. We then work through several simulations of a GLEAM subset using the flagging shapes listed in §5.2 to give a lower bound on the strength of this effect in measured power spectra. The simulations are centered on a right ascension of 0 and declination -27 degrees. For all power spectra shown in this work, we use the entire 30.72 MHz instantaneous bandwidth of the simulated visibilities. In practice, only a fraction of this bandwidth is used to as to ensure that only nearby epochs are probed, i.e. so that the measurement does not span too wide a redshift. We find the power spectrum features arising from chromatic flags are more clearly resolved using a wider bandwidth, and so we use the full band for the sake of demonstration.

5.3.1 A Single Flagged Source

We begin by examining the uv-plane resulting from a single point-source displaced from phase center under different flagging conditions. We consider three quantities:

1. The analytic uv-plane.

2. The reconstructed uv-plane using FHD, with no flags.

3. The reconstructed uv-plane using FHD as above, but with flags from 181-188 MHz that simulate digital television excision. The extra flags are on on all baselines for half the integrations in the simulation.

In Figure 5.1, we examine the difference between the analytic phase and the two reconstructed phases enumerated above at a single uv-point. The topmost panel of Figure 5.1 shows the difference between the analytic phase and the reconstructed phase with no flags applied, as well as the difference between the analytic phase and the reconstructed phase using DTV flags. Over the range of frequencies in the band, both residuals appear smoothly varying and nearly identical. The large-scale, smooth, but nonzero variation in the reconstructed phase error is what gives rise to the foreground wedge, via the mechanism explained in [33]. The bottom panel shows the difference between the two reconstructed phases. There is only disagreement for frequencies with flags, and the fractional discrepancy in the phase is $\sim 10^{-3}$. This small discontinuity accounts for power outside the wedge in the EoR window.

To show how this affects the power spectrum, we evaluate the modulus square of a windowed spectral Fourier transform of the reconstructed uvf-cubes at the uv-point depicted in Figure 5.1, using a Blackman-Harris spectral tapering function. We also do this with their difference to form a residual power spectrum. We show the results in Figure 5.2. We find that the flagged reconstruction has long tails out to high-order delay line-of-sight modes, while the unflagged reconstruction has favorable falloff. The tails of the residual power spectrum match the tails of the flagged power spectrum. The rate of falloff for the flagged data is

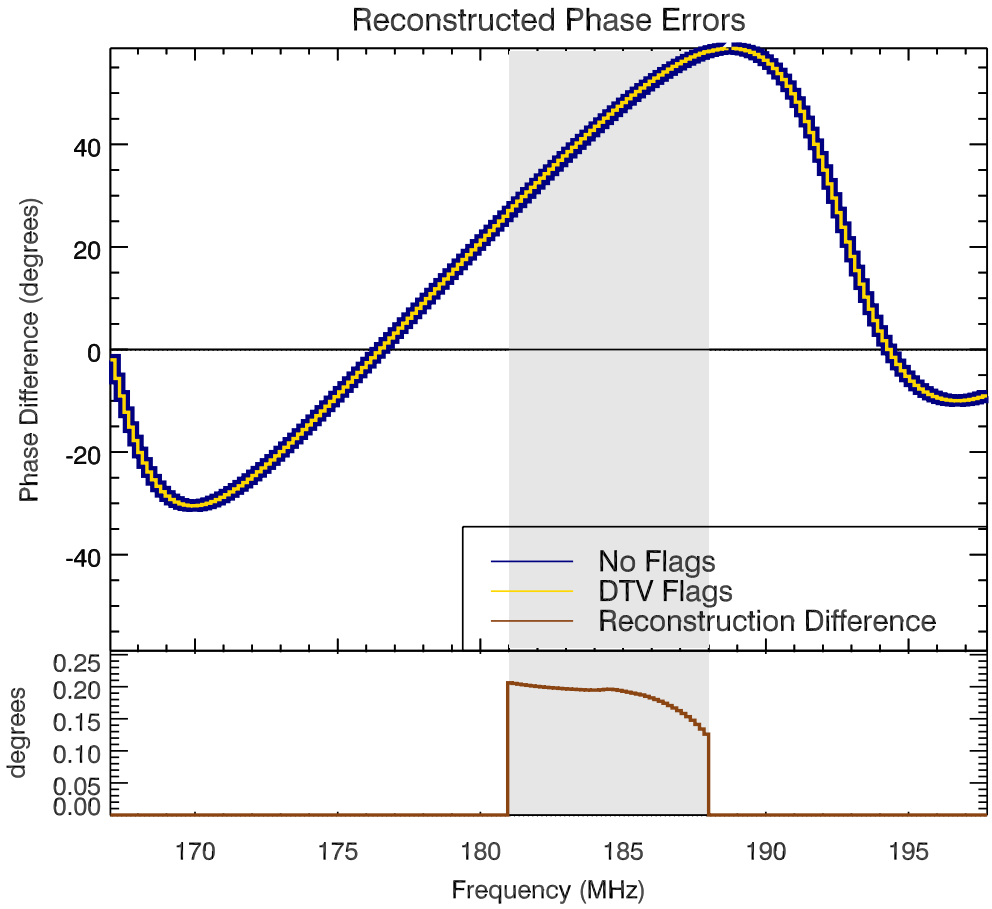


Figure 5.1: Top: Shows difference between analytic phase of a source at a given uv-point and the reconstructed phase using FHD with no flags (dark blue), and with DTV flags (gold). The smooth spectral variation in the phase reconstruction error is what gives rise to the foreground wedge. Bottom: Phase difference between the reconstructed phase with minimal flagging and with DTV flags. There is only disagreement where flags between the two reconstructions disagree, and the disagreement appears discontinuously at the boundaries of the flagged region. This sharp spectral structure gives rise to power outside the wedge, in the EoR window, cf. Figure 5.2.

exactly what would be expected for a sharp discontinuity, similar to the RFI power spectra in Chapter 4.

The size of the discontinuity is consistent with the relative baseline center offset between flagged and unflagged data. The discrepancy in baseline position can be understood by examining the trajectory of a baseline center as a function of time. The arcspeed of a baseline along its elliptical track is fastest when it its projected displacement vector is aligned with $u = 0$. For intervals of fixed length and a given baseline, centering observations on this LST produces the greatest discrepancies between flagged data and unflagged data. For such baselines and times, the arc length, Δs, swept out during a short time interval is approximately

$$\Delta s \approx X \Delta \tau, \tag{5.1}$$

to first order in $\Delta\tau$, where X is the East-West length of the baseline in question when projected from zenith at the given frequency, and $\Delta\tau$ is the length of the LST interval, measured in radians. For the unflagged data in the simulation, the average baseline locations are at the center of their respective arcs. For the flagged data, the average baseline centers are located away from the center of this arc. Since the first half of the two-minute interval is flagged, the average location of each baseline is approximately three-quarters of the way along the arc. This discrepancy is

$$\Delta s_{\max} \approx \frac{\pi}{1440} X \sim (2 \cdot 10^{-3}) X \tag{5.2}$$

A given baseline's arcspeed will be slowest when it projects along $v = 0$. The factor by which it is slower than its maximum speed is given by $\sin \delta_0$, where δ_0 is the declination of phase center. For the MWA EoR0 field, centered on a right ascension of 0 and declination of -27 degrees, this is about a factor of 2. Figure 5.1 shows that this discrepancy produces a phase difference between the unflagged and flagged, gridded visibilities on the order of 0.2 degrees that is spectrally discontinuous at the borders of the flagged regions. We note that the relative size of this effect matches the discrepancy in power between the unflagged and residual (difference between flagged reconstruction and unflagged) power spectra in the

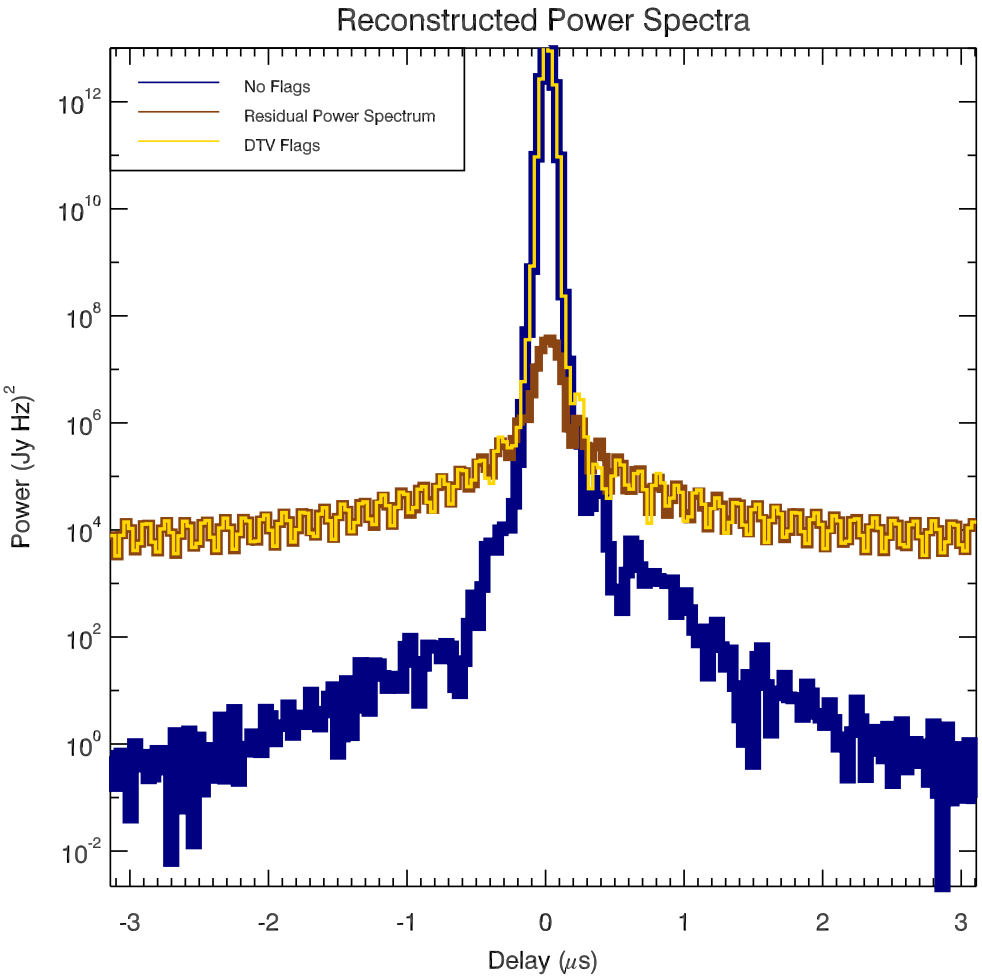


Figure 5.2: Power spectra of the two reconstructed uv-planes, evaluated at a single uv-point, along with the power spectrum of the residual. The power spectrum with no flags falls off at a similar rate as a Blackman-Harris window function's Fourier transform, as would be expected for a smooth-spectrum source. Adding flags gives a similar amount of power in the zero-delay mode, but with much slower falloff. The power spectrum of the residual is low in the zero-delay mode since the discontinuity is small, however it has long tails that match the tails of the flagged reconstruction.

zero-delay mode; the discrepancy in zero-delay power is fractionally $3.39 \cdot 10^{-6}$, which is slightly smaller than the square of the size of the fractional baseline displacement for that particular uv-point based on equation 5.2.

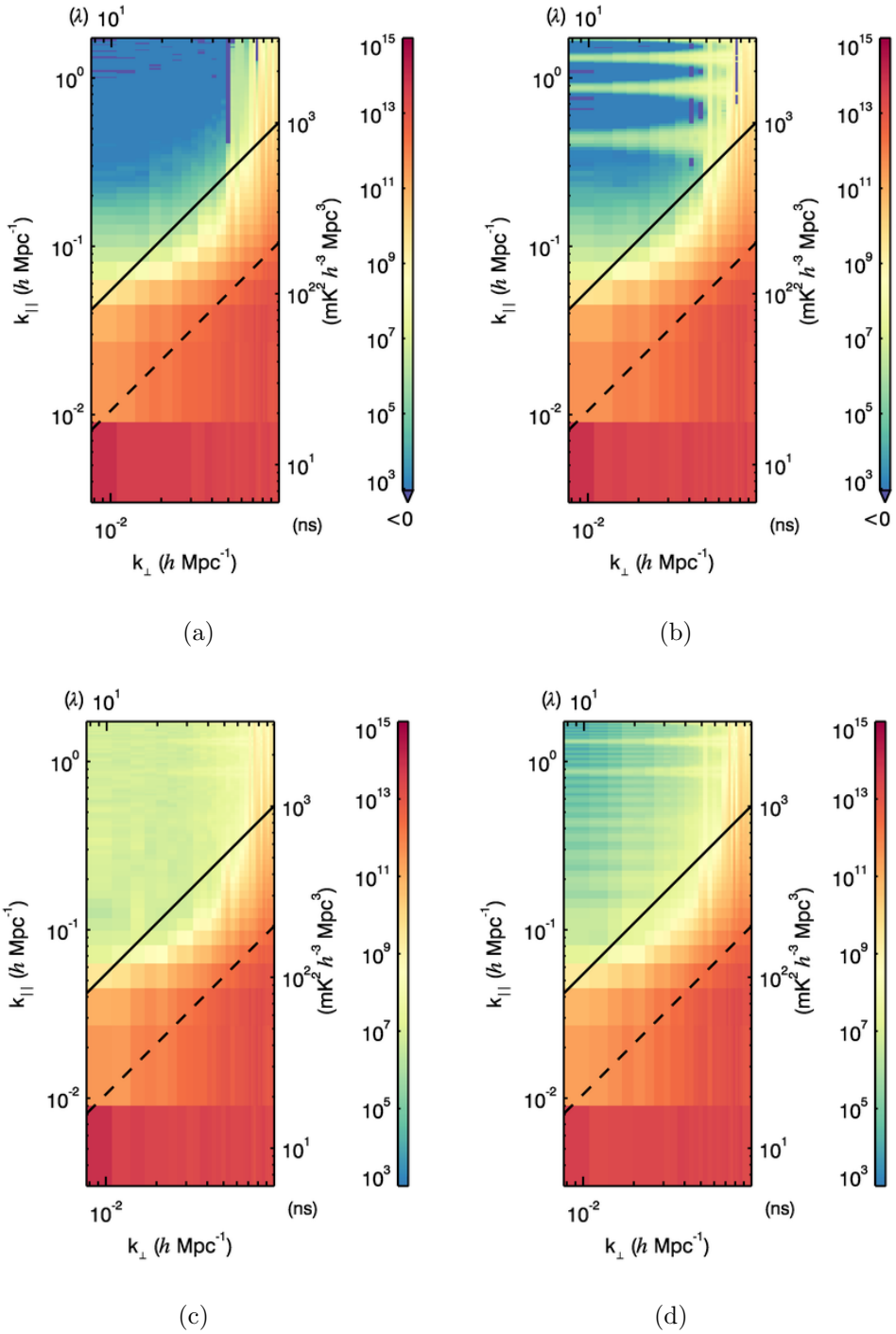
A discontinuity in the frequency domain results in power law contamination in the Fourier domain that is proportional to the size of the discontinuity. In Chapter 4 this was demonstrated by RFI sources introducing discontinuities. Here, it is the flags that produce the discontinuity, but the overall effect is the same. In practice, the sky signal is dominated by the foregrounds, and the size of the discontinuity resulting from the aberrant sampling function is proportional to their brightness. Since the foreground signal is orders of magnitude brighter than the EoR signal, the resulting contamination in the EoR window is significant. In the next section, we simulate visibilities and power spectra for GLEAM sources to get a rough estimate of the strength of this effect in a measured power spectrum. Since no source catalog is complete, and we do not model diffuse emission, the strength of this effect in simulated power spectra serves as a lower bound for the strength in measured power spectra.

5.3.2 Flagged GLEAM

In Figure 5.3.2, we show simulated power spectra of GLEAM sources with different flags applied. There are four chosen flagging patterns: no flags, coarse band edge flags only, DTV flags overtop coarse band edge flags, and random flags overtop coarse band edge flags. Each flagging pattern produces a characteristic shape in the EoR window. The purpose of the latter two patterns is to simulate what is found in MWA data. The simulated data are flagged and gridded at 80 kHz. Then HEALPix cubes are formed and averaged to 160 kHz for passage to εPPSILON. This typically results in a mixture of fully-flagged and partially-flagged visibilities.

The case without flags sets the ground truth for the GLEAM power spectrum with our current analysis techniques. The window is free of foreground contamination, with a floor given by the Fourier transform of the various tapering functions applied to the data [10].

Coarse band edge flags are periodic in frequency. We flag 80 kHz around each coarse



(a) (b)

(c) (d)

Figure 5.3: Simulated power spectra using GLEAM and different sets of flags: (a) no flags, (b) coarse band edge flags only, (c) random flags in addition to coarse band edge flags, and (d) 7 MHz corresponding to DTV channel 7 flagged in addition to coarse band edge flags.

channel edge, which means that the resulting 160 kHz coarse band edges of the HEALPix cubes are all half-flagged. This produces a harmonic structure in the Fourier domain along the line-of-sight modes. All power spectra with coarse band edge flags will have this harmonic structure present in them, and it is several orders of magnitude brighter than the expected EoR signal strength. Each harmonic has a non-negligible width in the line-of-sight modes, and affects all baselines, making a wide range of modes in the EoR window totally inaccessible.

The bottom-left panel of Figure 5.3.2 shows the results of applying random flags to the GLEAM visibilities at a rate of approximately 0.2%, in addition to coarse band edge flags. This simulates the effect of a small false positive ratio when flagging at the data resolution. AOFlagger is applied to MWA data at the finest resolution possible in order to have the greatest detection accuracy. The actual false positive rate at the finest resolution is higher than this rate. This rate reflects the number of fully flagged samples that typically remain after averaging from 40 kHz to 80 kHz, and ignores effects from partially flagged samples. These flags are distributed uniformly across all times, baselines, and non-edge frequencies. We see that for a two minutes simulation, this causes excess power with a noisy shape that is many orders of magnitude above the expected EoR signal.

The bottom-right panel of Figure 5.3.2 shows the results of flagging all baselines for the first half of the observation across all baselines for the frequencies 181-188 MHz. This pattern might be expected from running SSINS on an observation with DTV contamination resulting form aircraft reflection, as in Chapter 3. This produces a lobed structure as a function of $k_\parallel$, and in fact the size of the lobes corresponds to the inverse-width of the flagged region, as would be expected from basic Fourier transform reasoning and the calculations shown in Chapter 4.[1] This contamination goes approximately as $k_\parallel^{-2}$. Extrapolating this back to the lowest order $k_\parallel$ mode, we see that this is roughly 5-6 orders of magnitude beneath

[1]The lobes in this power spectrum appear slightly different than those in the simulated DTV power spectrum in Figure 4.2. This is a result of having chosen a different analysis bandwidth, which affects the width of the $k_\parallel$ bins for which the spectrum is calculated.

the GLEAM power, which corresponds to the excess power amplitude being quadratic in the sampling function discontinuity as in the single-source case. The contamination is also several orders of magnitude above the expected EoR signal strength in the window. The spacing of the lobes makes measurements in the window impossible without mitigation of this effect.

5.3.3 LST Replacement

We can associate a given sampling configuration with a local sidereal time and frequency combination. If missing LST/frequency combinations can be appropriately replaced by data from another night, then the sampling function can be restored to its original smoothness. Unfortunately, we find that naively averaging flagged data with corresponding unflagged data generically results in only a minor reduction of excess window power if special care is not taken.

To investigate averaging flagged nights with unflagged ones, we simulate two copies of a two-minute MWA observation: one with DTV flags as in Figure 5.3.2(d), and another with no flags. We then coherently average the resulting HEALPix cubes to observe the effect of exact LST/frequency replacement. Relative to other LST/frequency combinations, the flagged LST/frequencyes are still undersampled after the coherent average. While this reduces the strength of the discontinuity, it does not eliminate it. In Figure 5.4(a), we show the resulting power spectrum of the coherently averaged cubes. The excess power is only slightly diminished.

Curiously, if we create a copy of the two-minute interval with exactly complementary flags, i.e. with the second half flagged for DTV, and average it with the first, this produces a weighted average whose uv-plane is smooth as a function of frequency. The resulting power spectra have no excess window power. While we recognize that this method successfully removed the excess power resulting from fully flagged visibilities, this would be a costly mitigation strategy in terms of sensitivity. Since exactly complementary flags are required, this reduces the overall data volume of an analysis by at least a factor of two.

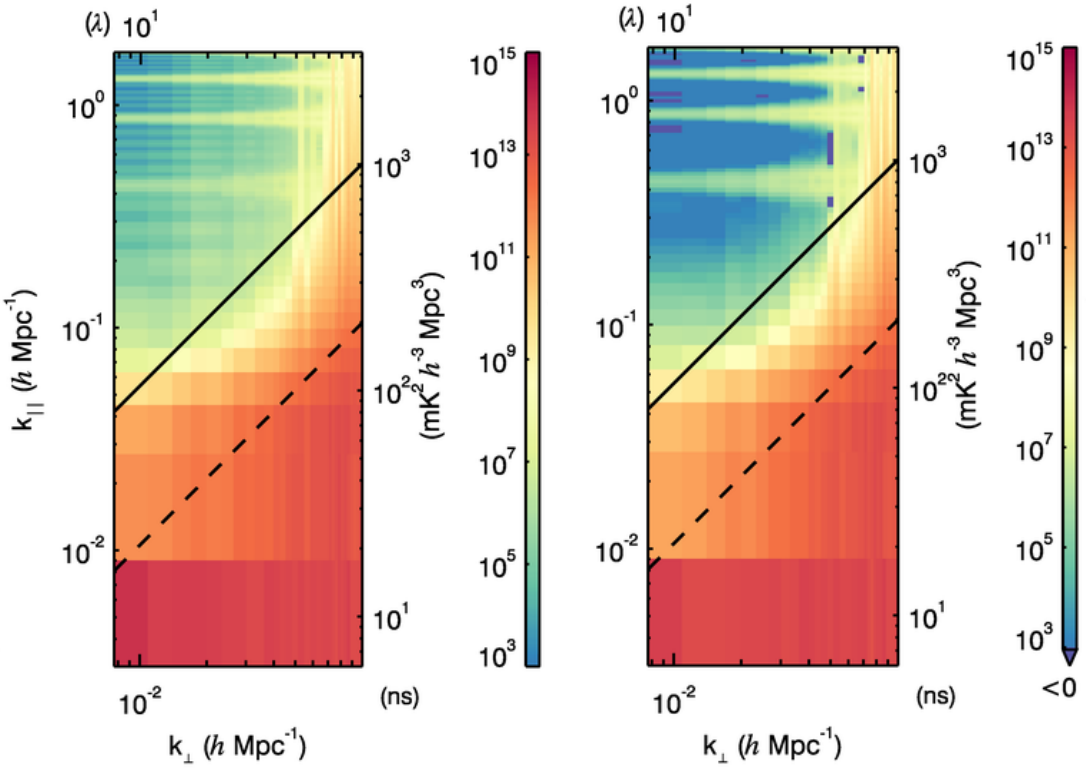


Figure 5.4: (a) Simulated power spectrum formed by flagging the first half of a two-minute interval with 7 MHz DTV flags and averaging it against the exact same observation but without any flagging. EoR window contamination levels are only mildly reduced and still too high. (b) Simulated power spectrum where the first half of the two minute interval is flagged across all frequencies, rather than just those corresponding to the DTV allocation. Coarse band edge flags are still present in the second half of the observation. No excess power in the window other than what arises from coarse band edge flags is added by this flagging strategy.

5.3.4 Achromatic Flags

We observe that if we flag all frequencies for any integration that is contaminated, then the window contamination is also removed. Since this is equivalent to simply not including any contaminated integrations, no additional chromaticity is injected into the uv-sampling function, resulting in no additional power other than what exists from the routine coarse band edge flagging. The ensuing power spectrum is shown in Figure 5.4(b). This potential mitigation strategy has an associated sensitivity cost that depends crucially on the flagging strategy. We propose this as the primary strategy for mitigating excess power from RFI flagging so long as the false positive rate of the flagger is sufficiently low. In the next section, we assess the impact of this mitigation strategy using simulations with flags inherited from measured data.

5.4 Flags Inherited from Data: Assessment and Mitigation

In this section, we examine the strength of the excess flagging power using one night of flags inherited from MWA data, as well as the impact of the mitigation strategy alluded to in the preceeding section. We also check to see if flagging power significantly reduces over the course of a night by analyzing model power spectra from in-situ simulations of GLEAM using the MWA as the dummy instrument, with flags calculated by SSINS and AOFlagger on just over two hours of measured MWA visibilities.

5.4.1 Integrating Over a Night

We show the baseline-averaged flagging patterns for the two-hour data set in Figure 5.5. The periodic flagging in time and frequency are not strictly RFI-related. The coarse channel edges and centers produce bandpass irregularities that need to be excluded from the analysis. The time boundaries of observations also tend to have irregularities due to misalignment of data coming from the correlator and occasional oddities when pointings are adjusted, among other effects. The regular 7 MHz blocks in the SSINS flags reflect the results from the

frequency-matched flagging algorithm, which specifically searched for DTV contamination. The flags from AOFlagger are mostly uniform except for noticeable DTV and narrowband events later in the night, corroborated by the SSINS flags. We show a closer view of these events in Figure 5.6. Since AOFlagger flags on a per-baseline basis and produces a small number of false positives, each baseline actually has a slight, random chromaticity relative to other baselines. Each of these types of chromatic structures, namely coarse band edge flagging, DTV, and uniform random, imprint visible structures in the EoR power spectrum according to their Fourier transform.

Individual dipole delays within an MWA tile can be adjusted so that the tile can point its beam. For EoR experiments with the MWA, this functionality is used to track a particular observing field as it rotates overhead. In Figure 5.5, we demarcate these transitions with the time axis tick marks. We show model power spectra for this night in Figures 5.7 and 5.8 on a per-pointing basis. The SSINS flags show much more variation between pointings, which is reflected in the power spectra of Figure 5.7. In contrast, the AOFlagger power spectra show little variation between pointings, and consequently there is little variation between them. In both cases, there is far more power in the EoR window than the expected EoR signal, and intervention is required.

The window power in the AOFlagger-informed power spectra appear to average down favorably. This may be due to the fact that this is a relatively benign night, and much of the RFI that does exist is probably fainter than the sensitivity of AOFlagger, i.e. the flags are dominated by false positives. In [65], AOFlagger reports more true positives, and so they see that the excess power from flagging does not average incoherently like thermal noise might be expected to. Instead, since certain frequencies are consistently occupied by certain transmitters, the power does not average down substantially. This is more like what is shown in Figure 5.7, where consistent DTV events are identified by SSINS, resulting in relatively poor average qualities in the EoR window.

AOFlagger consistently flags the frequency 194.455 MHz about 10% more often than other frequencies, fractionally. This ultimately puts a floor on how far down any averaging

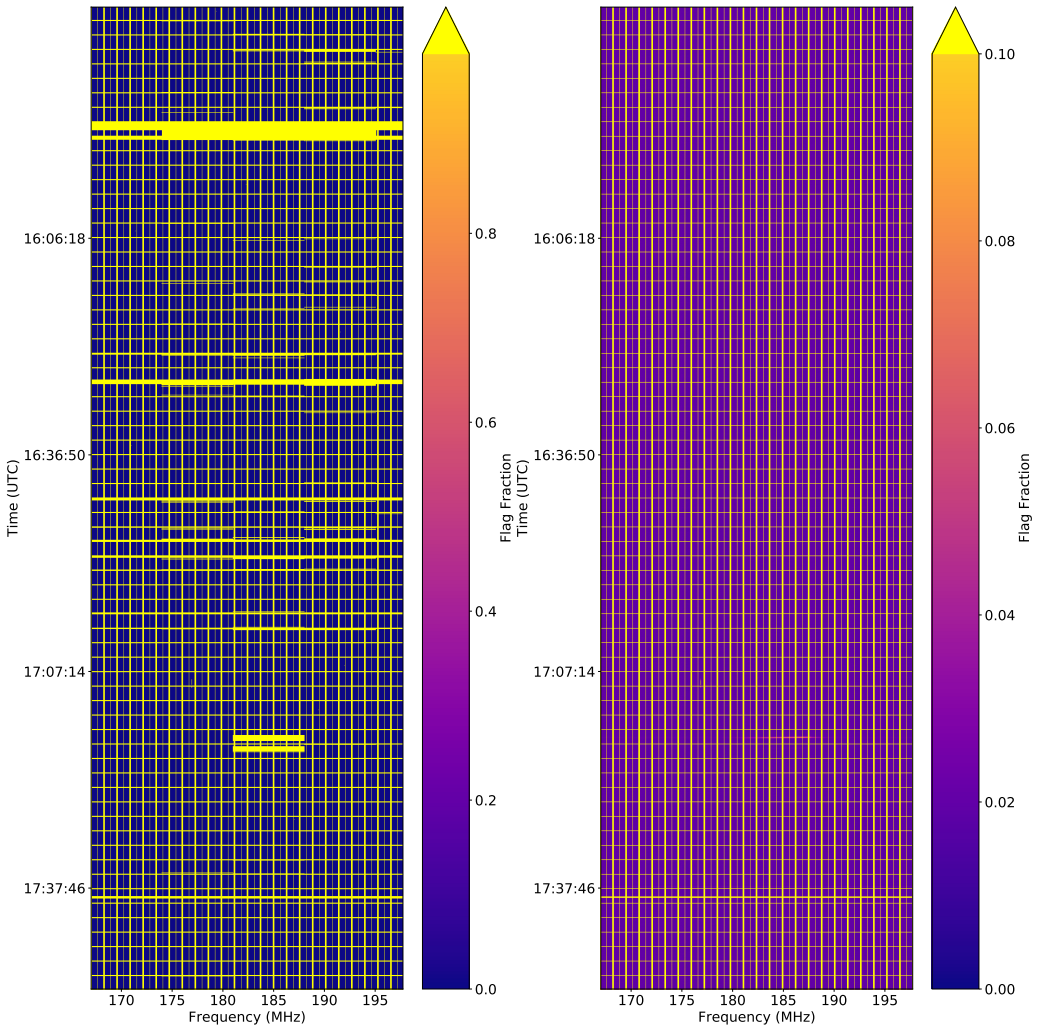


Figure 5.5: The fraction of baselines flagged using SSINS (left) and AOFlagger (right). Yellow-colored samples are flagged across all baselines. The vertical axis tick marks represent pointing boundaries. Since SSINS averages over baselines, either all baselines are flagged at a given time and frequency or none of them are. SSINS was set to specifically search for DTV, hence the somewhat regular 7 MHz blocks. AOFlagger tends to flag uniformly in the band, in part due to a small but noticeable false positive ratio, except some short DTV and narrowband events around 17:16 and 17:09 UTC, respectively.

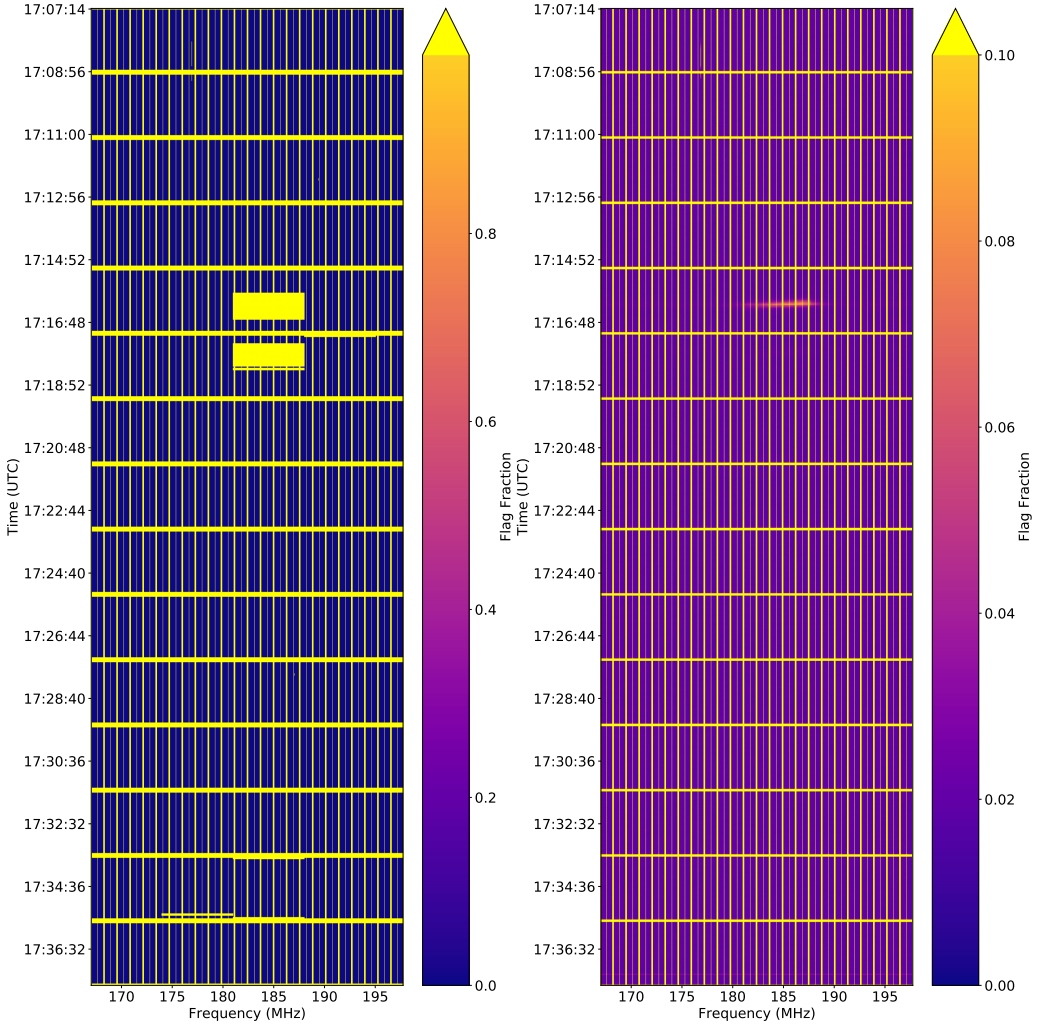


Figure 5.6: Closer view of the SSINS (left) and AOFlagger (right) flags from the fourth pointing in the night. Both SSINS and AOFlagger observe a narrowband event shortly before 17:09 UTC at about 177 MHz. There is also a DTV event around 17:16 UTC, due to an airplane reflection that both flaggers catch, though AOFlagger identifies a fewer contaminated times, and does not identify the event in the following observation. Very close to the end of the pointing, AOFlagger identifies a broadband event on a few baselines that SSINS does not identify. It is possible that this event is sufficiently diluted by the incoherent average in SSINS to become undetectable. While broadband flagging actually does not introduce additional chromaticity into the measurement since all frequencies are flagged equally, it is important to acknowledge possible false negatives in the SSINS pipeline.

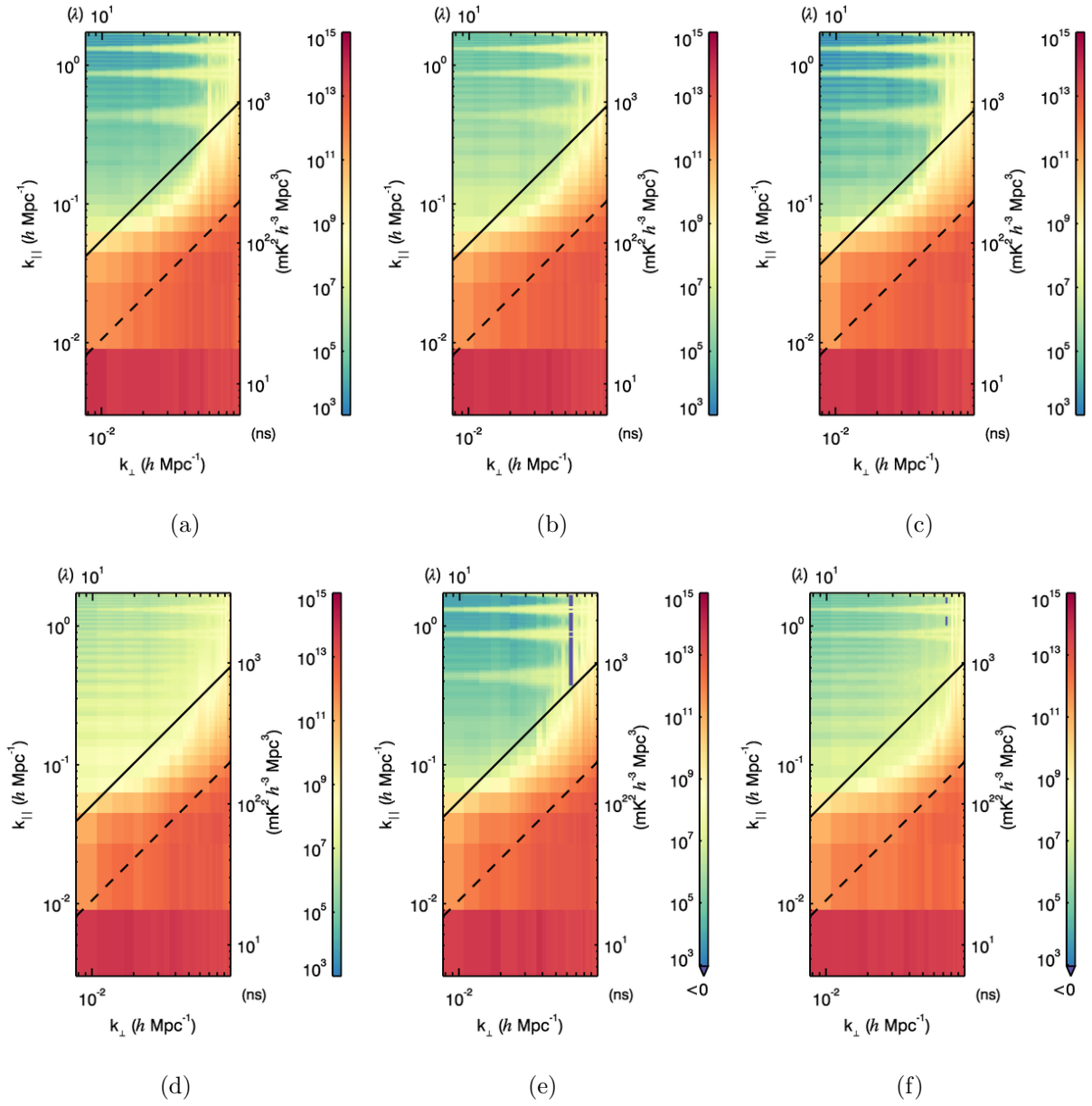
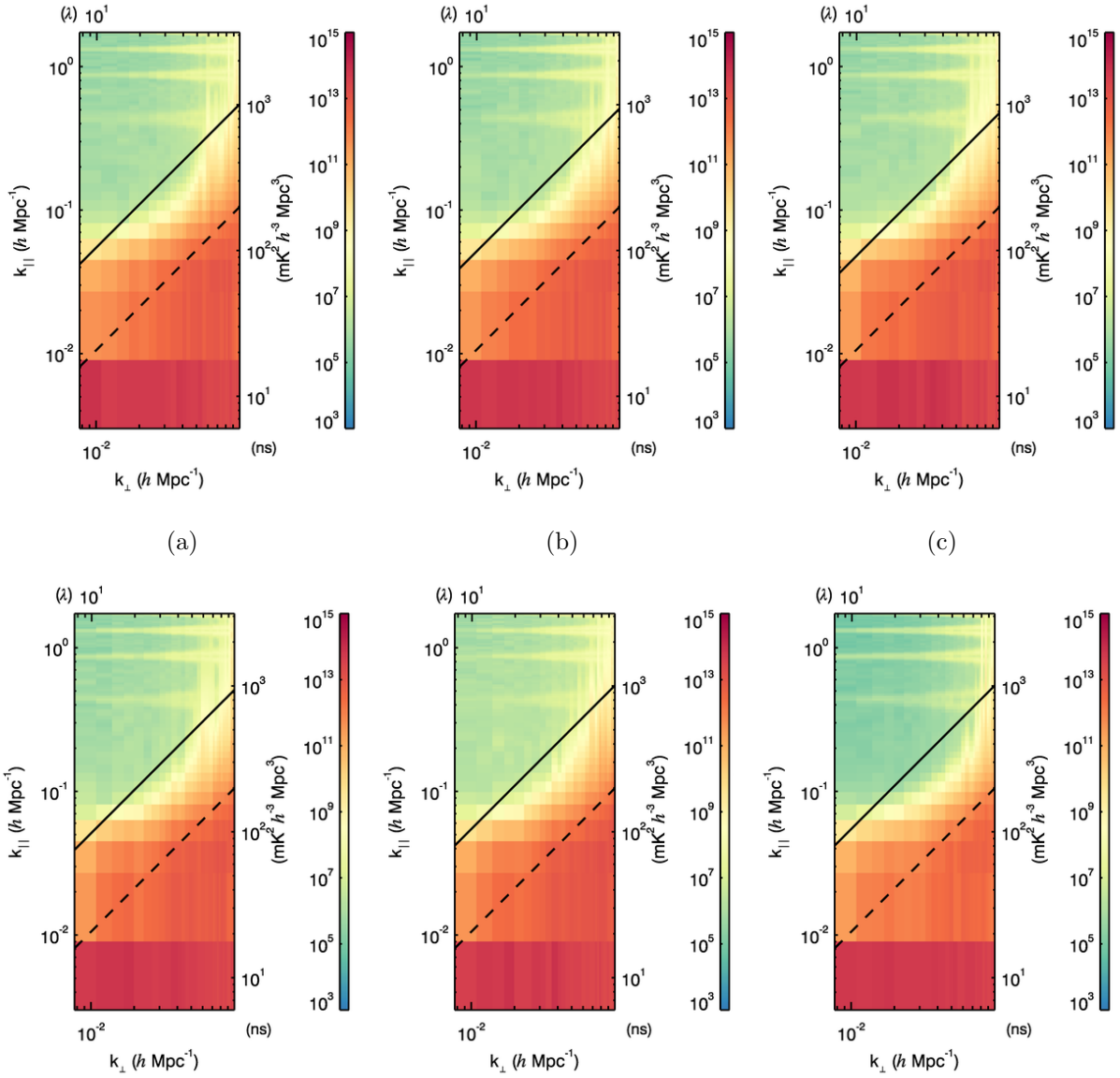


Figure 5.7: Simulated per-pointing power spectra using SSINS flags inherited from data. The five pointings are shown in chronological order left-to-right, top-to-bottom. The sixth power spectrum is the result from integrating the entire night. The fourth pointing is substantially worse than the others due to a bright DTV event that produced many DTV flags. The total integrated power spectrum contains contamination in the window that overwhelms the expected EoR signal by several orders of magnitude.



(a) (b) (c)

(d) (e) (f)

Figure 5.8: Simulated per-pointing power spectra using AOFlagger flags inherited from data. The five pointings are shown in chronological order left-to-right, top-to-bottom. The sixth power spectrum is the result from integrating the entire night. While AOFlagger did catch the DTV event in the fourth pointing on some baselines, the effect is not as discernible in these spectra as it was in the SSINS-flagged spectra. The EoR window contamination is still too high for a significant EoR detection.

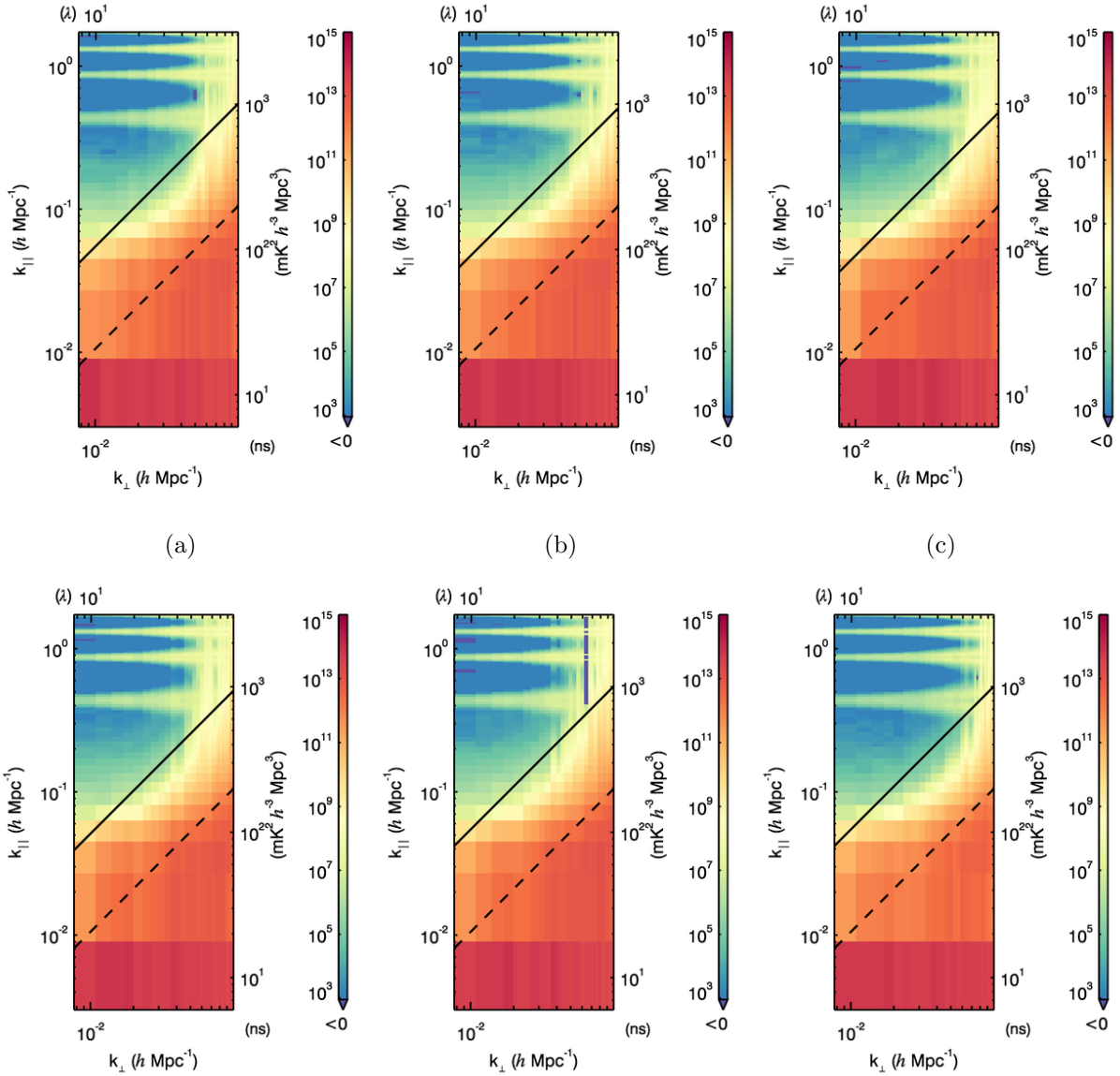
can reach, since a narrowband flagging shape will put constant excess power along the line-of-sight. The cause of this consistent flagging is so-far unknown. Since it is within a DTV allocation, it is confounding that the flags would be so narrow. It is close to the DTV guard band, and we do sometimes see AOFlagger flags around the guard band of extremely faint DTV signals observable by SSINS. However, this would suggest constant observation of a DTV signal that is not often identified by SSINS. Furthermore, the guard band flags from AOFlagger usually occupy more than one fine frequency channel, and this is localized to a particular channel. It is possibly locally generated RFI that only affects very few baselines, and thus is observed by AOFlagger but not by SSINS.

We see that with realistic flagging patterns, even relatively clean data will be too contaminated by excess flagging power for an EoR detection. For reference, the SSINS occupancy at high resolution is 4.4%, ignoring coarse band edges and the temporal boundaries of observations. Under the same conditions, the AOFlagger flags about 2% of the data at high resolution. The low resolution SSINS flags do not reduce the average occupancy by much since the vast majority of the flags are broadband. On the other hand, the AOFlagger flags are sparse, resulting in an order of magnitude decrease in the flag occupancy at low resolution; most of the flags result in partially flagged visibilities.

5.4.2 Mitigation via Flag Extension

In Figure 5.9, we show the results of applying a flag extension strategy to the SSINS flags before constructing power spectra. In this case, every integration with an RFI flag anywhere in the band is flagged entirely across the band. We see that this fully eliminates the excess power from flagging at the expense of a modest amount of data. Interestingly, the power could also be fully removed by setting the observing band equal to the occupied band. However, even comparatively broad RFI sources such as DTV are significantly narrower than typical bandwidths used in EoR power spectrum estimates, and changing the observing band to match this width would come at the cost of significant sensitivity.

The sensitivity loss for a flag extending strategy depends on the initial flags. For the



(a) (b) (c)

(d) (e) (f)

Figure 5.9: Simulated per-pointing power spectra using achromatic flags developed from the SSINS flags in §5.4. The five pointings are shown in chronological order left-to-right, top-to-bottom. The sixth power spectrum is the result from integrating the entire night. There is no excess power from these flags, with a relatively small decrease in overall data volume.

night used in §5.4, the SSINS occupancy is 4.4%, while the AOFlagger occupancy is 2%. Both of these figures ignore routine flagging at the coarse band edges and centers, as well as routine flagging at the beginning and end of observations. When flags are extended over the entire analysis band, the SSINS occupancy approximately doubles, resulting in a modest sensitivity decrease. Since AOFlagger has a false positive rate of order $1/N_\nu$, where N_ν is the number of frequencies in the band, extending the flags at the finest resolution results in substantial data loss. After extending the flags at the finest resolution, the effective AOFlagger occupancy is 86%. If instead the high-resolution flags are petrified into the data during downsampling and only the low-resolution flags are extended, then the ensuing flag occupancy is 15%. However, this will leave the high-resolution excess flagging power in the data. The exact balance of sensitivity loss and the excess flagging power in a measured power spectrum depends on the true power of the foregrounds in the data, which itself depends on the foreground subtraction strategy, if any.

This excess power is beneath the thermal sensitivity of a small data set, and would require as much data as a season to truly probe in an observational setting. In order to investigate the effect of forward model subtraction, we perform the same simulations, but with only a subset of the GLEAM catalog. This simulates perfect foreground subtraction up to a certain flux threshold. We show spherical power spectra in Figure 5.10. What we find is that about 99% of the flux in the GLEAM catalog needs to be removed before the power spectrum resulting from the flagged simulation resembles the unflagged case with no foreground subtraction. For most of the modes shown, this brightness is still too high. We also note that these simulations do not include diffuse structure. This suggests that chromatic flags could only be acceptable if a foreground removal strategy that accurately subtracts greater than 99% of the flux is available. On the other hand, if achromatic flags are used and 90% of GLEAM is subtracted perfectly, the remaining foreground power is sufficiently diminished in the modes shown for typical theoretical EoR signal estimates, with a potential margin for diffuse power. Inaccuracies in the removal strategy can result in EoR signal loss, and so it is imperative to reduce the demands of foreground subtraction as much as possible.

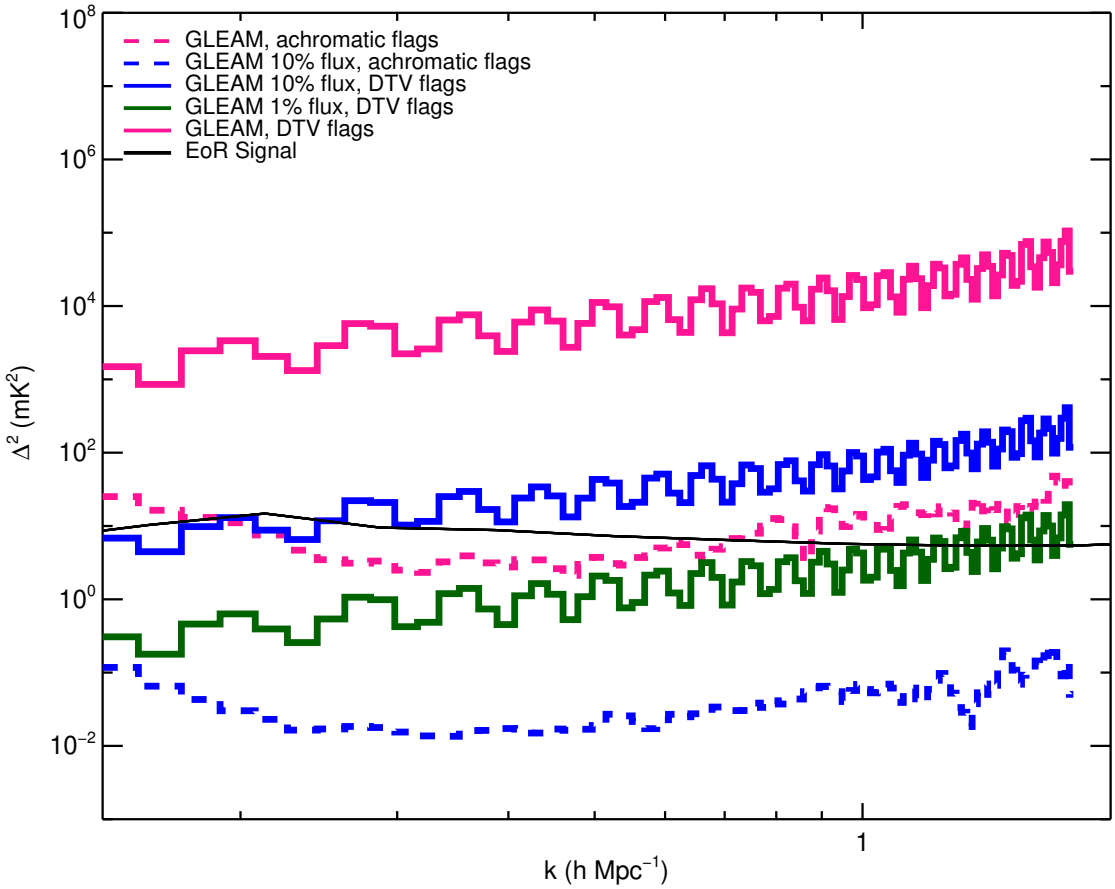


Figure 5.10: Spherical power spectra computed from simulations using GLEAM sources with various flagging and levels of foreground subtraction, compared to a somewhat optimistic fiducial EoR model (solid black). Coarse band edge flags have been removed for clarity, but would be present in current MWA EoR analyses. Solid lines show simulations with DTV flags, while dashed lines show simulations with achromatic flags. Pink, blue, and green lines show zero, 90%, and 99% foreground removal, respectively. Even with achromatic flags, some subtraction is required for a significant detection of this EoR model. With 90% of the flux removed, chromatic flagging leads to overwhelming power in the modes shown, whereas achromatic reduces the modeled foreground power to manageable levels. A significant detection of the fiducial model for most of these modes would require greater than 99% foreground removal, not including substantial contributions to foreground power from diffuse emission.

5.5 Conclusion

We find that chromatic data flagging produces excess power in the EoR window that in many cases is strong enough to prevent EoR signal detection. This effect arises because flagging disrupts the sampling function of the array by displacing baseline locations from their unflagged counterparts. The strength of the excess power is determined by both the strength of the foregrounds and the strength of the sampling function disruption. Without mitigation, chromatic flags will prevent a significant EoR detection.

The excess power is produced even in relatively clean data with overall RFI occupancy below 5%. Excess flagging power can occur in two stages. It can be petrified into the data from downsampling visibilities that have flags applied, producing partially flagged data, and it can also result when downsampled visibilities are fully flagged and excluded from the analysis. [65] discusses the case of partially flagged visibilities and provides several effective mitigation strategies. Most of these strategies are either unapplicable or fail when downsampling by only a small factor and when the dominant RFI sources are broad, producing few or no partially flagged samples.

In our investigation, the most successful strategy that balances simplicity with effectiveness is to extend the flags across the analysis band that result from a flagging strategy with a very low false positive rate. If we assume that the RFI brightness distribution is supported down to extremely faint levels, achieving a lower false positive rate by reducing the sensitivity of the flagging algorithm implies a higher false negative rate as well. As shown in Chapter 4, the cost of false negatives can be very high. The optimally sensitive flagging strategy will balance the cost of the false negative rate with the sensitivity loss from extending flags. An accurate evaluation of this cost function requires an understanding of the RFI statistics of the measurement set in terms of brightness and chromaticity.

Since the strength of the excess power depends on the foreground strength, a foreground subtraction strategy can also reduce the excess power. We find that for broad shapes such as DTV, nearly 99% of the total flux on the sky needs to be subtracted perfectly in order to

reduce the excess window power to a level comparable to extending the flags. Even at this level, contamination is still too high for constraining most EoR models, meaning substantially deeper foreground subtraction will be required if chromatic flags are used. Inaccurate foreground removal comes with the risk of EoR signal loss, which can lead to artificially low upper limits or false detections of the EoR signal if mischaracterized. Therefore, there is an array of experimental demands that must be balanced including thermal sensitivity, excess power from chromatic flags, the possible necessity for deep foreground subtraction, and the probability of missed RFI.

Chapter 6

DEVELOPING AN UPPER LIMIT ON THE 21-CM EOR POWER SPECTRUM SIGNAL PART I: PIPELINE DESCRIPTION

In this chapter we describe the power spectrum estimation pipeline for producing an upper limit on the 21-cm EoR power spectrum signal using the second season of MWA phase I data, beginning in the second half of 2014 AD. We also describe the data set, going into particular detail about its RFI occupants as reported by SSINS. Since developing a competitive limit necessitates analysis of a large data volume, this serves as an excellent opportunity to collect RFI statistics and perform tests that allow us to measure the effect of RFI on the 21-cm power spectrum.

We use an almost identical power spectrum analysis pipeline as was used in [10]. The main change between this analysis and that one is that we have now included more thorough RFI analysis. Additionally, we include a VanVleck correction for digital nonlinearities in the receiver chain [15], and some minor changes to the calibration pipeline. Since this exercise invokes the entire experimental pipeline, we review a few key concepts below, referring to more complete descriptions when appropriate.

6.1 Introduction

The 21-cm EoR power spectrum signal is extremely faint. Typical theoretical estimates of the EoR signal brightness temperature are ∼mK, while the astrophysical foregrounds have a brightness temperature of hundreds of Kelvin. This poses many challenges for the task of estimating the EoR signal. Since the EoR signal is expected to have sharp spectral features, and the foregrounds are spectrally smooth, there is a region of power spectrum space expected

to be simultaneously clean of foregrounds and yet possessing useful information about the EoR signal, called the EoR window [44]. The region of the cylindrically averaged power spectrum contaminated by foregrounds is wedge-shaped, hence it is called the foreground wedge [23, 33, 56, 95].

Measurements of the 21-cm signal for modes in the wedge will require extremely accurate foreground subtraction. Subtracting inaccurate forward models of the foregrounds will leave residuals that still overwhelm the signal [23, 95]. Furthermore, subtraction or projection of models that mix foreground and EoR components results in signal loss [22, 69]. Alternatively, measurements for modes in the EoR window have comparatively relaxed requirements since the absolute dynamic range is set by the floor of the spectral tapering function, for which there are several choices that give the required range [31]. For a succinct summary of foreground mitigation and avoidance techniques, including a discussion of signal loss, see [46]. The power spectrum measurements in this work generally operate with a foreground avoidance strategy, although as seen in Chapter 5, some amount of foreground subtraction may be necessary even for modes in the spherically averaged power spectrum that primarily receive contributions from the EoR window.

21-cm EoR power spectrum measurements are elaborate. They require precision synthesis of radio interferometer data, which can be corrupted by a plethora of systematic effects. Whenever a systematic produces spectral irregularities in the data that do not match the true foreground signal, power is produced in the EoR window, which is particularly deleterious for foreground avoidance strategies. There are many avenues whereby additional chromaticity is introduced into the data in a radio interferometer. In theory, proper calibration can remove many instrumental artifacts that produce excess chromaticity. However, calibration requirements are strict [94], and errors in calibration can produce additional structure that contaminates the EoR window [9, 19, 56]. Furthermore, not all chromatic structure in the data can be removed by standard calibration routines, e.g. digital quantization nonlinearities and RFI. Therefore, calibration cannot be the only defense against spectrally varying systematics. Much of the analysis in this work is dedicated to effects that cannot or should

not strictly be solved through calibration.

In the rest of this chapter, we describe the overall structure of the analysis pipeline. Much of this is described in more detail in [8]. We summarize the key features here very briefly, referring to that reference for detailed information.

6.2 Analysis Overview

Figure 6.1 shows a flowchart of the analysis pipeline. Different colors indicate different legs of the pipeline: data acquisition (§6.3), pre-processing (§6.4), RFI analysis (Chapter 7), calibration, imaging, integration, and power spectrum estimation (§6.5-§6.8). Data are filtered for quality at each step. The data in this analysis were collected between July 17 and September 14, 2014. They are pointed at the MWA EoR0 field, centered on a right ascension of 0 and declination of -27 degrees. Previous jackknife tests have shown that only certain pointings produce high-quality data due to the setting galaxy [14, 43].

6.3 Data Acquisition

The first leg of the pipeline is raw data acquisition, whose steps belong to blue nodes in the chart. The analog signal from each dipole within a given tile is combined at a beamformer with switchable electronic delays to allow pointing of the tile. For this data set, we use five pointings centered on the zenith pointing that progressively shift westward to coarsely track the observing field. The combined analog signal is then passed to a receiver, each of which hosts 8 tiles. Here, it undergoes analog bandpass filtration to only allow signals from 80-300 MHz, as well as subsequent analog-to-digital conversion.

The MWA applies an "FX" correlation strategy. First the signals are channelized by a series of filters and Fourier transforms in the "F-engine," and then they are cross-multiplied at the "X-engine." Specifically, the digital signals are passed to a two-stage polyphase filter bank (PFB), where they are first broken into 256 1.28 MHz coarse channels, each of which has a different digital gain applied, and onlt 24 of which are selected for further processing. For each selected coarse channel, a time series is created, which is then filtered and channelized

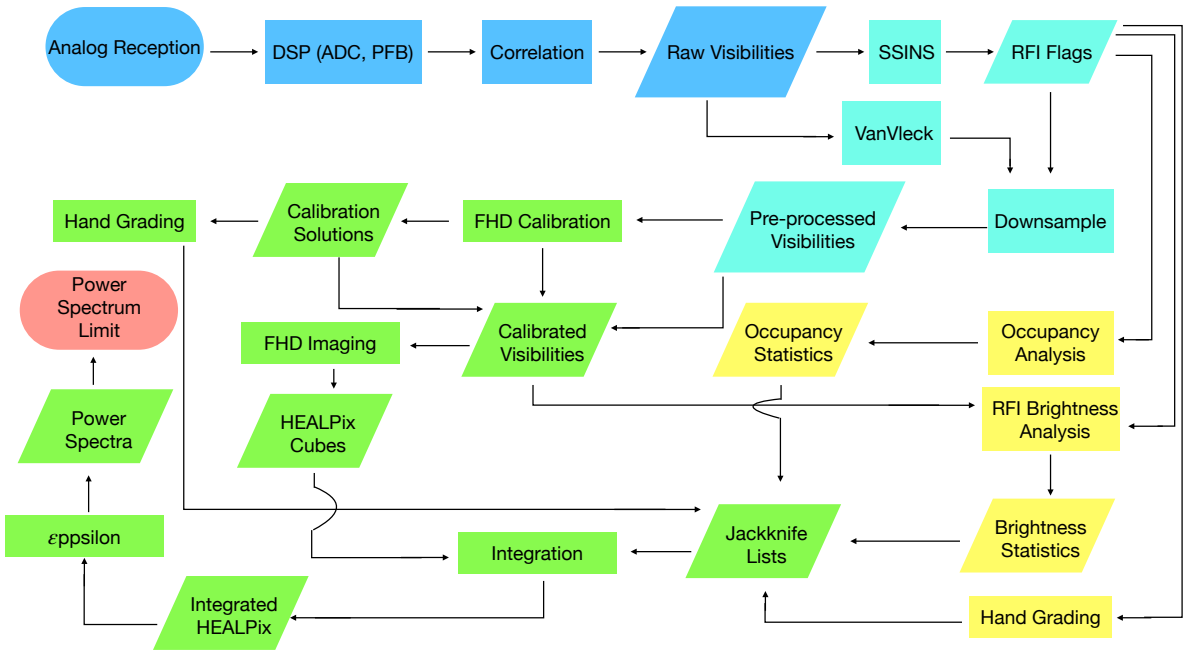


Figure 6.1: Flowchart representing the data acquisition and analysis pipeline. Dark blue nodes represent data acquisition steps. Cyan represents pre-processing. Green nodes are standard analysis steps, while yellow nodes are analysis nonstandard steps specific to this work. The red node is the final product, which is the limit.

once more, producing fine channels. The final channelization results in 10 kHz resolution with a total instantaneous bandwidth of 30.72 MHz. For more specific details about the filter shape and conventions, see [50, 79]. The signal of each tile is then cross-multiplied with each other tile signal at the X-engine, along with subsequent time-frequency averaging to produce the raw visibilities.

The raw visibilities for this analysis are written to disk at a time resolution of two seconds and a frequency resolution of 40 kHz in two minute chunks, referred to as "observations." Since there are 128 tiles to correlate, this results in 8128 cross-correlations and 128 autocorrelations for each of the 768 fine frequency channels and 4 polarizations (2 per tile). At this

resolution, each observation takes roughly 10.7 GB of disk space. The raw data is stored on magnetic tape at the Pawsey Supercomputing Centre in Perth, Australia.

6.4 Pre-Processing

Several aspects of the raw visibilities are corrected during pre-processing. The phases of the visibilities are adjusted according to the differences in cable length between the tiles and receivers, and also to specify a phase center on the sky. Since the alteration to the bandpass resulting from the PFB is known, this shape and the digital gains for each coarse channel are divided out. Due to quantization nonlinearities, this does not fully correct the bandpass, and residual frequency structure is left over. Usually, excess instrumental frequency structure, such as spectral variations in antenna gains, is taken care of during calibration. However, the digital nonlinearities require special handling.

Traditional calibration routines implicitly assume that the signal from each receiving element is a linear function of the true analog signal. For data acquisition, the analog signal in the receiving elements is digitized. Digitization is an inherently nonlinear procedure. The strength of the nonlinearity depends on the strength of the incoming signal. Due to the two-stage PFB, the instrument bandpass possesses a comb-like structure, where the teeth of the comb appear at the edges of coarse channels. Since the strength of the nonlinearity in the coarse band edges is sufficiently strong, this results in calibration errors in the coarse band edges that overwhelm the calibration error budget [11].

For the data used in this limit, we apply a VanVleck correction [15] for the nonlinear distortion present in the raw visibilities during pre-processing. The fine details of the particular implementation we employed is beyond the scope of this work,[1] so we describe it only briefly. The MWA digital signal pipeline involves several stages of quantization. The VanVleck correction employed in this work corrects the nonlinearity associated with the final round of quantization. Corrections for other quantization stages are forthcoming. The

[1] and is also not work the author should take credit for. See `https://github.com/EoRImaging/Memos/blob/master/PDFs/007_Van_Vleck_A.pdf`.

actual correction involves solving an integral equation that relates the quantized and analog visibilities. This equation has no closed-form inverse for the MWA bit depth. In order to generate fast solutions, the equation is approximately inverted using Chebyshev polynomials. This helps reduce calibration errors in the coarse band edges, but does not eliminate them entirely. For this reason, we still flag coarse band edges in the analysis.

In [10], a major improvement to the limit came from excluding observations that contained DTV RFI. The actual RFI flags from SSINS were not applied. Rather, observations with known RFI contaminants were excluded from the analysis altogether, including data within those observations that was not classified as contaminated by SSINS. In this analysis, we apply the SSINS flags to the data, extending them in frequency so as to avoid the effects discussed in Chapter 5. Actually applying the SSINS flags could allow for a measurement with reduced thermal uncertainty compared to [10]. We then compare this limit with one made where RFI-contaminated observations are excluded to varying levels. This allows us to probe the balance between thermal noise and systematic contamination introduced by RFI, which we theoretically characterized in Chapter 4. We discuss this portion of the analysis thoroughly in §§7.2 and 7.3.

In this limit, we do not apply AOFlagger flags, unlike every other MWA EoR limit so far. Upon examining the outputs of each flagging pipeline, we found a modest number of RFI events that were caught by AOFlagger and not by the pre-extended SSINS flags. After frequency extension, many of these events were recovered. Additionally, some samples that were flagged by AOFlagger and not by SSINS may be false positives that result from AOFlagger's morphological detection algorithm, which can overextend flags for broad features [62]. Due to the inability to generically extend the AOFlagger flags without causing massive data loss, we opt to entirely forego them at the possible expense of a few extra false negatives in the entire season.

6.5 Calibration

Standard calibration routines assume that the measured and true visibilities are related by

$$v_{ab}(f,p) = g_a(f,p)g_b^*(f,p)u_{ab} + n_{ab}, \tag{6.1}$$

where $v_{ab}(f,p)$ represents the measured visibilities for antennas a and b as a function of frequency and polarization, g_a and g_b are the respective gains for those antennas, u_{ab} are the true visibilities (apparent spatial coherence function samples), and n_{ab} is a noise term. We use traditional sky calibration in this work, implemented in FHD, as described in [8]. Traditional sky calibration works by substituting a forward model for the true visibilities in equation 6.1 and finding the gains that minimize the squared residuals between the gain-multiplied forward model and the measured visibilities. This is often opposed to traditional redundant calibration, which leverages redundant baselines within an array. Perfectly redundant baselines would give exactly the same measurement of the sky in the absence of noise. Traditional redundant calibration does not employ a sky model, which results in degeneracies that can only be broken by including a sky model [19]. In [20], these traditionally opposed calibration frameworks are unified by Bayesian considerations, showing that sky calibration and redundant calibration reflect particular prior assumptions about the fidelity of the sky model. In particular, sky calibration assumes that the model is perfect, i.e. it has no uncertainty, while redundant calibration assumes infinite uncertainty in the model. Their work also shows that calibration is more robust when uncertainty in the model is properly quantified, and thus the calibration in this work could be improved in such a way.

We summarize the calibration steps here, drawing mainly on the discussion in [8], and explaining a few minor updates. First, per-frequency, per-antenna, per-polarization gain solutions are found by minimizing the aforementioned squared residuals using an alternating direction implicit method [81]. To speed calibration relative to what was used in [10], we employ Kalman filter during calibration. Note that the physical antenna gains are spectrally smooth, while these gain solutions are noisy. To produce a smooth solution, we divide out an estimate of the bandpass, and then fit a low-order polynomial for each gain amplitude, as

well as a phase ramp. Additionally, we fit for a cable reflection mode in the spectrum, which to first order multiplies the gain by a complex sinusoid [40]. We then divide by these smooth fit solutions rather than the initial noisy estimates to obtain the calibrated visibilities.

There are a number of data-driven ways to estimate the bandpass. We use the autocorrelations in the style of [43]. We time-average the autocorrelation ratios of each tile, which acts as an approximation of the true antenna gain amplitudes. Denoting the time-averaged autocorrelation ratio of antennas a and b as A_a^b, we have equation (12) from [43]:

$$A_a^b(\nu) \approx \frac{|g_a(\nu)|}{|g_b(\nu)|}. \tag{6.2}$$

Denoting the (noisy) gain solution from the per-frequency step as $\hat{g}_a$, we can correct for the bandpass via

$$g_a(\nu) \approx \left\langle A_a^b(\nu)|\hat{g}_b(\nu)| \right\rangle_b \tag{6.3}$$

i.e. we average the product of the autocorrelation ratios and estimated gains over the tiles in the array. This averaging avoids pitfalls associated with using a particular reference antenna, which may cause idiosyncrasies of that antenna (e.g. imperfect cable reflection fitting) to be imprinted on the other tiles. For more details, please see [43].

We hand-grade calibration solution for quality. Figures 6.2 and 6.3 show example gain amplitude solutions from the same night. These are each for two-minute observations. We generally notice in this data set that calibration solutions from different pointings can appear characteristic. Figure 6.2 belongs to the -2 pointing while Figure 6.3 belongs to the -1 pointing. The solutions in Figure 6.2 have more noticeable frequency structure than those in Figure 6.3. This trend is common across all nights in the data set used for this limit. We observe that the solutions for tile 44 are consistently outlying compared to other solutions in that its gain amplitudes in the Y polarization have significantly more chromatic structure than other tiles. This is indicative that something may be malfunctioning with the tile. For this reason, we remove tile 44 from the rest of processing.

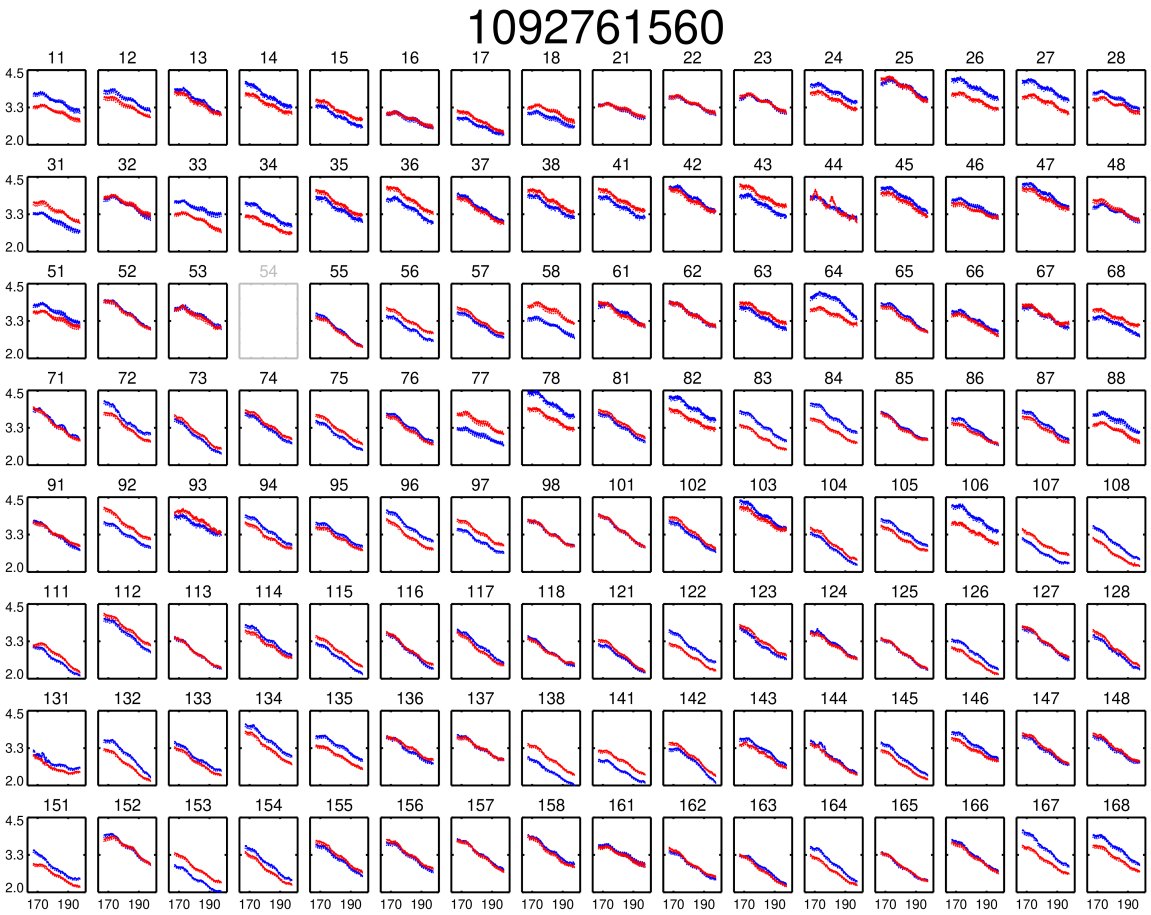


Figure 6.2: Gain amplitudes for an observation in the -2 pointing. Each panel belongs to a particular tile. The blue line is the X (East-West) polarization while the red line is the Y (North-South) polarization. Tile 54 was excluded from the data set during data acquisition. Tile 44 exhibits an unusual chromaticity in its North-South solution.

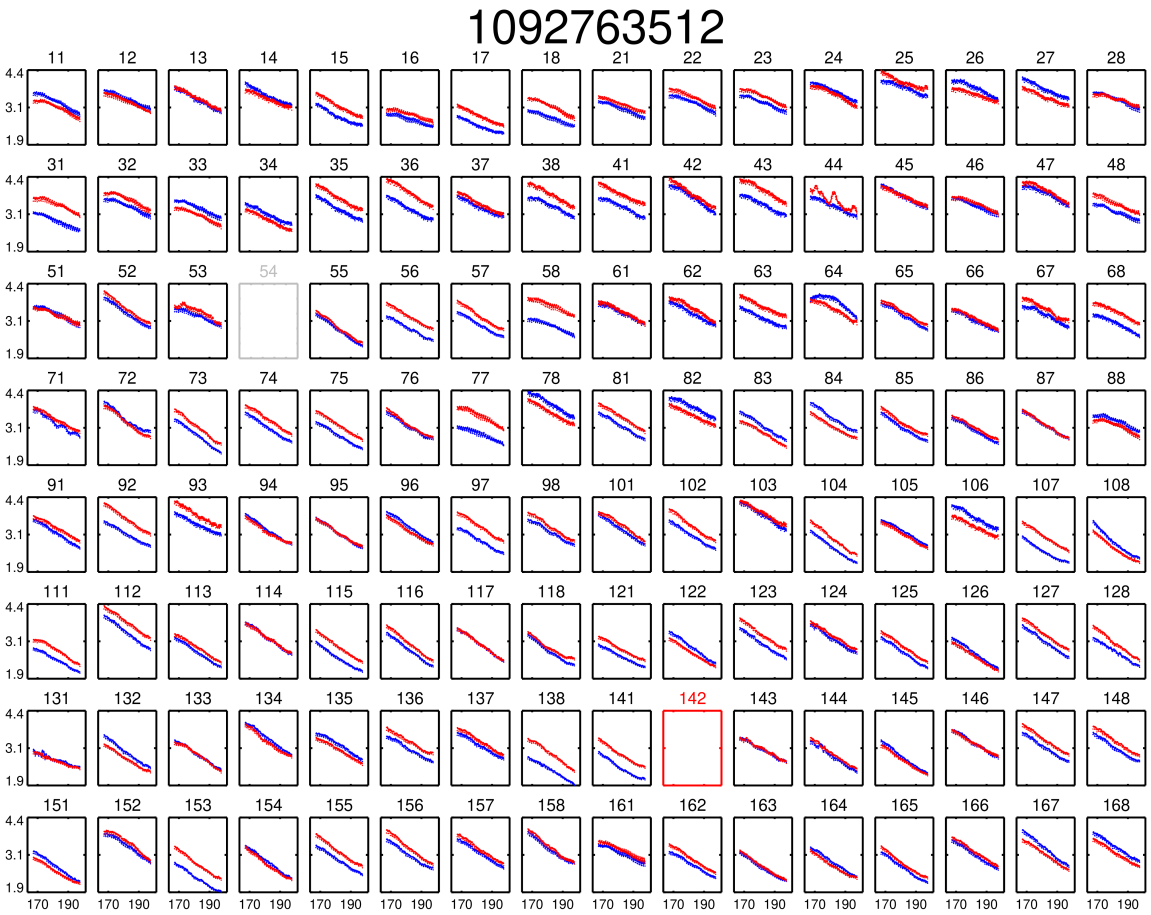


Figure 6.3: Gain amplitudes for an observation in the -1 pointing. There is noticeably less frequency structure in these solutions compared to those shown in Figure 6.2. Tile 44 also exhibits excess chromaticity in this observation. Additionally, Tile 142 was flagged by the calibration routine as an outlier, and is excluded.

6.6 Imaging

After calibration, the data and model are imaged to create HEALPix cubes [29] for passage to εPPSILON. We again draw on [8] for this description, providing essential details here. To begin this process, the calibrated data are gridded to the uv-plane. Gridding is the process of propagating the visibilities to a finely discretized uv-plane in order to generate an estimate of the apparent spatial coherence function. The gridded data are then Fourier transformed to a fixed sky frame to form the HEALPix cubes. We grid a different uv-plane for each frequency so that we can make separate HEALPix maps at each frequency and therefore probe the cosmological line-of-sight distance.

Choices made during gridding affect the properties of the resulting map. For mapmaking purposes, i.e. to make the most faithful map of the sky given the lossy measurement process, it is provably optimal (in terms of a signal-to-noise ratio) to grid with the conjugate of the Fourier transform of the receiving element's beam pattern [58, 89], henceforth referred to as the uv-beam. Each visibility is smeared around the baseline-center location according to the uv-beam pattern in order to estimate the apparent spatial coherence function. Since this reflects a particular choice of optimization, and the true spatial coherence function can never be recovered, it is not uncommon to use a different gridding kernel than the uv-beam when reconstructing the uv-plane for the advantages that it might bring. Precision requirements for interferometric gridding in 21-cm power spectrum estimation are specifically discussed in [66], where a range of parameters regarding the gridding kernel are analyzed.

[10] found that using a modified gridding kernel increased the dynamic range of the measurement without signal loss, allowing for a deeper power spectrum limit. The choice of kernel was equivalent to windowing the apparent sky with the square of a Blackman-Harris window function. This is implemented by convolving the uv-beam with the Fourier transform of the aforementioned window function and gridding with the resulting kernel. This greatly reduces the survey volume and leads to longer correlation lengths between uv-points [8].

To construct the model, we Fourier transform point-sources of the GLEAM [36] catalog

to generate their discretized spatial coherence function. These are then degridded[2] with the instrumental beam to produce model visibilities. The model visibilities are then regridded using the modified gridding kernel. This accomplishes an in-situ simulation or forward model of the measurement pipeline with a set of known radio point sources on the sky. Other source distributions can also be used. We refer to the calibrated visibilities and resulting products as "dirty." When we subtract the model from the "dirty," we refer to the result as a "residual." The gridded *uv*-data are Fourier transformed to the sky domain to form the HEALPix cubes, which are embedded in a fixed sky-frame and thus allow for coherent integration.

6.7 Integration

Since the noise in each observation is independent and the signal is coherent as long as the integration frame remains fixed on the sky, we can integrate multiple HEALPix cubes together in order to boost our signal-to-noise ratio. This step serves as a useful point to set up jackknife tests, where we split the data into distinct populations and examine the effect on the measurement. We can integrate each subset separately and experiment with combinations. This allows for the isolation of scientific variables so that we can examine whether certain systematic effects have an effect on our power spectrum measurement. We discuss our jackknife tests in detail in §§7.2-7.4.

6.8 Power Spectrum Estimation

We ultimately want to measure the power spectrum of the 21-cm brightness temperature map. We use the Error Propagated Power Spectrum with Interleaved Observed Noise (εPPSILON) software package for this purpose. εPPSILON provides power spectrum estimates along with uncertainty estimates for the measurement. To estimate the power spectrum, the integrated HEALPix cubes must be transformed back to the Fourier domain.

If a Fourier transform and square operation is performed on a single coherently integrated

[2]Gridding in reverse, i.e. reproducing the visibilities from an apparent spatial coherence function.

HEALPix cube, the noise will self-correlate, resulting in a noise pedestal in the power spectrum measurement. To avoid this, interleaved even and odd time steps within an observation are gridded separately and turned into different HEALPix cubes. All even and odd HEALPix cubes are integrated separately so that when the power spectrum is formed, it can be formed by cross multiplying the even and odd cubes rather than squaring an individual cube. Since the 21-cm signal is expected to integrate coherently and the HEALPix frame is fixed, this produces the same signal term while producing different noise terms. Most importantly, the term involving only noise is now zero-mean, meaning the noise pedestal vanishes.

Due to missing data and the desire for uncorrelated errors between the cosine and sine modes in the Fourier domain, the Fourier transform along the frequency axis is given special treatment. Instead of an ordinary discrete Fourier transform, a generalized form of a Lomb-Scargle periodogram is used [10], which phase-rotates the resulting cosine and sine modes of the Fourier transform such that the error distributions for cosine and sine modes in the same power spectrum bin are uncorrelated. The transition to this rotated mode is still linear, which is subtly important for analytic error propagation, detailed next.

Error propagation is a major driver of the design of εPPSILON. The thermal noise in the visibilities is a colored circular gaussian process. The entire pipeline involves gridding, Fourier transforms, and weighted averages, which are all linear transformations of the visibility data. Since circular Gaussian noise remains circular Gaussian under linear transformations [28], the thermal noise term is Gaussian for most of the pipeline. The power spectrum is quadratic in the data, which is a strictly nonlinear transformation. It can be shown that the thermal term in each 3d voxel of the εPPSILON power spectrum estimate is biexponentially distributed (Appendix E).

To form spherical power spectra, voxels are averaged together with inverse-variance weights according to the array sampling function. To determine the weights, a value of 1 is gridded everywhere that a visibility is reported using the modified gridding kernel. Similarly, the sample variance of the visibilities is gridded with the square of the modified gridding kernel. To propagate the weights and uncertainties, HEALPix cubes are made for the weights

and variances, and integrated along with the data cubes. It can be shown that averages of independent biexponential random variables converge to Gaussian random variables under the central limit theorem, and it can be shown that this happens fairly quickly. Strictly speaking, not all voxels that contribute to a spherical power spectrum bin are independent, and some bins have few-enough contributors that a Gaussian uncertainty distribution is inappropriate. These bins tend to be on lower-order k-modes. For the k-modes shown in spherical power spectra in this work, the uncertainty distribution is expected to be very Gaussian. Nonetheless, εPPSILON uses internal simulations to approximate the non-gaussian distribution and subsequent distributions for the weighted averages. Implementation of a more exact calculation of the error distribution as in Appendix E is a subject of future work, and is expected to make a very small effect on the error estimate.

A second form of noise estimation that is employed in εPPSILON is to use the difference between the even and odd cubes. If all significant RFI is flagged, the difference cube should only contain thermal noise. Therefore, the difference power spectrum serves as a proxy for the uncertainties in that their expected amplitude should be analytically related to the actual uncertainties elucidated above. These error estimates are used during validation to ensure that the noise is being propagated correctly.

The in-situ simulation capabilities of FHD allow for robust tests of signal loss. By simulating visibilities for a fiducial EoR signal and estimating a power spectrum from them under various conditions, [10] verifies that there is no appreciable signal loss in this power spectrum estimation pipeline. This and our end-to-end error propagation allows us to confidently place upper limits on the EoR power spectrum signal in our measured power spectra.

Chapter 7

DEVELOPING AN UPPER LIMIT ON THE 21-CM POWER SPECTRUM SIGNAL PART II: DATA CURATION AND POWER SPECTRUM RESULTS

In this chapter, we present power spectra made using the pipeline described in Chapter 6. The purpose of this analysis is twofold. Examining an entire season data allows us to more clearly understand the RFI statistics of the MWA than with a smaller data set. We can also probe the effect of RFI on power spectrum measurements at varying levels of integration depth. Thus, the first purpose of this analysis is to measure the effect of RFI on power spectrum limits. By comparing power spectra of known contaminated observations to matched observations identified as clean by SSINS, we generally find that excess window power is correlated with the presence of RFI, suggesting that there is faint residual RFI undetected by SSINS. We also find a small number of supposedly clean power spectra that contain similar excess window power. Since much of this analysis already involves deep power spectrum integrations, we take the opportunity to use these power spectra to construct a final integration list for a 21-cm EoR power spectrum upper limit.

In §7.1-§7.3, we describe the data curation process. We begin with preliminary data cuts related to the ionosphere. We then describe the RFI statistics of the entire season in terms of reported SSINS occupancy and brightness. This analysis provides a framework for the power spectrum jackknife testing that we present in §7.4. While we present RFI statistics for the entire season, only data that passes the ionospheric cut is used for the power spectrum tests. Using the results of these tests, we further reduce the data set to a final integration list for setting the upper limit, presented in §7.5. In §7.6, we summarize this work and present our final conclusions.

7.1 Preliminary Data Cuts

The MWA beam pattern changes for different pointings. Our observations in this work are centered on the same field. However, as the beam changes pointings, different sources enter the sidelobes. Certain sources are difficult to model, and this can lead to calibration errors that produce poor power spectrum measurements. In [14], it was found that the central five pointings around the zenith pointing gave the best power spectrum measurements. This was also corroborated in [43] for phase II data. We follow suit in this work, only using the central five pointings. Some of these observations failed pre-processing, so we do not include them. The total number of observations in this set is 3168.

It is well-known that the Earth's ionosphere can affect radio wave propagation. Using ionospheric metrics developed in [37] that capture source position offsets that result from ionospheric activity, [93] produces several deep EoR power spectrum limits that are improved by removing observations of poor ionospheric quality. We use this ionospheric quality metric to reject all observations with ionospheric quality metric greater than 5, which is a relatively harsh cut. Observations need to be calibrated by the MWA Real Time System [54] before their ionospheric quality can be judged. If they fail to calibrate, they will not have reported values. Observations that do not calibrate are suspicious, so we cut all observations that do not have reported ionospheric quality values. This cut out a large number of observations, leaving only 1285 observations.

We show a graphical representation of the initial data selection process in Figure 7.1. Relatively few full nights made it past the ionospheric data cut. This impacts the uv-coverage of the final measurement since rotation synthesis allows a given physical baseline to probe more spatial modes of a given field of view. The remaining observations after the ionopsheric cut constitute 40 hours of data. Over 1000 hours of MWA data will need to be integrated to make a significant EoR measurement [14]. However, systematic effects are still strong enough that smaller integrations yield valuable lessons.

All remaining data cuts are based on the RFI statistics of the data. This is the primary

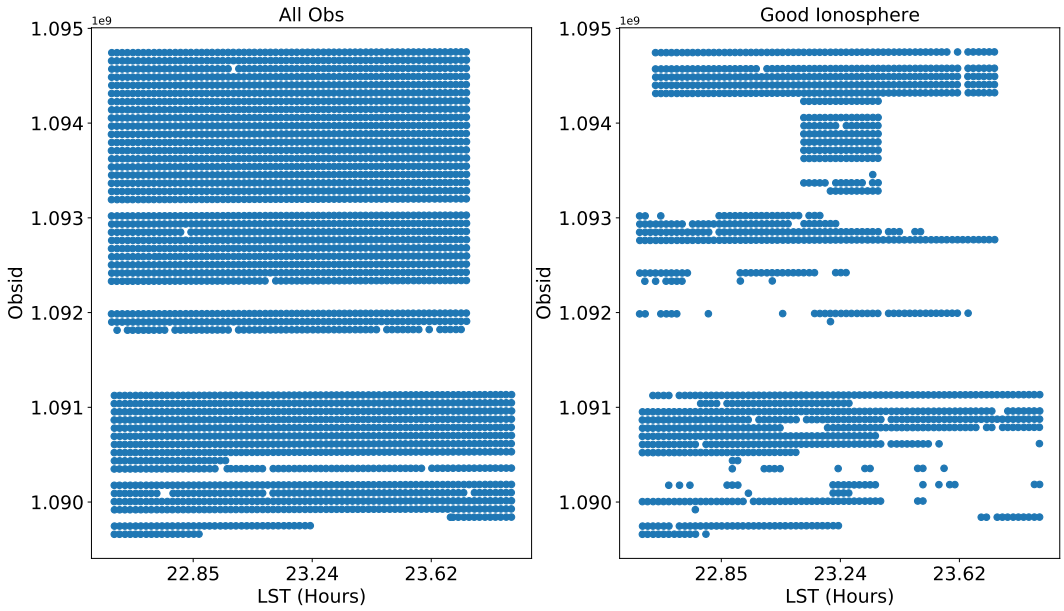


Figure 7.1: Scatter plots of each observation's ID (GPS start time) and its starting local sidereal time. The left plot shows all observations in the five central pointings. The right plot shows all observations remaining after cutting out observations with poor ionospheric quality. Most of the data in the third quarter of the season was removed by the ionosphere cut. Interestingly, the zenith-pointed observations in the cluster near GPS time 1094000000 seemed to pass the cut despite all other pointings failing on those nights.

aspect of the analysis meant to be probed by this limit, and so we devote the next two sections to describing these cuts.

7.2 RFI Occupancy Analysis

Between iterations of the SSINS frequency-matched flagging algorithm, the time, frequencies, and shape label of the most recently identified RFI event are recorded. We analyze this flagging metadata to understand the occupancy statistics of the data set used for the limit in order to better understand the RFI content. First we discuss a few important caveats about this flagging metadata related to misclassification. Then we present the occupancy results.

7.2.1 RFI Type Misclassification

While SSINS does use prior information about possible RFI contaminants and does associate each event with a label, it is optimized for flagging rather than classification. This manifests when the shape dictionary has entries that overlap in frequency. A particular common example is broadband streaks versus any DTV contaminant. The frequency-matched flagging algorithm does not compute a probability for when two overlapping shapes might occur simultaneously, hence there is no discrimination for when additional excess intensity from a streak event boosts the calculated z-scores for a DTV event, and there is no event in the output metadata that reports such a combination. Due to the iterative nature of the flagger, the most common outcome when this occurs is that both shapes are caught in separate iterations and are reported separately. Note that the computed flags will be identical regardless of the resulting classification. An instance of a misclassification is when several (3-4) DTV events occur simultaneously. Since combinations of DTV signals are not checked, this can sometimes be reported as a broadband streak event. Due to our flag-extension strategy (Chapter 5), the calculated flags in this case are also identical to the properly classified case. Adding more accurate shape discrimination and subsequent classification significantly increases the computational complexity of each SSINS iteration (Appendix A) without producing more effective flags. For the purposes of producing an EoR power

spectrum limit, we consider this sufficient. However, a more exhaustive analysis of the RFI environment would require enhancing the classification capabilities of SSINS, and we may do this at a later time.

We also briefly remark on a choice of implementation regarding flag extension. Flag extension can be done between flagging iterations or after all flagging stops. Extending between iterations affects the calculated z-scores and subsequently the final calculated flags and metadata. We find that in about 10% of the data, this results in slightly different calculated flags, with no obvious false negatives in either case. The occupancy statistics are more clearly understood when flag extension is performed after the fact, and the overall occupancy is very nearly equal. So, for the purposes of a more clearly understood metadata analysis, we compute these flags. However, the flags applied to the data involved flag extension between iterations.

7.2.2 Nonstationary Thermal Noise and Insufficient Time Sampling

SSINS relies on two assumptions to accurately detect RFI according to the significance threshold. The first assumption is that the thermal noise is weakly stationary, i.e. that the mean and covariance function are not changing in time. The other is that a sufficient number of time samples exist in the observation such that the uncertainty in the sample mean does not disrupt the Gaussianity of the z-scores. Violation of these two assumptions tends to cause overflagging and underflagging, respectively. We demonstrate that in this section by employing the SSINS match filter on simulated gaussian spectra with these properties.

The thermal noise in the visibilities, which is summarized by the system temperature, can be non-stationary for several reasons. The system temperature has contributions from the sky and the instrument. If a particularly bright source moves into the field of view, then this can make a meaningful change on the overall system temperature. This is more noticeable for instruments with narrower fields of view that are not source tracking, or for instruments that suddenly change their field of view and introduce new sources. Many hardware components often have temperature-dependent performance. Some of the amplifiers

in the signal chain have a temperature-dependent gain [11]. As the telescope functions, it generates heat, and must be refrigerated to maintain optimal performance. This refrigeration cycle can cause percent level fluctuations in the average thermal noise amplitude over the course of an observation, which is visible in the MWA SSINS. This effect is relatively frequency-independent, causing SSINS to over-report broadband RFI events.

We capture the nonstationary behavior of the SSINS background in Figure 7.2, where we show MWA SSINS averaged below the TV band for a night of data alongside PPD values for the ORBCOMM frequencies. The PPD[1] is a measure of total tile power evaluated after coarse channelization but before the channel select. This allows for examination of the broad spectrum all the way from the DC mode to the nyquist frequency. One possibility for broad SSINS variation is the tails of extremely bright, out-of-band RFI. In Figure 7.2, the PPD values in the ORBCOMM frequency only synchronize with a handful of small SSINS variations. The slow, rolling features do not correlate with out-of-band RFI including ORBCOMM, FM radio, and the military band above 223 MHz. We also observe that ionospheric behavior as measured by the metric in the previous section is uncorrelated with these variations.

We suggest that these arise due to slow variations in the system temperature as a result of heating and cooling of the receivers, which affects the gain in the receiver amplifiers. The physical receiver temperatures are not kept for archival data, however they take a very characteristic shape while the array is operating. They appear roughly periodic on a ~ 10 minute timescale with a clear exponential rise/decay pattern within a period. Each receiver hosts 8 tiles. In the top panel of Figure 7.3, we show tile autocorrelations averaged over the same frequency band as the SSINS in Figure 7.2 and averaged by receiver. We see that each receiver has a somewhat coherent expoential rise/decay time profile with a high signal-to-noise ratio, however each receiver has a slightly different average amplitude and time offset. When the total average of these is computed (bottom panel), we recover the

[1]Unfortunately the meaning of this acronym has been lost to the sands of time.

SSINS time profile of Figure 7.2 using only the autocorrelations (recall that the SSINS are computed only from the cross-correlations).

In order to quantify the extent to which nonstationary noise generates false positives, we used a simulation. We began by taking the night of SSINS data shown in Figures 7.2 and 7.3, and removing any obviously RFI-contaminated observations. We then fit a line to each frequency channel and polarization in the SSINS of each observation. Next, we sampled a gaussian spectrum whose mean (and therefore standard deviation given the properties in chapter 3) was determined by this line at each time, frequency, and polarization. We then ran the SSINS match filter, as we would on the data, and determined how many samples were flagged. Finally, we averaged this process over many trials to obtain the average false positive ratio for that spectrum. In Figure 7.4, we show a scatter plot of the average false positive ratio against the absolute value of the average SSINS slope for each observation. The variation makes it difficult to tell precise details, but there is a strong positive correlation. In order to reduce sensitivity to this particular effect, we raise the significance threshold of the streak flagger to $10\hat{\sigma}$, as opposed to the typical $5\hat{\sigma}$ cut used for other shapes. With these settings, we find that the median false positive rate is 0.04%.

Having fewer time integrations with which to estimate the mean per frequency and polarization prevents the z-scores from being approximately Gaussian. A closed form expression exists for this distribution under certain conditions, however it is unilluminating.[2] What we generally find is that the z-scores calculated by SSINS for simulated stationary gaussian noise are more concentrated than a standard normal random variable, meaning that the z-scores are underreported relative to the significance threshold. This leads to underflagging. We summarize this in Figure 7.5. For this reason, we completely flag all observations where greater than 60% of integrations have flagged.

[2]Since the estimated standard deviation is proportional to the estimated mean, it is not the Student-T distribution. It is instead the quotient distribution of two Gaussian variables with identical mean but different standard deviation, offset by a constant. If there are sufficiently few samples such that the correlation between the estimated mean and any given sample cannot be ignored, then even the previous statement is not true.

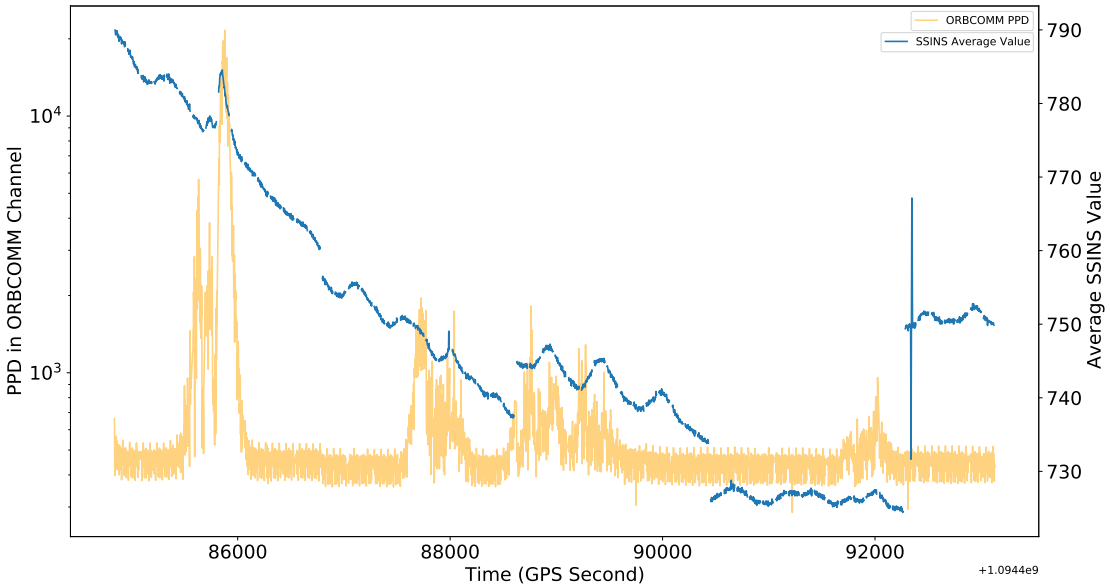
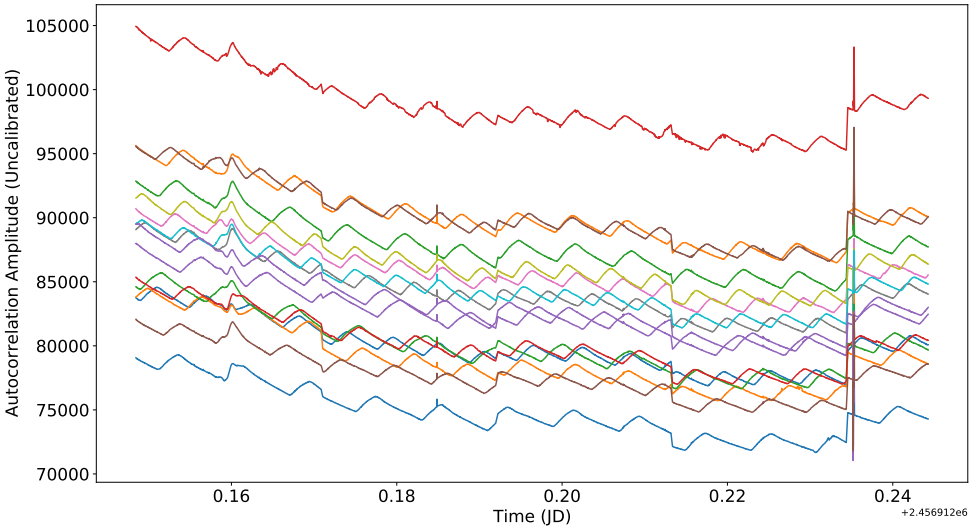
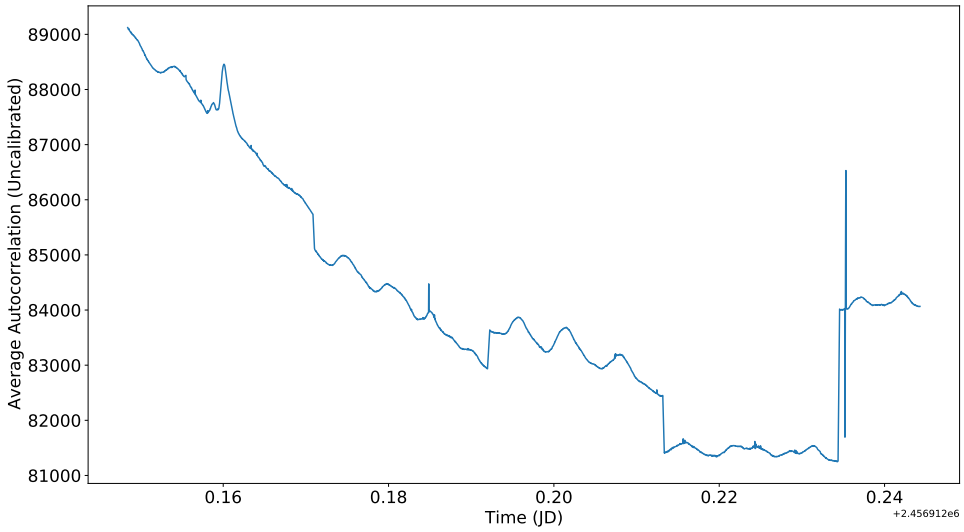


Figure 7.2: MWA SSINS for a night, averaged over frequencies below the TV band (blue), along with PPD values at ORBCOMM frequencies (transparent orange). We see that some sharp variations in the SSINS are correlated with extremely bright ORBCOMM calues, however most variations are uncorrelated with ORBCOMM. Note that the ORBCOMM values are on a log axis, while the SSINS values are on a linear axis. Discontinuities occur at pointing boundaries, which may be due to the pointing-dependent beam pattern gathering different sources in the sidelobes.



(a)



(b)

Figure 7.3: Top: Receiver-averaged autocorrelations for a night of MWA data. Each receiver has a time profile reflective of a refrigeration cycle occurring on regular intervals, though the refrigeration waveforms are not in phase between different receivers. Bottom: Average of the top panel over receivers, which recovers the SSINS shape in Figure 7.2.

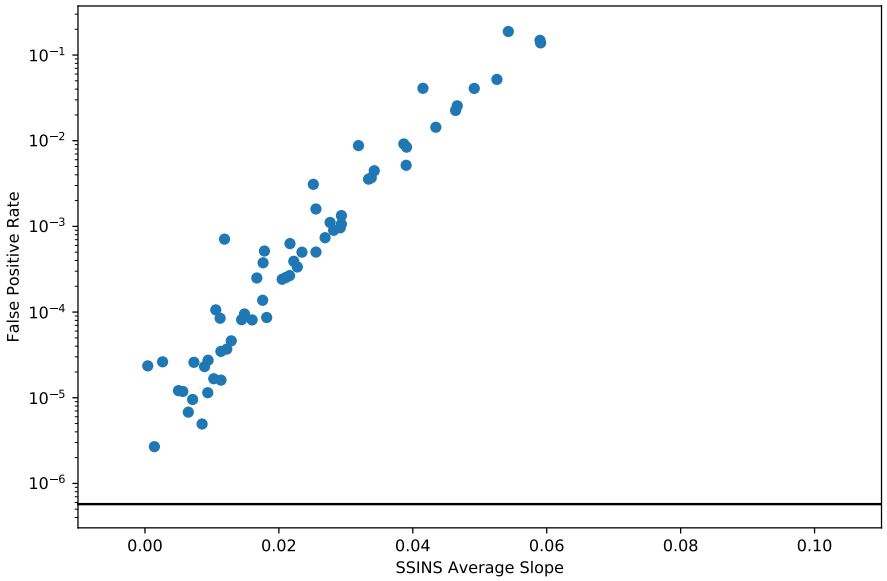


Figure 7.4: Average SSINS slope (change in SSINS per integration, averaged over frequency and polarization) vs. false positive ratio on a semilog plot. The solid horizontal line indicates the theoretical false positive ratio for stationary gaussian noise, which is well-matched by simulations of stationary noise (not shown).

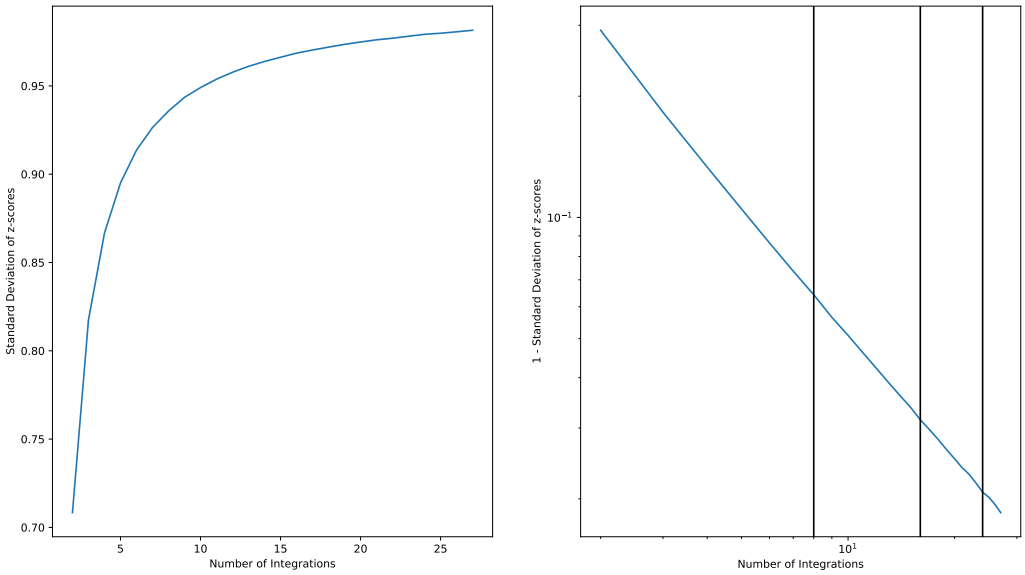


Figure 7.5: Left: The standard deviations of the z-scores calculated by SSINS, which should be close to 1 when the Gaussian approximation holds, plotted against integration number. Right: The difference between 1 and the standard deviations of the z-scores, with vertical lines at 8, 16, and 24 integrations, on a log-log plot. It clearly follows a power law. This suggests that at least $\sim$ 16 samples should be used for flagging purpose, and that if the number of unclassified samples in a channel falls below this number, the remainder should probably be flagged more aggressively.

7.2.3 Summary of Flagging Settings

We search for four 7-MHz-wide DTV signals, all adjacent, starting at 174 MHz: channels 6-9. Channel 9 begins at 195 MHz, so we only observe 2.8 MHz of its allocated bandwidth. We also search for narrowband occupants and broadband streaks. For all shapes other than broadband streaks, we use a significance threshold of 5. For broadband streaks, we

Shape Label	Frequencies (MHz)	Significance Threshold
TV6	174-181	5
TV7	181-188	5
TV8	188-195	5
TV9	195-202 (197.8)	5
Narrowband	Various	5
Streak	167.1-197.8	10

Table 7.1: Association of shapes in the match filter, their frequencies, and the significance thresholds used in the filter. The TV9 shape is allocated all the way to 202 MHz, but our observing band cuts off at 197.8 MHz. Narrowband shapes are 1 fine channel wide and allowed anywhere in the band.

use a significance threshold of 10. Between iterations, we flag a frequency channel for the entire observation if its occupancy is greater than 60%. We also extend flags in frequency between iterations. This combination means that if any RFI occupant is found for 60% of an observation, the entire observation is flagged. We summarize this in Table 7.1.

7.2.4 Occupancy Results

In this section we discuss the RFI occupancy of the data set as reported by SSINS. We find that the overall distribution of RFI occupancy is relatively unaffected by the ionospheric cut, and so only show distributions before the ionospheric cut.

In Figure 7.6, we show a histogram of TV occupancies on a per-observation basis, before the ionospheric cut. We observe that channels 6, 7, and 8 have very similar occupancy distributions, while channel 9 is much rarer to observe, with 96.5% of its mass contained at 0 occupancy. Examining broadcasting stations in Western Australia, we find that channels 6, 7, and 8 are broadcast relatively equally, while channel 9 is broadcast less often. Since we only observe 2.8 MHz of channel 9's allocation, we also expect that SSINS is less sensitive

to its presence. While the ionospheric cut reduces the strength of the tails, the overall shape of the distribution appears similar, and this is likely a manifestation of the fact that the ionospheric cut excluded about 2/3 of the data.

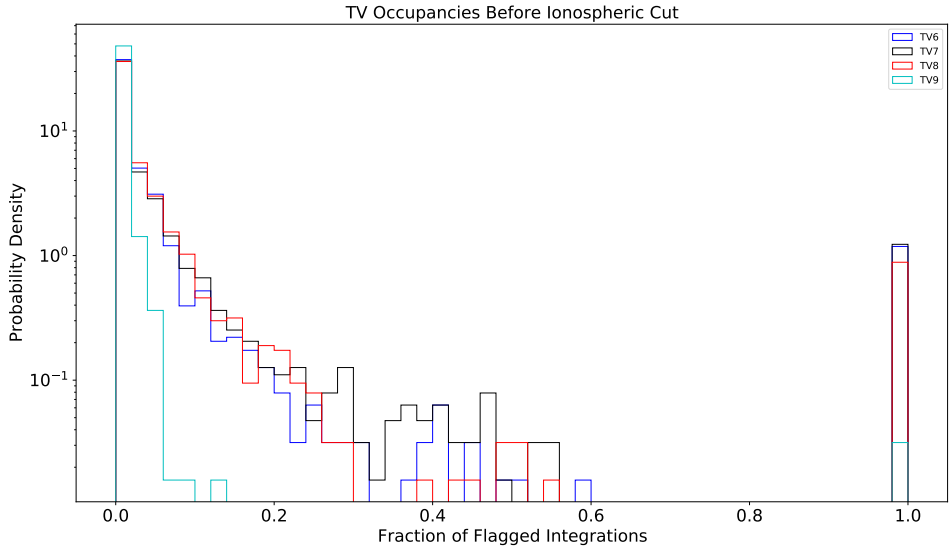


Figure 7.6: Mass histogram of DTV occupancies per observation, before the ionospheric cut. See text for discussion.

Since the majority of the observing band is allocated for DTV, we expect very few narrowband events, which is consistent with what we observe. Curiously, most narrowband events that are clearly true positives are within the DTV allocation. Some of these events are extremely bright, and may be a result of out-of-band RFI clipping the ADC or causing intermodulation products. Alternatively, there could be locally generated RFI.

The streak shape is the most prone to false positives due to nonstationary thermal noise. Most anthropogenic signals at these frequencies have their power narrowly concentrated relative to the observing band. Extremely bright out-of-band RFI can produce broad fluctuations in the SSINS. For instance, Figure 7.7 shows the mean-subtracted SSINS in the

E-W polarization for the observation in Figure 7.2 with the peak that strongly associates with strong ORBCOMM power. Observations of this quality tend to be entirely flagged by SSINS. Occasionally the RFI event will straddle the boundaries of observations, or proceed through multiple observations, and so observations surrounding ones such as these will also have high streak occupancy. Non-anthropogenic signals, such as lightning, can also be broad. There does appear to be a noticeable difference in distributional shape after the ionospheric cut. We have not evaluated if this is statistically significant. From physical considerations, it is possible that ionospheric activity is correlated with atmospheric conditions prone to long-range transmission of RFI or lightning events, e.g. sproadic E propagation or tropospheric ducting, however a more careful study would be necessary to assess such a claim. We note that when we scatter plot streak occupancy versus the ionospheric metric, no obvious correlation appears.

Since we suspect a significant number of the RFI events recorded by SSINS are due to reflections from aircraft, we seek to understand the timelike properties of the RFI occupancy. These properties are more clearly understood with the full season data set before the

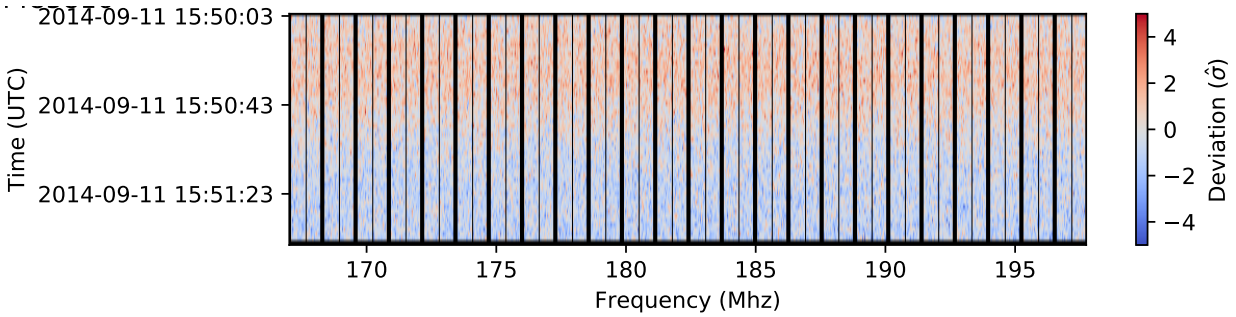


Figure 7.7: Uncalibrated mean-subtracted SSINS for an observation during a bright ORBCOMM event. The z-scores have a strong gradient in the time direction, causing the observation to be fully flagged for broadband streak events. This is likely the tails of the extremely bright ORBCOMM signal.

ionospheric cut, and since the ionospheric cut does not seem to signficantly alter the RFI statistics, we only show the time statistics of the full season.

First, we notice that the different TV channels are sometimes flagged simultaneously, and flags tend to appear in clusters within a night. By constructing the autocorrelation function of the time series of flags belonging to a given channel, we sometimes observe that flags for different TV channels are strongly correlated within a night, however not all nights have such correlations. Since not all broadcasting stations in Western Australia transmit at the same frequencies, and there is probably a range of aircraft trajectories that intercept these various stations. Furthermore, occasionally direct transmission of the signals is possible. For instance, there is a night of times with extremely high TV occupancy on August 27, 2014, clearly visible in the occupancy scatter in Figure 7.8 near GPS time $1.093 \cdot 10^8$. The SSINS of this night show extremely strong, persistent DTV interference. This is probably a ducting event allowing for long range direct reception of the RFI. Since the RTS could not calibrate these observations, no ionospheric metric could be gathered. Therefore, we expect that occasionally we sometimes observe different DTV signals simultaneously with regularity, but sometimes not, and that the flags might reflect this in their two-point statistics. This expectation is corroborated by manual inspection of the SSINS.

In Figure 7.8, we show the total occupancy of each observation with the flags extended. Here, the clustering behavior of the flags is more apparent. To also illustrate this clustering, we show a histogram of interarrival times between SSINS flags in Figure 7.9, not including the night with strong ducting, along with a Poisson process Markov chain counterparts. To understand the motivation behind such a plot, consider a perfectly periodic arrival process where continuous pulses of length t appear spaced apart by T, $T \gg t$. Supposing this process is sampled at rate f_0, a histogram of the arrivals would be concentrated on two values: $1/f_0$ and $(T - t)$. On the other hand, a Poisson process where arrivals occur randomly but with a mean arrival rate that applies equally well to all time increments of a given size will have exponentially distributed arrival times. Since reflectors arrive over the array irregularly, but tend to reflect RFI into the array continuously when their position is appropriate, we expect

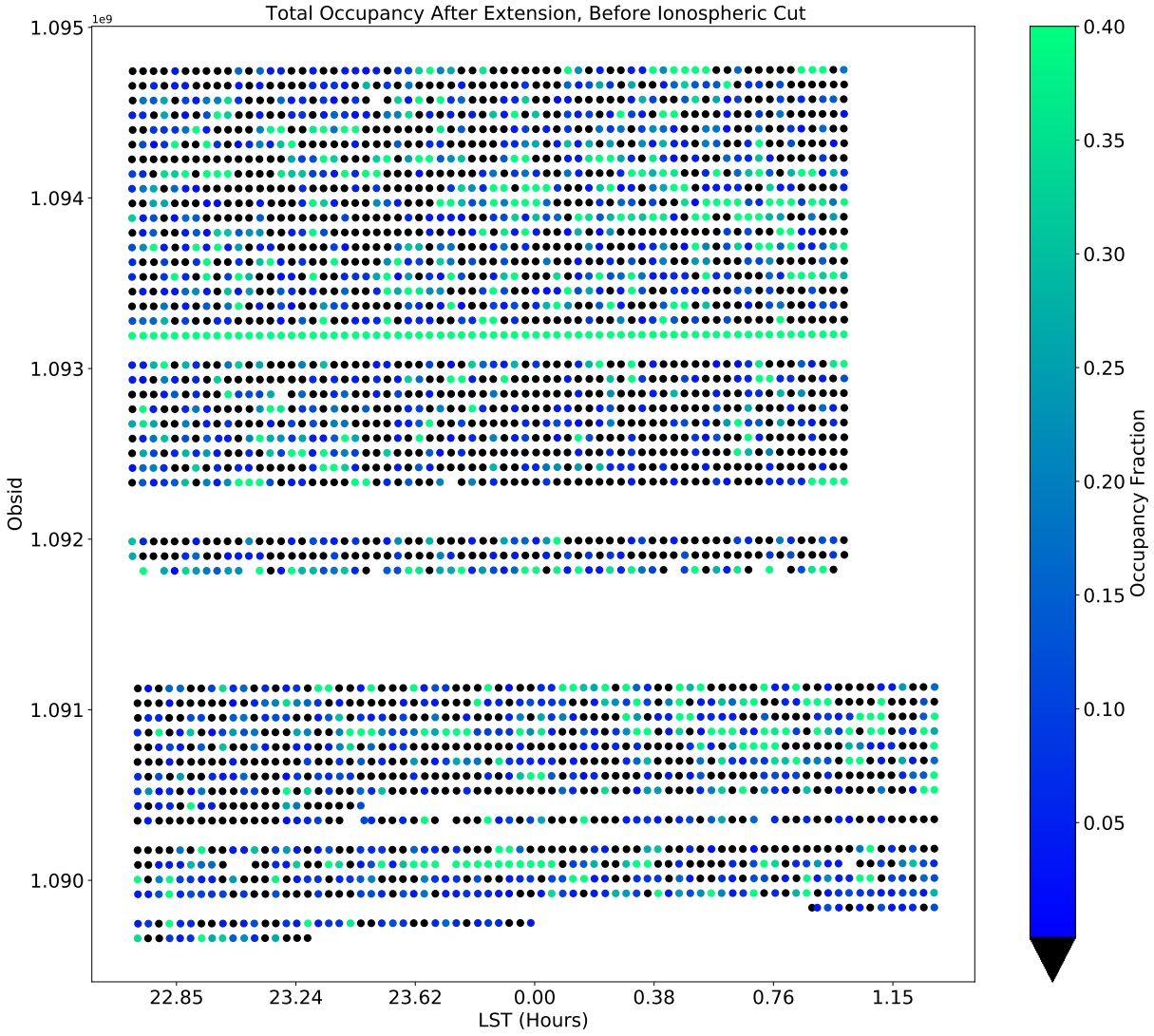


Figure 7.8: Total Occupancy scatter plot after extending flags across frequency, which roughly doubles the total occupancy in the data set. Each marker represents a 2-minute observation, where its horizontal position is its starting sidereal time and its vertical position is its observation ID (starting GPS time). Here, clusters of flags are apparent.

a strong concentration of interarrival times at the sample spacing, and a range of gaps that can potentially be very long depending on air traffic on the given day. In Figure 7.9, over 80% of the mass of this histogram is contained at the sample spacing, with a very powerful tail compared to an exponential distribution with the same mean. This suggests clustered flagging with a range of gap sizes rather than a simple Poisson process.

Another way to probe the clumping properties of the flags is to examine the transition probabilities in the time series. What we find is that unflagged data is almost always followed by more unflagged data with a probability of 98%, while flagged data is followed by flagged data with a probability of 80%. Since the flag occupancy is relatively low ($\sim$ 10%), this manifests as long runs of unflagged data with small-medium sized clumps of flagged data. We find that such clumps can be created randomly using a discrete time Markov chain with consistent transition probabilities as the flagged data. Each night we generate a time series using that night's transition probabilities, and then we histogram the resulting interarrival times. While the realization is noisy, it appears to consistently overpredict interarrivals between 2 and 5 minutes, and consistently underpredict more extreme interarrival times (Figure 7.9).

The RFI flag time series for each night is not entirely consistent with the Markov chain; they differ in their two point statistics. The Markov chain is memoryless, and so any given realization generally does not produce an autocorrelation function with strong peaks away from zero lag. On the other hand, many of the RFI flags exhibit discernible peaks at a range of lags, though some nights with comparatively few flags appear somewhat consistent with the Markov realization. We show examples in Figure 7.10. This suggests that flags from SSINS are not entirely random point processes with only short-range correlations. These additional peaks in the autocorrelation function may occur when reflectors transit different peaks of the primary beam, or may suggest multiple reflectors with some lag between them. The additional peaks also appear to have a characteristic width from night-to-night, possibly indicating a typical clump size related to the speed of reflector transit through the beam.

It is clear from the transition probabilities, interarrival times, and two-point statistics

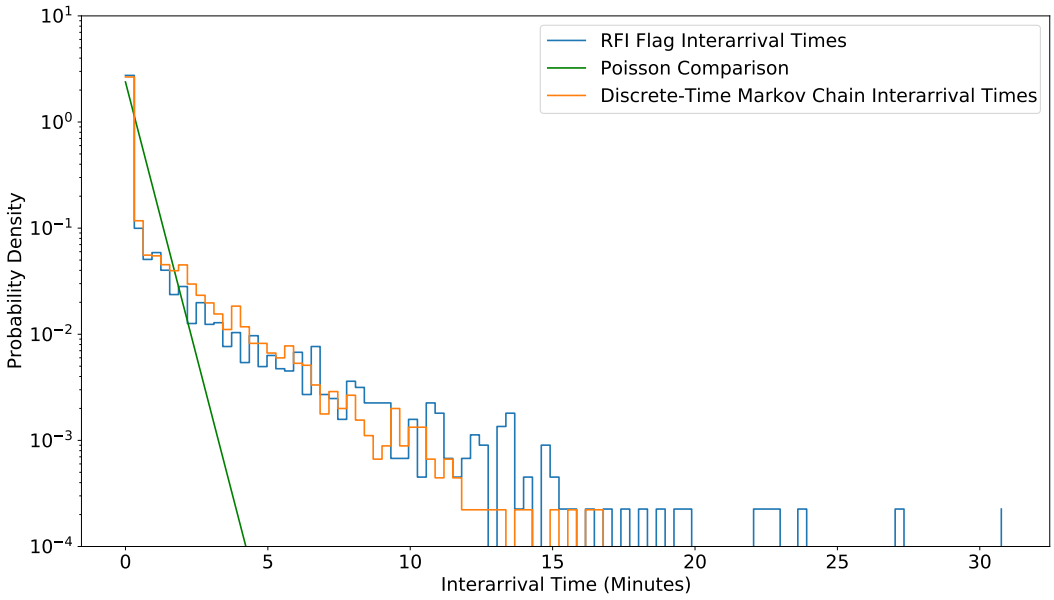
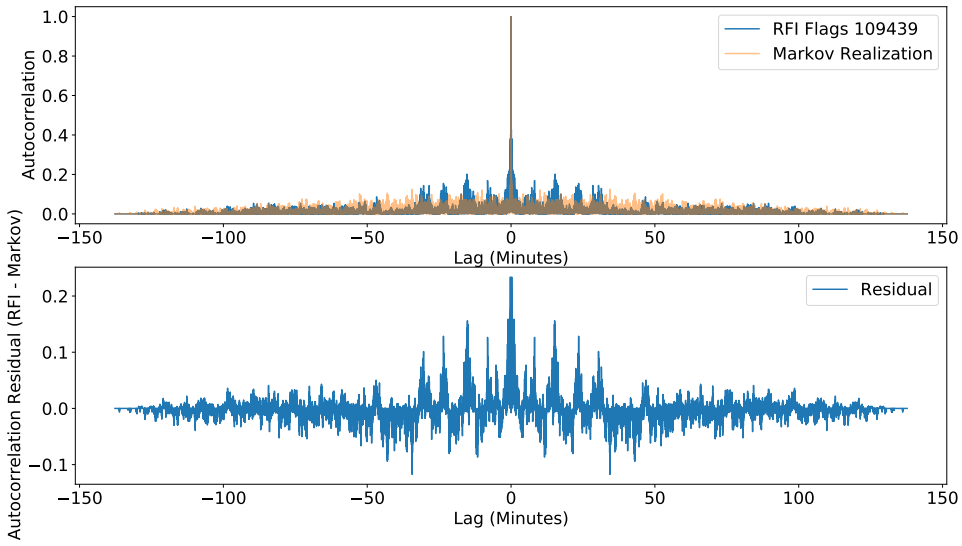
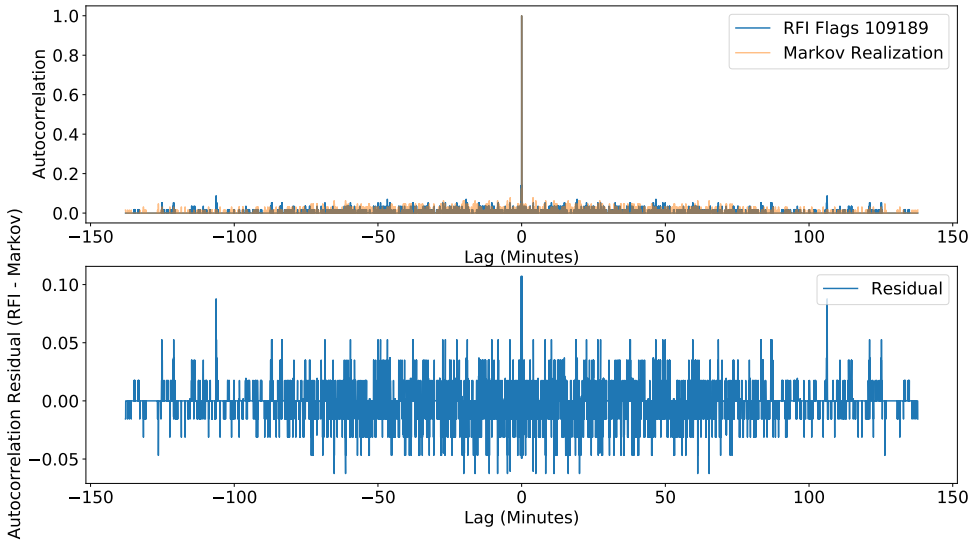


Figure 7.9: Histogram of flag interarrival times over all nights for the broadcast SSINS flags. The "Poisson Equivalent" is an exponential distribution with consistent mean interarrival time as the data. The Markov chain line is built from Markov chain realizations of each night with consistent transition probabilities as the night of RFI flags they are built from. Interarrival times from the Markov chain realizations are roughly consistent with those from the RFI flag time series, though not perfectly so, e.g. overpredictions from 2-5 minutes and underpredictions for times greater than 5 minutes.



(a)



(b)

Figure 7.10: (a) Top Panel: Autocorrelation of RFI flag time series for a particular night (blue) along with the autocorrelation function for a Markov chain realization (transparent orange). Bottom Panel: Residual of the top panel. The Markov chain realization is clearly missing the peaks at lags between 0 and 50 minutes. (b) Same as (a), but for a night with relatively few flags. The autocorrelation functions are more or less consistent except near 0.

of the RFI flag time series that RFI events tend to temporally cluster. This motivates a particular style of jackknife test, where observations with no RFI flags are compared to observations that contained some RFI flags. The logic is that an observation with RFI flags is more likely to contain RFI just beneath the sensitivity of SSINS than observations with no flags at all. In the next section, we develop a series of jackknife tests based on the brightness of the flagged emitters with this physically motivated assumption in mind.

7.3 RFI Brightness Analysis

As shown in Chapter 4, the excess contamination from RFI responds quadratically to its flux. Since SSINS uses incoherently averaged time-differenced visibilities, it is not immediately apparent how to exactly relate the brghtness in the SSINS to the brightness in the visibilities. However, the amplitude of a visibility difference can be no more than twice the brightness of the brighter of the two samples that were differenced. Therefore, we can use the SSINS amplitudes as a proxy for the brightness in the visibilities.

The contaminated samples receive contributions from both thermal noise and RFI. In the SSINS, this usually manifests as a small excess amplitude above the thermal background. When the RFI is much brighter than the thermal noise, we can probe the effect of RFI-contaminated observations using the SSINS amplitudes directly. However, when the thermal noise contribution is larger than the RFI contribution, such as extremely faint RFI left uncaught by SSINS, selecting observations with high SSINS amplitudes can actually select on noise levels rather than RFI contamination. While selecting out observations with higher noise might produce a deeper limit, it does not directly investigate the effect of RFI on power spectra in an observational setting.

In order to probe the effects of extremely faint RFI beneath the sensitivity of SSINS, we need a hierarchy that subtracts out the effects of thermal noise variance across observations. We use the SSINS z-scores for this purpose. Since a mean is subtracted out and divided per frequency, the z-score reflects the excess amplitude that results from the RFI contribution in a dimensionless and statistically well-defined manner. SSINS flags based on the z-scores in the

uncalibrated data. Since the gain between obervations is not always consistent, particularly across a season, we choose to make SSINS from calibrated data and recalculate the z-scores for these jackknife tests. We only use data that was not flagged to estimate the per-frequency mean involved in making the z-scores.

In Figure 7.11, we show histograms of SSINS amplitudes as well as z-scores for the entire season, separated by those samples identified as contaminated (flagged) and those that were not (unflagged). In the top panel we show SSINS amplitudes. Due to the central limit theorem, we expect that the per-frequency distribution of the purely thermal samples in the SSINS follow a Gaussian distribution. However, when all frequencies and many observations are combined in mixture, we observe a highly non-gaussian distribution. In the bottom panel, the unflagged z-scores, which are calculated per frequency, are highly Gaussian, reflecting the statistical assumptions underlying the SSINS pipeline. The unflagged samples have brightness between 33 and 52 Jy. On the other hand, the flagged samples have amplitudes that range from 30 to 450 Jy. The z-scores of the flagged samples exhibit a highly non-Gaussian distribution.

Since there is significant overlap between the flagged and unflagged brightness distributions, hierarchically ranking the observations requires the z-scores. We note that a tiny minority of the recalculated unflagged z-scores are greater than the significance threshold set by the SSINS flagger. Since SSINS is run on the uncalibrated data, the z-scores respond to the gain fluctuations that result from the air conditionning cycle. These are partially corrected during calibration. Calibration solutions are constant over an observation, however, the variations between antennas due to varying amplification are corrected. This changes the relative balance of the waveforms in Figure 7.3(a), which changes how the calibrated version of 7.3(b) looks. This changes the time-dependence of the SSINS amplitudes which can mildly redistribute the z-scores.

Many of the flagged z-scores are far below the significance threshold. This occurs when the sample was flagged as a part of a shape in the match filter. In other words, broader RFI can be found at a lower average flux density compared to narrower shapes due to the

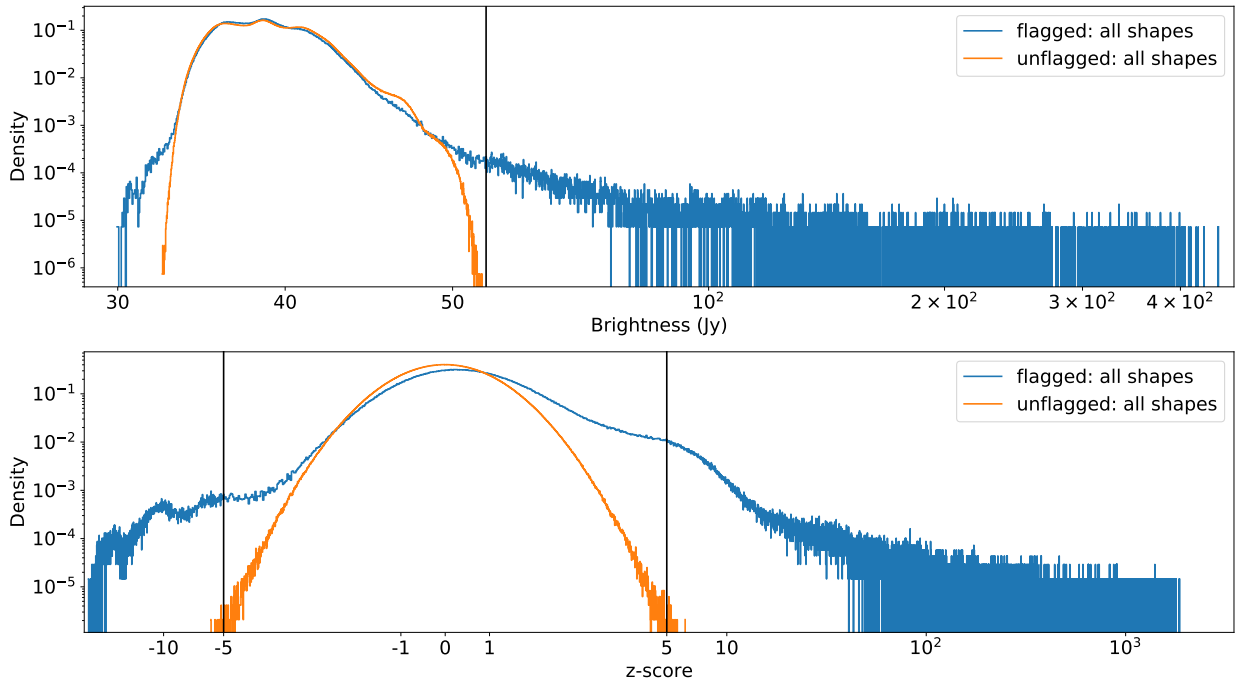


Figure 7.11: Histograms of SSINS amplitudes and z-scores for the entire season, separated by flagged and unflagged data. The horizontal axis is a hybrid linear-logarithmic scale, where the boundary between scales is demarcated by the black vertical lines. The z-scores of the unflagged samples (bottom orange) are highly Gaussian, as expected from the assumptions about the thermal noise. The flagged samples have a highly non-Gaussian z-score distribution. Their brightness distribution has significant overlap with the unflagged brightness distribution, and cannot be separated sheerly by drawing an amplitude cut.

fact that broad clusters of positive (or negative) outliers would exist only with extremely low probability in a purely thermal SSINS with stationary noise. The match filter sums z-scores over frequency bands and finds the samples with maximum absolute deviation, iteratively. All polarizations are flagged if a significant outlier is found in any one polarization. In Figure 7.12, we show the same quantities as in Figure 7.11, but emulating the match filter process by summing over frequency and histogramming the (averaged) samples with maximum absolute deviation over the polarization axis. The thermal background in the bottom panel now takes the form of an extreme value distribution for a Gaussian of sample length equal to 4 (number of polarizations). The recalculated flagged z-scores are now offset from zero. Many still exist below the significance threshold as a consequence of the z-score reshuffling.

We argue that these are still likely to be true positives. While the sample size of a given SSINS is very large when considering single samples, the frequency-summing step in the SSINS match filter reduces the sample size by the number of frequency channels in the observation. Accordingly, this means that substantially fewer extreme values are expected in sub-band sums, and so a significance threshold of 5 gives a wider margin of outliers compared to what is expected from the thermal background. We also find that power spectra from observations with only these low significance values are still correlated with excess window power compared to their unflagged counterparts. We show these in the next section.

7.4 Power Spectrum Jackknife Test Results

The level to which residual RFI contaminates a power spectrum is determined by its apparent brightness and frequency structure. The apparent brightness is determined by the transmitter brightness, the method of propagation, and the relative position of the telescope primary beam. Since transmitters tend to be located close to population centers, we expect that RFI contamination should generally increase when the telescope is pointed more to the West.

To probe these different variables, we divide the contaminated observations into integration subsets separated by RFI shape reported by the match filter (narrowband, streak, or

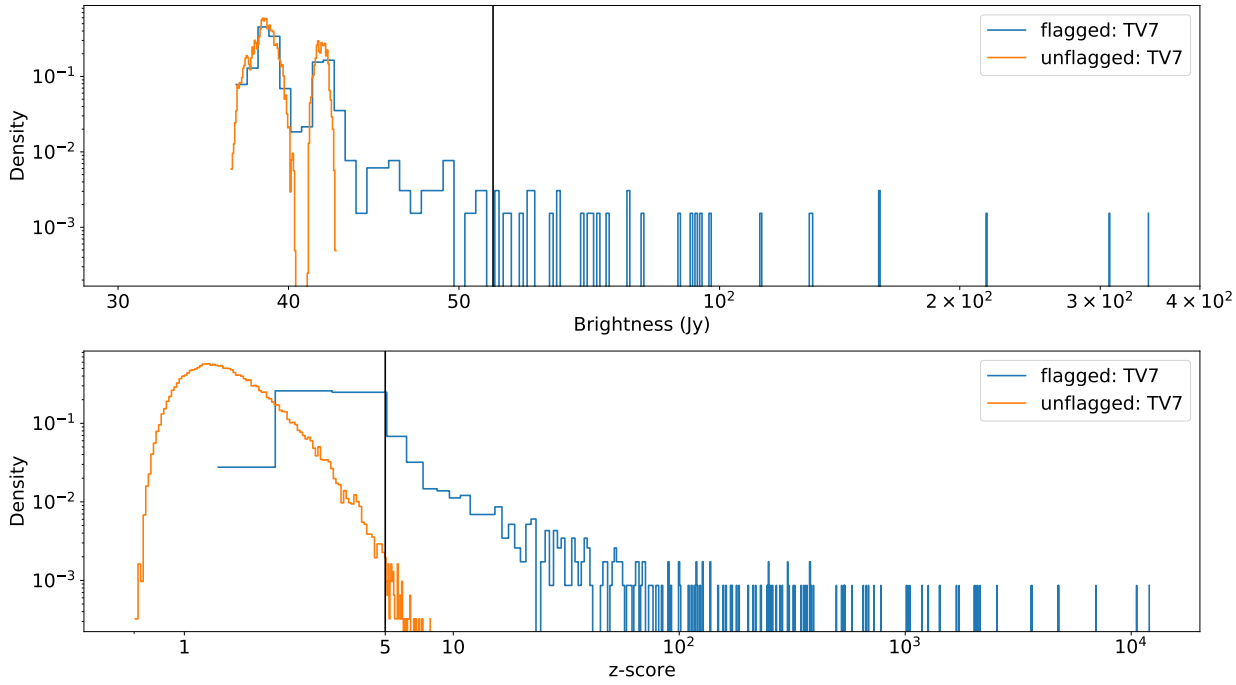


Figure 7.12: Same as Figure 7.11, but with brightnesses and z-scores calculated by doing the SSINS sub-band sum that is employed in the match filter over the TV7 frequencies. Interestingly, the unflagged and flagged brightness samples appear bimodal.

any TV channel), telescope pointing (integers from -2 to +2), and reported z-scores of the RFI events (bins with edges 3, 5, 10, 100, 1000, 10000). For each contaminated observation in a subset, we find a matching observation at a similar sidereal time with no RFI reported by SSINS. We then integrate these subsets, form power spectra, and examine them side by side.

We see a common trend among all RFI types and z-score bins, which is that observations possessing SSINS flags, i.e. observations that were identified as having some contamination, tend to show excess window power compared to observations that were not identified as having any contamination, despite having applied the flags. This trend becomes more no-

ticeable as the array points more westerly. As an example, we show a sequence of pointings for integrations over observations with any RFI type in them whose z-scores were all between 10 and 100 in Figure 7.13. The excess window power suggests that observations identified as being contaminated possess residual RFI uncaught by SSINS that makes a noticeable effect on the power spectrum measurement. In the full set of integrated power spectra, we see the pointing trend occur regardless of RFI type, and in most z-score bands. We expect this physically since there are more transmitter sites and population centers towards the West of the array than towards the East, and so pointing the telescope beam more West will put associated RFI events in a more sensitive location in the beam.

We generally find that the excess power is far more noticeable in the East-West polarization than the North-South. We know from inspection of the SSINS that RFI is often seen as unpolarized, i.e. roughly equal strength in all four instrumental polarizations. In some specific instances it can appear polarized. Usually if it appears polarized, it appears stronger in the East-West dipoles. Occasionally it will appear stronger in the North-South dipoles. The apparent polarization of the RFI source is a combination of its broadcast polarization, some propagation effects, and perhaps most importantly, its geometric location relative to the array.

Since a dipole has no sensitivity along its axis, RFI from the Southern or Northern horizon will appear to the array as if it is East-West polarized, and vice-versa for the Western and Eastern horizons. There are a number of digital television transmitter sites in Western Australia, and therefore a range of possible propagation directions for DTV RFI. The strongest transmitters are in Perth, which is to the South. However, direct reception is extremely unlikely due to the extreme remoteness of the MWA. It is more likely that reflections off of aircraft or other transient phenomena make up the bulk of DTV receptions. For example, the largest population center near the MWA is Perth, and most flights that could reflect DTV to the MWA are probably headed towards there (or perhaps towards Geraldton, which is slightly Southwest of the array). Some of these reflections occur when the aircraft is near its destination, and so the RFI source will appear in the Southern sidelobes, e.g. what we

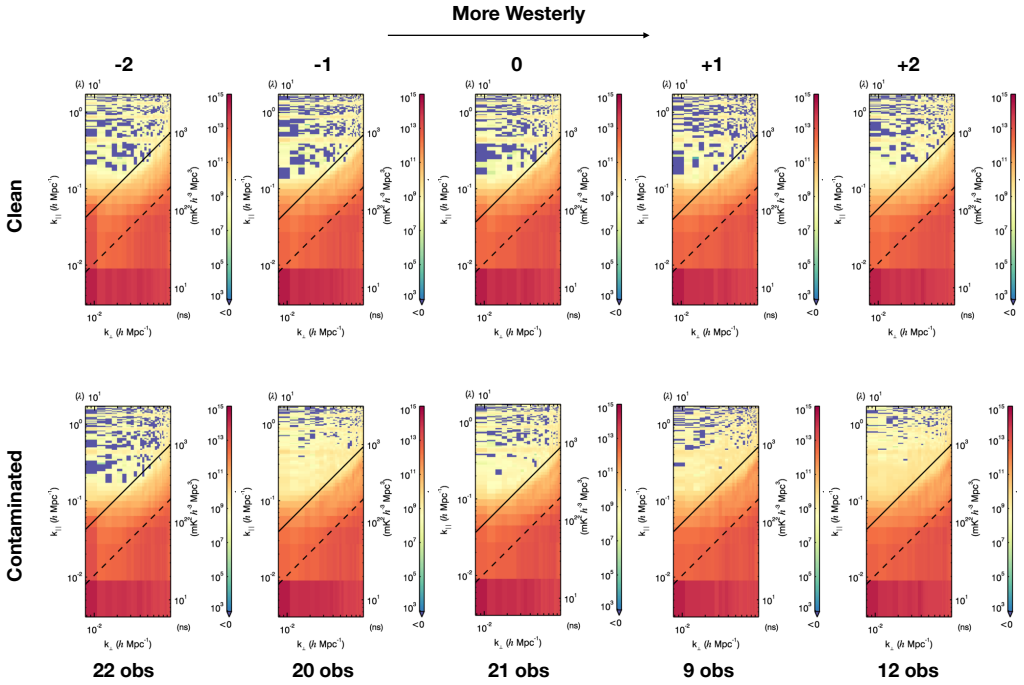


Figure 7.13: Top: East-West polarized power spectra for observations that SSINS identified as clean over the five pointings present in the data. The array points more Westerly towards the right of the figure. The number of observations in each integration is annotated at the bottom of the figure. The depth of integration ranges by about a factor of 2.5. Bottom: Corresponding contaminated power spectra, where any RFI type is included so long as the observation has RFI z-scores that all lie between 10 and 100. We see excess window contamination in the RFI-contamianted spectra, suggesting residual RFI that was uncaught by SSINS. This contamination is generally worse in the later pointings that point more West. There is no clear correlation between integration depth and contamination levels.

saw in Chapter 3. Since this is also where the strongest transmitters are, we expect more noticeable power spectrum contamination in the East-West polarization.

In Figure 7.14, we show power spectra in the +2 pointing, which is the most commonly contaminated pointing in our jackknife tests, separated by the SSINS z-scores of the identified events without discriminating between RFI types. Interestingly, we see that integrations of observations with high SSINS z-scores do not display the characteristic excess power in the window. However, this may be due to the fact that such high z-scores are relatively rare, and that there is significant variation in the noise-levels between different members of the jackknife test. It may be that the noise must be integrated down before an obvious RFI signal appears in the window. We also remark that all integrations with RFI z-scores below 100 have a corresponding "clean" power spectra that demonstrate some excess power in the window, though not as severely as the power spectrum in which SSINS identified RFI. This suggests the presence of false negatives in the SSINS pipeline that can have a significant effect on the power spectrum measurement.

From the theoretical discussion in Chapter 4, we expect that different RFI types might have different power spectrum contamination shapes. In Figure 7.15, we show a jackknife test over RFI shape as determined by SSINS. In general, we find no obvious power spectrum difference among the shapes, contrary to expectation. For instance, we expect narrowband contamination to be approximately constant as a function of $k_{\parallel}$. Even in more dramatic cases with substantially more noticeable residual interference than is shown in Figure 7.15, the narrowband RFI does not produce a noticeable constant pattern. If we compare to the theoretical calculations in Chapter 4, we see that narrowband RFI power spectra at a given flux density is generally lower compared to their DTV counterparts, but also that it is constant in $k_{\parallel}$. Since the noise levels are approximately constant in $k_{\parallel}$, we expect that if this were only due to narrowband contamination, then the entire EoR window would show clear contamination. Due to this inconsistency, we suggest that the residual RFI in this power spectrum is actually not narrow in frequency, but rather something broader that SSINS misses altogether. It could be a relatively stationary source of DTV interference, such as a

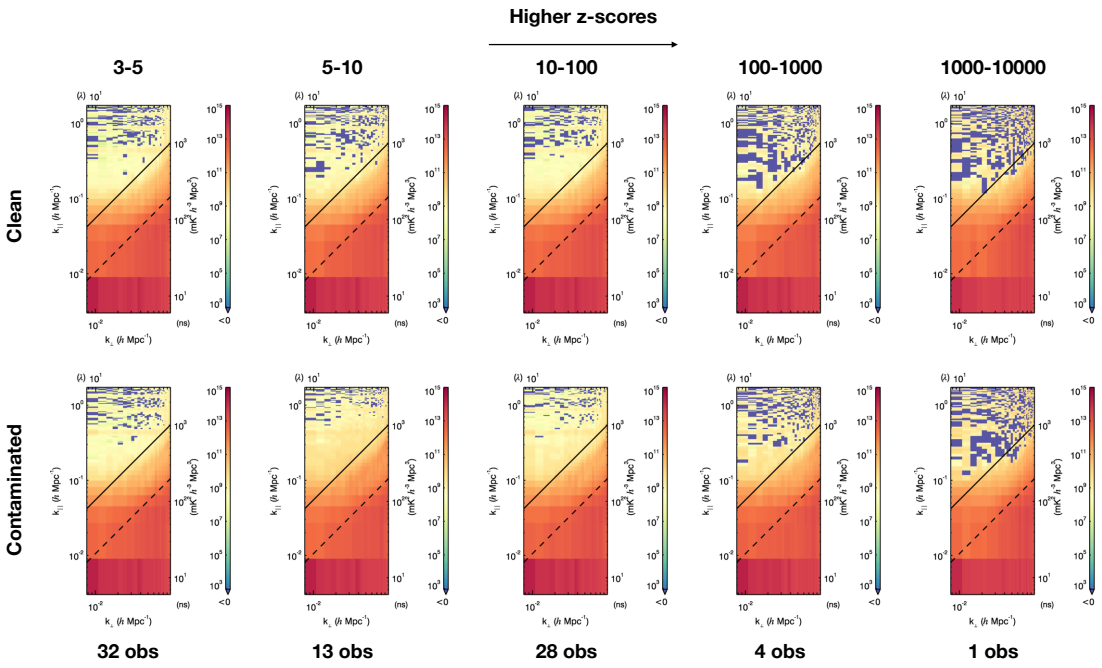


Figure 7.14: Similar power spectrum jackknife test as Figure 7.13, except this time comparing only the +2 (most Westerly) pointing for RFI of any type in the annotated z-score bins at the top of the figure. Excess power is more noticeable for observations in which the identified contaminants had low z-scores. Since the number of contributing observations ranges from 1 to 32, there is significant variation in the noise levels of each integration, which makes excess power from residual RFI harder to discern. We note that the corresponding "clean" integration for the power spectrum whose SSINS z-scores fall between 10 and 100 appears to have moderate excess window contamination as well, through less than its contaminated counterpart.

distant reflector that is coincidentally moving towards the array or nearly so.

Since the streak events in the jackknife test shown in 7.15 showed no obvious excess power, we show a power spectrum for a group of observations in the -1 pointing classified as containing broadband interference flagged by SSINS, along with the LST-matched observations in Figure 7.16. These z-scores were between 10 and 100, indicating that the interference was moderately bright in the SSINS compared to the thermal background. The excess window power matches that for other shapes shown in Figure 7.15. It is nontrivial to theorize about the expected contamination of broadband interference due to the fact that it could come from several different types of sources. For instance, we have evidence that it could be the sidebands of bright ORBCOMM interference, 40-70 MHz away from the central frequency. The ORBCOMM signal would need to be understood in extreme detail since these sidebands are so far from the allocation. We know from Chapter 4 that if it were approximately flat-spectrum over the observing band, that it would give no excess power in the window relative to the foregrounds. Therefore, supposing these signals are indeed responsible for the excess power, there must be some nontrivial structure in the sidebands of these signals. Another possible broadband emitter is lightning. Lightning produces emissions over a range of radiofrequency scales [99]. Some theoretical studies suggest that certain types of lightning events can produce relatively flat emissions over the bands considered in this work [48, 83], although [48] shows some nontrivial structure in this range. In summary, since the particular signal structure behind the broadband events in the MWA SSINS is as of yet undetermined, we find it difficult to hypothesize about the expected shape of power spectrum contamination.

The similarity in contamination between the different shapes could also be due to other factors. First, as noted in §7.2, SSINS is more optimized for flagging than classification, and the occasional misclassification does occur. Second, it is possible that a given reflector is reflecting multiple different transmissions simultaneously or with a short gap between them. This means that any given observation classified as containing some type of interference does not exclude it from containing another type of interference. In other words, there is some

amount of overlap between the subsets shown in Figure 7.15, and this may be responsible for the similarity in contamination shape.

To more directly investigate the effects of applying the SSINS flags, we examine an integration in which residual RFI was present by forming the power spectrum without applying the SSINS-calculated flags and then subtracting the original power spectrum that had flags applied. Note that we are using the same observations in each case, and only choosing whether or not to apply the flags. We show the result in Figure 7.17. While the change in the power spectrum is slight, when we plot the difference we see a clear signature in the window matching the shape of excess power that we have seen so far. We also see that power was consistently removed in the region of the wedge between the dashed and solid lines. The corresponding LST-match of this observation set is shown on the far left of the figure to allow for a comparison in the style of the previous jackknife tests.

The dashed and solid lines in the wedge represent the extent in $k_\parallel$ to which sources at the edge of the primary beam and at the horizon throw power due to the chromatic point-spread-function. The region between these two lines corresponds to the sidelobes. In this example, we chose a handful of observations from the +1 pointing that were identified as containing DTV interference. Considering Figure 3.1, we expect that a DTV source has a smooth component that appears in the power spectrum wedge, and a spectrally sharp component that throws power into the higher $k_\parallel$ modes as in Chapter 4. This power spectrum suggests that the RFI sources are consistently located in our sidelobes, corroborating our imaging experiments.

These results potentially suggest that a new RFI mitigation strategy may be required for a significant EoR detection. SSINS only finds RFI that exists after a time-differencing procedure. The goal of this procedure is to subtract out the coherent sky signal from the flagging statistic. However, It appears that some residual RFI exists in these power spectra that goes uncaught by SSINS. It is possible that this RFI is more like the sky in that it does not subtract out in a 2-second difference. One way to balance this weakness of SSINS is to complement it with the following imaging strategy.

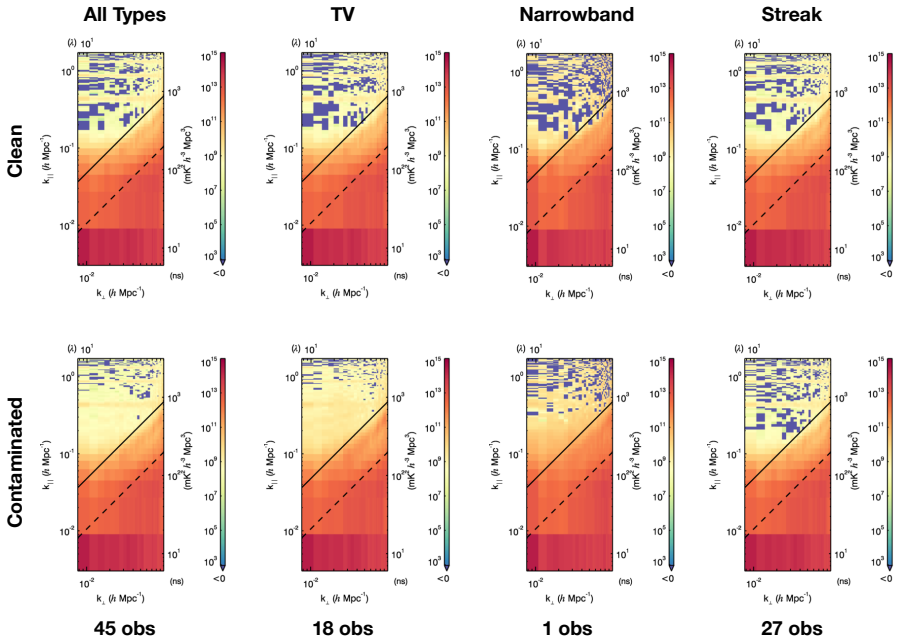


Figure 7.15: A power spectrum jackknife test for zenith-pointed observations, where spectra are separated by RFI type. These all have RFI events with z-scores between 10 and 100. We find that the region of contamination seems unrelated to the type of RFI. This may be due to the fact that multiple RFI types cohabit some of these observations.

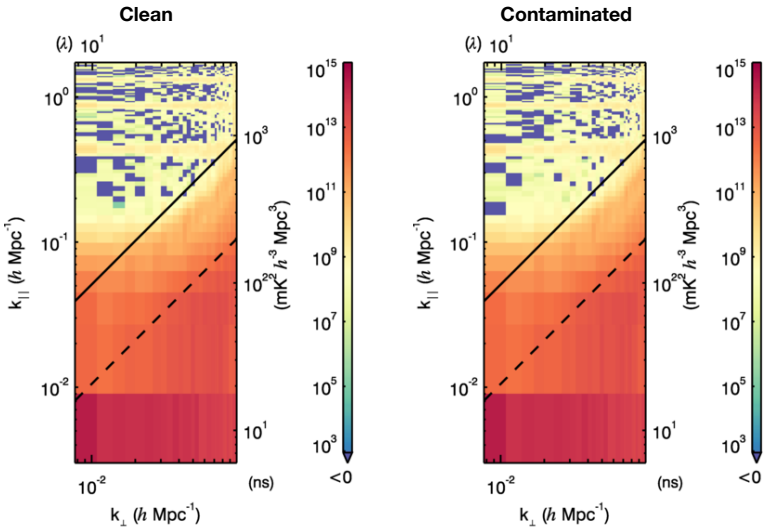


Figure 7.16: Power spectrum jackknife test for observations classified as containing broadband interference. Excess window power takes similar shape as RFI events in Figure 7.15.

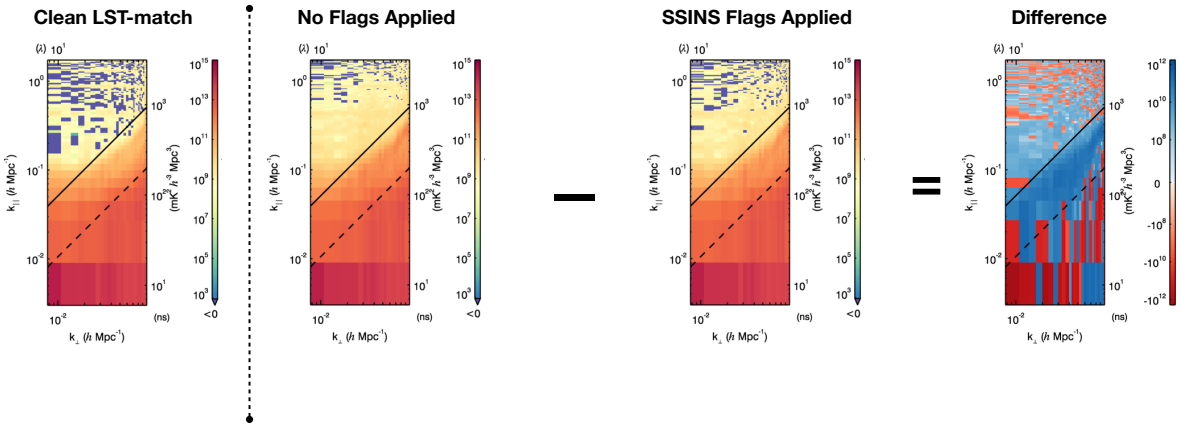


Figure 7.17: A power spectrum test in which two separate power spectra were made for the same observations, once with flags applied and once without flags applied. We find that applying the SSINS flags removes power from the window in the exact shape as seen in other RFI-contaminated power spectra. Additionally, power is removed in the region of the wedge corresponding to the sidelobes, which is where we observe DTV sources in our images. The clean LST-matched observations are shown on the far left as a reference.

While some RFI might be stationary within an observation, most RFI will not be stable in celestial coordinates from observation to observation, particularly if those observations are significantly lagged with respect to civil time. Therefore, one could difference images of appropriate observations and looks for strongly outlying pixels. This could be done at varying integration depths for relatively cheap computational cost within the current pipeline. This would simultaneously allow for RFI identification and localization of the RFI sources, which will be important for developing models of the RFI statistics of the observatory. These models can ultimately be used for determining the depth to which RFI excision needs to be performed. In fact, a similar strategy for RFI identification was implemented in [78], though only at a 2-second integration depth.

7.5 Upper Limits on the 21-cm Epoch of Reionization Power Spectrum Signal

Before extracting an upper limit on the 21-cm EoR power spectrum signal, we perform a final reduction based on the results of the jackknife tests. These final eliminations are based on a qualitative assessment of the EoR window in the power spectrum of various integrated subsets. Since our aim is to make a power spectrum measurement, we perform these cuts conservatively so as to avoid selection bias as much as possible.

First, we remove all observations that were identified as contaminated by SSINS. This is motivated by the fact that a substantial number of power spectra involving only contaminated observations expressed excess window power compared to their clean, LST-matched counterparts. Furthermore, the shape of this excess power matches the shape of the power spectrum difference in the jackknife test shown in Figure 7.17. We term this shape as "the RFI footprint." Finally, the idea that residual faint RFI might temporally neighbour SSINS-identified samples is motivated both by physical considerations and the occupancy study in 7.2. This leaves 591 observations, which is about 18.4 hours of data.

Next, we note that some integrations involving only observations identified as clean by SSINS also display this excess window power, although this is substantially less common than in their counterparts. We start by removing all (supposedly) clean integrations with

fewer than 20 observations that have obvious excess window power. In order to produce the cleanest limit possible, we reintegrate the remaining clean observations, separating them roughly by pointing and day, such that there are between 12 and 20 observations per integration. This integration depth is based on balancing the interference-to-noise ratio with the desire to be as fine-grained as possible in removing observations with this metric. We then remove integrations whose resulting power spectra show obvious excess window contamination. Since we are wary of selection bias, we attempt to be as conservative as possible during this step. We show the power spectra of all integrations that are removed during this stage in Appendix F. This removes a further 119 observations, which we designate as the "Wall of Shame" set, and leaves 472 observations, which we call the "Limit Set." We also note that no observations from the +2 (most westerly) pointing survive this test.

We show deep 2d power spectra from each of the three sets (wall of shame, limit set, and coherent average of the two) calculated over the entire observing band in Figure 7.18. The power spectrum estimated from the East-West dipole data for the Wall of Shame set displays the characteristic RFI footprint. On the other hand, the power spectrum made using only the North-South dipole data appears noise-limited in the window. Thus, we see that even when we deeply integrate observations known to have RFI contamination, the footprint still appears polarized. This can be understood by considering that the brightest DTV transmitters in Western Australia are nearly due South of the array. Since the North-South dipoles are not sensitive to the Southern horizon, we generally expect the RFI flux to be stronger in the East-West dipoles than the North-South ones.

In contrast, the Limit Set displays power spectra that look very similar between the different dipoles. They both appear to have a systematic source of contamination in the lower-left corner of the EoR window. What distinguishes this systematic from the characteristic RFI footprint is that it does not extend to higher $k_\perp$ modes. Furthermore, it is relatively equal in power across the polarizations, which would be peculiar for RFI given the Wall of Shame power spectra. We suggest that this is possibly a different systematic effect that is unaccounted for in this analysis. Note that when we coherently average the Limit Set and

Wall of Shame, we observe the contamination extend to higher $k_\perp$ in the power spectrum from E-W data, but not in the one from N-S data.

As a counterargument, we remark that there are many other DTV transmission sites in Western Australia than the aforementioned ones South of the array, distributed in varying directions. Most of these transmission sites are significantly less powerful than the one in Perth, some by multiple orders of magnitude. Since this is a complex scattering problem, it is difficult to deduce exactly how bright each transmitter should appear. However, some basic scaling arguments suggest that the RFI flux should still be dominated by by the Southern transmitters. Therefore, it is possible that this could be extremely faint RFI from the dimmer transmitters, while the flux from the brighter transmitters was removed by our repeated selections. However, for this to be true, there would need to be an explanation for why the longer baselines (higher $k_\perp$ modes) display less power from these hypothetical RFI events. None of the transmitters are close enough for regular direct reception, and therefore the nature of the scatter would have to be different such that the long baselines do not respond as strongly to the RFI. This is physically unusal, since the scatterers are likely to be the same class of object regardless of where the transmitter is. One possibility is that the closer transmitters tend to scatter closer to the array. If the scatterer is sufficiently close, it might appear less point-like and thus the longer baselines might not respond as strongly to the interference. The truth of the matter remains to be seen.

We report the final limits using the 472-observation set. We also integrate the 119 observation set separately to examine how the residual RFI integrates in the power spectrum. Finally, We make an integration with the entire set of 591 observations to see how this final cut affects the power spectrum. We draw limits for three different redshifts: 6.5, 6.8, and 7.1. These correspond to three different bands, each of which is exactly half the bandwidth of the MWA EoR highband. They are centered 1/4, 1/2, and 3/4 the way through the band. We display the spherical power spectra along with the 2-σ upper limits in Figure 7.19. Our lowest upper limit that we report is $\Delta^2 \leq 1.61 \cdot 10^4$ mK^2 at $k = 0.258$ h Mpc^{-1} using the East-West polarized dipoles at a redshift of 7.1.

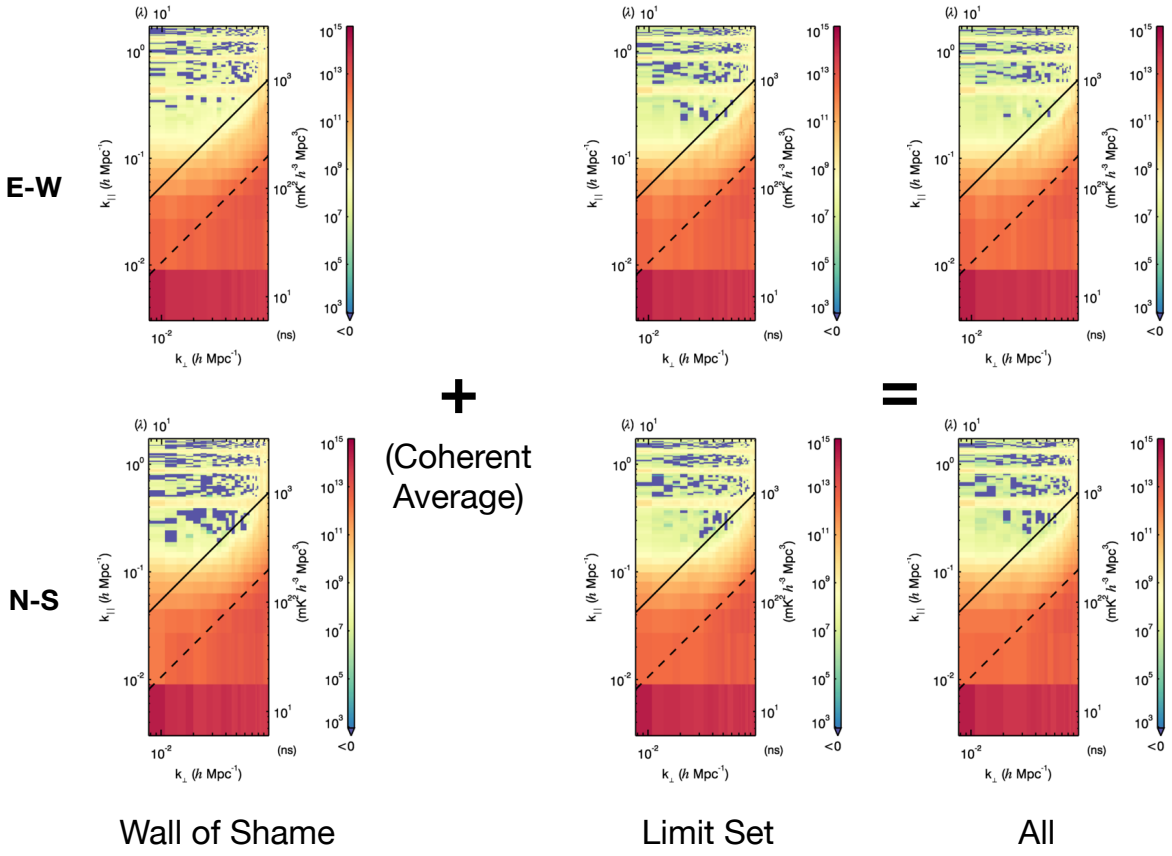


Figure 7.18: Deep 2d power spectra from three different sets of data, all of which were deemed uncontaminated by SSINS. The Wall of Shame set shows a clear RFI footprint in the spectrum made from the E-W dipoles, but appears noise-limited in the one from the N-S dipoles. On the other hand, the Limit Set power spectra appear similarly regardless of which data is used. The coherent average of the two performs to expectation in that adding the Wall of Shame set does not strongly affect the appearance of the power spectrum from the N-S data, and imprints the characteristic RFI shame in the spectrum made from the E-W data.

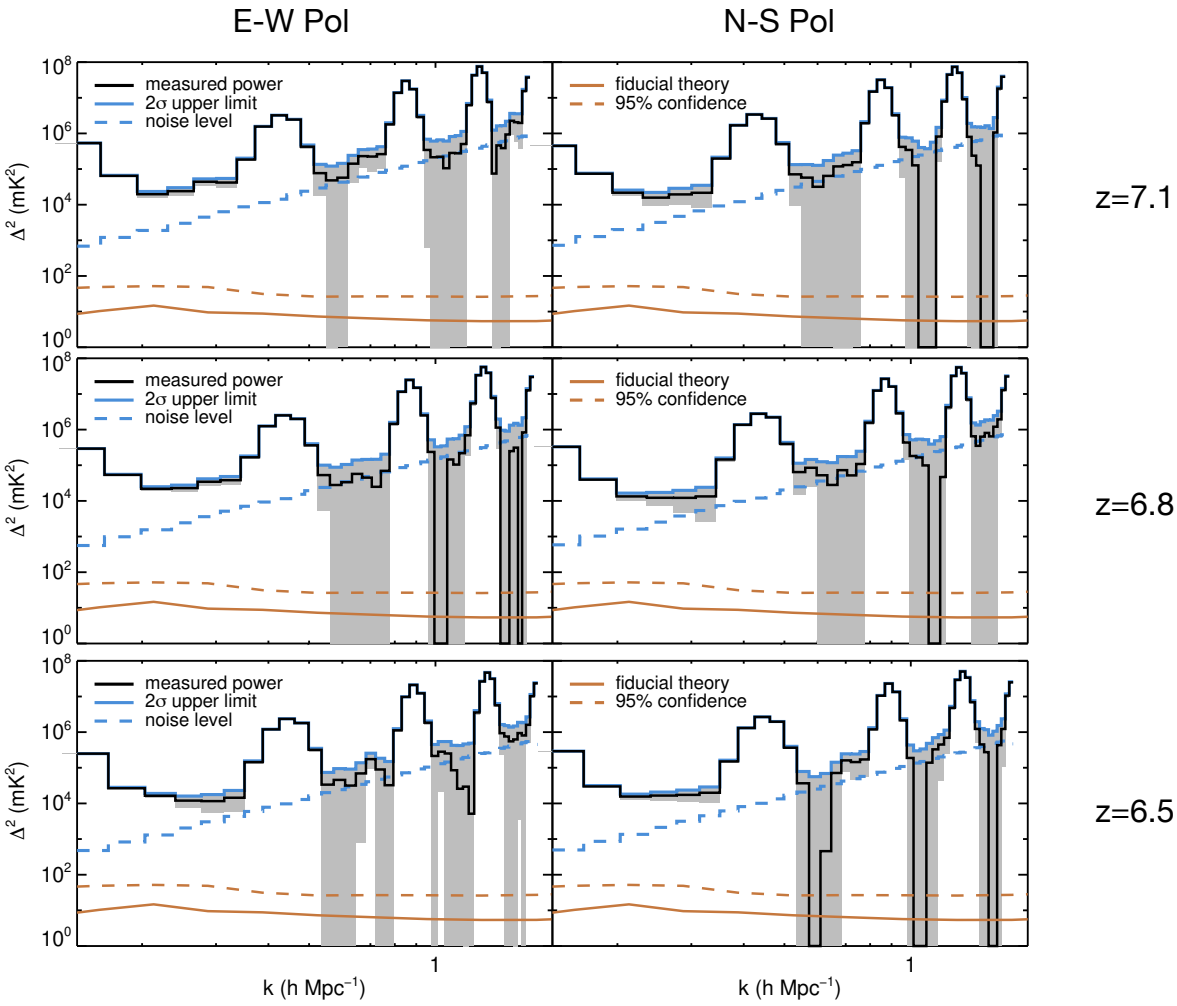


Figure 7.19: Spherical power spectra for the Limit Set, along with $2-\sigma$ upper limits. We notice that the range of wave modes between the first and second coarse band harmonics often appears noise-limited.

We observe that the range of wave modes between the first and second coarse band harmonics is noise-limited in most of the presented spherical power spectra. This suggests that adding more high-quality data would probably improve the limit. Interestingly, if we evaluate the upper limit using the coherent average of the Wall of Shame and Limit Set, we see that it degrades the limit systematically between the first and second coarse band harmonics for redhifts 6.5 and 6.8 in the East-West polarization. The effect is extremely slight, but significantly consistent across the stated bins. The lowest limit that we observed, which we do not report, occurs in this coherent average below the first coarse band harmonic. We do not report that limit since the coherent average was known to be RFI-contaminated before we constructed it.

7.6 Final Conclusions

This work was devoted to understanding the role of radio frequency interference in measurements of the 21-cm Epoch of Reionization power spectrum signal. We attacked the problem from several different angles. We first built a detection method, SSINS, to identify faint RFI that eluded standard RFI flagging measures. We then demonstrated that informing power spectrum integration sets using information from SSINS improves EoR power spectrum limits, suggesting that RFI was an important systematic effect and thereby justifying ultra-faint RFI detection efforts [10].

As a means to verify that the upper limit improvements were indeed related to RFI, we theoretically calculated and simulated the power spectrum of a variety of RFI sources that affect our measurements. We found that even comparatively dim RFI sources can overwhelm the EoR signal, and that it was extremely plausible that residual RFI would produce excess power in the EoR window to the level that removal of faint sources using SSINS would improve the power spectrum upper limit at the depth of current upper limit studies. In other words, we were able to theoretically explain an effect for which we had observational evidence.

While the theoretical study was intended to describe RFI sources, ultimately it works by

characterizing particular types of spectral structure within the data and the effect of that structure on the power spectrum. While RFI may be the most obvious source of spectral structure for which to apply the study, the results can be applied wherever that structure exists, in principle. As an example, we characterized the effect of chromatic flagging on power spectrum measurements in much the same way that we characterized RFI contamination. In this study, we found that chromatic flags produced spectral variations in the data in much the same way as the flagged RFI. This particular effect makes its way into the power spectrum via the subtleties of interferometry. Chromatic flags produce spectral irregularities in the array sampling function, which allow the RFI flags to couple to the foregrounds in such a way as to produce meaningful excess power in the EoR window despite that they typically produce only tiny variations in the array sampling function. Moreover, the functional form of the excess power in the window closely followed what we expect from RFI sources of the same spectral shape as the flags, thus demonstrating the point.

Armed with the understanding of how RFI ought to affect a power spectrum measurement, we studied a large data set taken by the MWA during the 2014 observing season. Using SSINS, we examined the RFI statistics of this large data set, both in terms of its temporal structure as well as its brightness within the data. We found that RFI flags from SSINS tend to cluster in time, and show temporal correlations in some instances. This is physically reasonable for aircraft reflection events in which aircraft move at consistent speeds through the beam in relatively consistent locations, though perhaps in somewhat inconsistent intervals. We observe RFI over a range of brightness, and much of the RFI that we flag is comparable and oftentimes much less than the average brightness of the thermal noise. Based on these statistics, we estimate that the faintest event detectable by SSINS can still make a meaningful impact on the power spectrum measurement, suggesting that even better RFI mitigation strategies might be necessary.

We verified this claim by performing an analysis of the power spectra of carefully selected subsets of the 2014 data set. This produced powerful evidence that our understanding of RFI is accurate. By comparing power spectrum integrations involving observations deemed

uncontaminated by SSINS to LST-matched observations that we identified as contaminated by SSINS (and had the flags applied), we found that there was extremely convincing evidence for residual RFI contamination that was uncaught by SSINS. This is consistent with the notion that RFI events temporally cluster, and suggests that when constructing limits, one may need to excise entire observations where SSINS finds any contamination.

Building off of this analysis, we constructed an upper limit on the 21-cm EoR power spectrum signal using this data set. We resolved to exclude all observations that had any contamination identified by SSINS. We also identified a range of observations that were not identified by SSINS as containing RFI that showed contamination in the EoR window in the same manner as RFI. We chose to exclude only those integrations where this effect was considered obvious so as to minimize the effects of selection bias. We then made power spectra over a selection of redshifts, and found our lowest reported limit to be $\Delta^2 \leq 1.61 \cdot 10^4$ mK^2 at $k = 0.258$ h Mpc^{-1}. By cross-examining deep power spectra constructed from this limit set and an additional set known to contain RFI contamination, we found that RFI beneath the sensitivity of SSINS can consistently degrade the upper limit for bins between the first and second coarse band harmonics in our power spectra. The effect is small, and we see additional systematic contamination in the EoR window that does not clearly match what we have come to expect from RFI contamination. This suggests that, after SSINS, RFI may no longer be the dominant systematic effect on our power spectrum measurement.

Appendix A

COMPUTATIONAL COMPLEXITY OF SSINS

In this section we theoretically determine the computational complexity of the SSINS iterative frequency-matched flagging algorithm given a dictionary of shapes. We simply calculate the minimum number of operations necessary and add their associated times. This therefore reflects the minimum amount of computation time, which will not be satisfied due to implementation details. This will at least tell us how various sections of the algorithm scale relative to one another, and how different shape dictionaries scale relative to one another.

A.1 The General Complexity

The total time is the sum of the times per iteration, indexed with j. Each iteration can be broken into a step where z-scores are calculated per shape, a search step where the maximum z-score is found, and a final step to recalculate the z-scores. We will denote the non-negligible times in the order of the list above with t_z, t_s, and $t_{r,j}$. Times without the subscript j are independent of iteration, while $t_{r,j}$ depends on the results of the max search. Summarizing, the total time, T, is

$$T = N_i(t_z + t_s) + \sum_{j=1}^{N_i} t_{r,j} \tag{A.1}$$

where N_i is the total number of iterations.[1]

First we calculate t_z. Each shape, B, has a number of frequencies associated with it, N_B. Given the initial z-scores for the iteration, the z-score for each shape at each time and polarization, $Z_B(t_n, p)$ of equation 3.15, requires N_B multiplies from the bit mask for pre-existing flags, $N_B - 1$ additions to sum the single-sample z-scores, $N_B - 1$ additions to

[1]The loop always terminates since in the event that only one time remains for a given channel, the z-score for that sample will be 0, which is less than the significance threshold.

sum the bit mask (producing the effective N_B factor in equation 3.15), a reciprocal square root operation, and a final multiplication operation. Denoting the times for addition, multiplication, and reciprocal square root with t_a, t_m, and t_q, respectively, we can write the total time to produce the all the z-scores:

$$t_z = (N_t - 1)N_p \sum_{B=1}^{K} 2t_a(N_B - 1) + t_m(N_B + 1) + t_q, \tag{A.2}$$

where K is the total number of shapes, N_t is the total number of integrations in the original data set, and N_p is the total number of polarizations. Rearranging terms, one sees

$$t_z = (N_t - 1)N_p \left(K(t_q - 2t_a + t_m) + (2t_a + t_m) \sum_{B=1}^{K} N_B \right). \tag{A.3}$$

Since each shape has at least one frequency associated with it, and broader shapes such as DTV usually require many frequencies ($\sim 10^2$ for DTV in MWA and HERA), the sum on the far-right is at least as large as K but can be much larger. If no broad shapes are used and each channel is searched individually, then $K = N_f$, the total number of frequencies, and $N_B = 1$ for all B. This leads to

$$t_z^{(\text{single})} = (N_t - 1)N_p N_f (2t_m + t_q) \tag{A.4}$$

In other words, the basic flagging algorithm without any additional shapes is linear in the total data volume. No additions are necessary since only single-samples are used. While in principle one can achieve a lower total computational complexity with a coarser shape dictionary that ignores single-frequency shapes and still covers the whole band, this runs the risk of diluting sensitivity to truly narrow RFI which is a common type. With this constraint, equation A.4 reflects the minimum computational complexity of the sub-band z-score calculation.

In tests with the MWA and HERA, usually there are a few broad shapes worth testing along with the narrowband contaminants. In these cases, the sum in equation A.3 is $\approx 3N_f$, while $K \approx N_f$, meaning that the sum dominates the compute cost for this step. In general,

the total "volume" of the shape dictionary (how many frequencies it commands to check, including overlapping shapes, calculated by this sum) usually drives the total compute cost of the SSINS algorithm, which we demonstrate by comparing to the terms calculated below.

Since the shapes and data are in no particular order, the max search must check every element resulting from the calculation of $Z_B(t_n, p)$ once using a comparison operation. We denote the time for a comparison operation with t_c. There is one comparison per shape, time, and polarization. There is also one final comparison operation to see if the max is greater than the significance threshold. We have

$$t_s = \big((N_t - 1)N_p K + 1\big)t_c. \tag{A.5}$$

Finally we compute $t_{r,j}$, which depends on the result of the max search. The only z-scores that need to be recomputed are those belonging in channels of the shape that was maximally outlying. For each channel and polarization, there are $(N_t - 1)$ multiplies for the bit mask, $(N_t - 2)$ additions to compute the sum, another $(N_t - 2)$ additions to compute the total number of contributing samples using the bit mask, and a division by this number to get the mean estimate. This mean needs to be subtracted from the $(N_t - 1)$ times (applying the mask as appropriate). A single multiplication gets the standard deviation estimate, and a division per time finally produces the z-scores. Denoting division time with t_d, we have

$$t_{r,j} = N_{B,j} N_p (N_t(t_m + t_d) + (3N_t - 5)t_a). \tag{A.6}$$

Substituting the individual terms into equation A.1, we have

$$T = N_i \Bigg\{ N_p(N_t - 1)\Bigg(K(t_q - 2t_a + t_m) + (2t_a + t_m)\sum_{B=1}^{K} N_B \Bigg) + \Bigg(N_p(N_t - 1)K + 1 \Bigg) t_c \Bigg\} + N_p \Bigg(N_t(t_m + t_d) + (3N_t - 5)t_a \Bigg) \sum_{j=1}^{N_i} N_{B,j}. \tag{A.7}$$

This expression is needlessly exact. The following statements allow for a useful approximation. First, SSINS requires that the number of integrations in the data is much greater than a small integer (see §7.2.2). Next, the total data volume is always much greater than

Operation	Latency (clock cycles)
FADD	3
FMUL	5
FDIV	14-16
FCOM(P) FUCOM (comparison)	3
FSQRT	14-21

Table A.1: Latencies for floating point operations used by the SSINS algorithm, in terms of clock cycles, for the Intel "Skylake" architecture.

1. Finally, floating point operations take different amounts of times, which can be both hardware and software dependent. We list the latency in terms of clock cycles for the Intel "Skylake" architecture[2] in Table A.1, which reflect the minimum time for these operations to be performed. Assuming other implementation details affecting the compute time affect all operations equally, we can use these figures and approximations to understand the minimum scaling of the algorithm in terms of the clock cycle time, τ. We will assume that a reciprocal square root is a square root followed by a divide, which is not true for all architectures. Combining all this, using median values where appropriate, yields:

$$T \approx \tau N_p N_t N_i \left\{ 35K + 11 \sum_{B=1}^{K} N_B + 29 \langle N_{B,j} \rangle \right\}, \tag{A.8}$$

where the brackets indicate an average. In practice, the total volume of the shape dictionary generally outweighs the volume of any given shape by a factor of 2-3 since usually 2-3 shapes overlap in a given sub-band, viz. the narrow, broadband streaks, and possible specific allocated shapes of chapter 3. In this case, when broadband streaks are the dominant occupant, t_r and t_z are comparable. Conversely, if narrowband occupants are the dominant contaminant, then t_z strongly dominates t_r. In either case, the scaling of the algorithm is

[2]These are reported measurements from `https://www.agner.org/optimize/instruction_tables.pdf`

roughly linear in the number of frequencies when shapes are chosen by eye.

In the next section we will examine t_z specifically under three shape dictionaries that cover the entire band with different generalized shapes, rather than physically motivated ones. Each dictionary has extremely high volume compared to the total number of frequencies in the SSINS, meaning the t_z term will totally dominate the compute time.

A.2 Unknown Shape Hunting

The current method used to determine which shapes belong in the match filter relies on a combination of knowledge about the RFI environment along with an exhaustive hand-graded examination of a sizable data set. In the event that there is truly too much data for an exhaustive search to be feasible, one might consider shape dictionaries that essentially allow SSINS to hunt for unknown shapes.

One way to automatically search for shapes is simply to put every possible shape into the match filter. If shapes are allowed to be discontiguous, for instance if two RFI signals (or other suspicious instrumental effects) in different subbands are always broadcast with one another, then every combination of frequencies ought to be sought for. The total number of shapes of length k for a full band of N_f frequencies is given by $\binom{N_f}{k}$, and so the total number of shapes in this dictionary is

$$\sum_{k=1}^{N_f} \binom{N_f}{k} = 2^{N_f} - 1, \tag{A.9}$$

which can be recognized by considering the expansion of $(x+y)^{N_f}$ for $x = y = 1$. The total volume of the shape dictionary for this scheme is

$$\sum_{k=1}^{N_f} k\binom{N_f}{k} = N_f 2^{N_f - 1} \tag{A.10}$$

which can be shown by reindexing k from 0 and using the equation A.9. Exponential scaling is extremely undesirable given that modern radio interferometers often operate on hundreds to thousands of channels at a time.

On the other hand, if we search only for all the contiguous shapes, then the number of shapes of length $N_f - k + 1$ is simply k, so the total number of shapes is

$$\sum_{k=1}^{N_f} k = \frac{N_f(N_f+1)}{2} \tag{A.11}$$

while the total volume is equal to

$$\sum_{k=1}^{N_f} k(N_f - k + 1) = \frac{N_f(N_f+1)(N_f+2)}{6}, \tag{A.12}$$

In other words this makes the compute time cubic in the number of frequencies. For HERA, this brings the compute outside of the real time regime, i.e. M minutes of data takes more than M minutes to flag.

Yet another step down would be to repeatedly divide the band in powers of 2, or whatever other integers exist in the prime factorization of N_f, and search for all contiguous shapes of those lengths. If N_f is a power of 2, then the total number of shapes is given by

$$\sum_{k=0}^{\log_2 N_f} 2^k = 2N_f - 1, \tag{A.13}$$

while the total volume is given by

$$\sum_{k=0}^{\log_2 N_f} 2^k \left(\frac{N_f}{2^k}\right) = N_f(\log_2 N_f + 1) \tag{A.14}$$

which is a much more desirable scaling for most applications.

Such a shape scheme might be desirable for an instrument such as HERA, where the instantaneous bandwidth covers many different types of allocations. Analysis of the metadata from this strategy might also help the analyst discover shapes they were unaware of. One could even imagine programming a learning algorithm using this metadata as an input. However, generally speaking, this is still roughly an order of magnitude slower than the basic hand-made shape dictionary for the MWA and HERA, and the boost in performance, if any, is yet to be tested thoroughly.

Appendix B

SSINS AS A BAYESIAN MAXIMUM A POSTERIORI DECISION RULE

A maximum a posteriori (MAP) decision rule is one that acts based on the outcome with the highest posterior probability. In RFI flagging, one makes a binary decision about whether a sample is likely to be contaminated. A MAP rule for flagging would flag any sample that is more likely to be contaminated than uncontaminated. In this section, we will show that the SSINS flagging algorithm is equivalent to a MAP decision rule for detecting square pulses in the presence of additive white Gaussian noise with a particular prior.

For this demonstration we follow [49]. Suppose we are interested in receiving a signal, denoted by a vector $\mathbf{x}_s$, over a channel of some discrete length, N. Suppose also that the channel is affected by additive white Gaussian noise of mean equal to zero and standard deviation, 1. Then, the probability of receiving some measurement, $\mathbf{y}$, given that the signal was transmitted (say, $s = 1$, the first signal of interest) is

$$P(\mathbf{y}|s = 1) = (2\pi)^{-N/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{x}_1)^T(\mathbf{y} - \mathbf{x}_1)\right). \tag{B.1}$$

Similarly, the probability of receiving $\mathbf{y}$ given that no signal was transmitted ($s = 0$) is

$$P(\mathbf{y}|s = 0) = (2\pi)^{-N/2} \exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{y}\right). \tag{B.2}$$

We can determine the probability of the presence of a signal, given $\mathbf{y}$, using Bayes' rule:

$$P(s = s'|\mathbf{y}) = \frac{P(\mathbf{y}|s = s')P(s = s')}{P(\mathbf{y})} \tag{B.3}$$

where $P(s = s')$ and $P(\mathbf{y})$ reflect our prior assumptions about the probabilities of receiving the signal s' and $\mathbf{y}$. A MAP rule to decide whether the first signal of interest was received or

no signal of interest was received can be formulated in terms of the ratio of the two posterior probabilities reflecting these scenarios:

$$\frac{P(s=1|\mathbf{y})}{P(s=0|\mathbf{y})} = \frac{P(\mathbf{y}|s=1)P(s=1)}{P(\mathbf{y}|s=0)P(s=0)}. \tag{B.4}$$

If this ratio is greater than 1, then it is more likely that a signal was received than not, and vice-versa if the ratio is less than 1. Using the explicit probabilities mentioned above, we have

$$\frac{P(s=1|\mathbf{y})}{P(s=0|\mathbf{y})} = \exp\left(\mathbf{y}^T\mathbf{x}_1 - \frac{1}{2}\mathbf{x}_1^T\mathbf{x}_1 + \ln\frac{P(s=1)}{P(s=0)}\right) \tag{B.5}$$

This ratio is greater than 1 when

$$\mathbf{y}^T\mathbf{x}_1 > \frac{1}{2}\mathbf{x}_1^T\mathbf{x}_1 - \ln\frac{P(s=1)}{P(s=0)}. \tag{B.6}$$

Now suppose that the signal $\mathbf{x}_1$ is a square pulse that is 0 except for some number of elements M_1, and for those nonzero elements it takes the value $1/\sqrt{M_1}$. Substituting this into the expression above yields

$$\frac{1}{\sqrt{M_1}}\sum_{j=l_1}^{u_1} y_j > \frac{1}{2} - \ln\frac{P(s=1)}{P(s=0)}, \tag{B.7}$$

where l_1 and u_1 are the indices bounding the nonzero elements of $\mathbf{x}_1$ and y_j is the jth component of $\mathbf{y}$. With the appropriate relabelings, this is identical to asking if the right-hand-side of equation 3.15 is greater than a significance threshold, τ_{01}, set by

$$\tau_{01} = \frac{1}{2} - \ln\frac{P(s=1)}{P(s=0)}, \tag{B.8}$$

which is exactly the formula for the SSINS decision rule regarding whether a particular sub-band in the SSINS is contaminated with RFI. However, encoded in this formula is an assumption about how likely a particular signal is to be received compared to not receiving it.

To discriminate between different possible received signals in this Bayesian framework, say $s=1$ and $s=2$, one would evaluate

$$\frac{P(s=1|\mathbf{y})}{P(s=2|\mathbf{y})} = \frac{P(s=1|\mathbf{y})}{P(s=0|\mathbf{y})}\frac{P(s=0|\mathbf{y})}{P(s=2|\mathbf{y})}. \tag{B.9}$$

SSINS discriminates between potential signals by asking which measurement was least likely to arise from pure noise by asking what the probability of a given sample arising from a standard normal distribution would be. The ratio above asks the exact same question in the sense that if the degree to which $P(s = 1|\mathbf{y})$ is greater than $P(s = 0|\mathbf{y})$ is higher than the degree to which $P(s = 2|\mathbf{y})$ is greater than $P(s = 0|\mathbf{y})$, then $P(s = 1|\mathbf{y}) > P(s = 2|\mathbf{y})$, and vice-versa. The advantage to the Bayesian framework is that it demands explicit elucidation about the assumptions of the problem. Evaluating equation B.9 explicitly,

$$\frac{P(s = 1|\mathbf{y})}{P(s = 2|\mathbf{y})} = \exp\left(\mathbf{y}^T(\mathbf{x}_1 - \mathbf{x}_2) - (\tau_{01} - \tau_{02})\right) \tag{B.10}$$

where τ_{02} has been defined similarly as τ_{01}. In the basic SSINS framework of chapter 3, $\tau_{01} = \tau_{02}$, which essentially expresses a uniform prior over RFI types. If this ratio is greater than 1, then $s = 1$ is more likely than $s = 2$, given $\mathbf{y}$. Proceeding as before, with the uniform prior just mentioned, this occurs when

$$\frac{1}{\sqrt{M_1}} \sum_{j=l_1}^{u_1} y_j > \frac{1}{\sqrt{M_2}} \sum_{k=l_2}^{u_2} y_k \tag{B.11}$$

with M_2, l_2, and u_2 defined similarly as their counterparts. This is precisely the evaluation that SSINS makes during its frequency-matched flagging procedure. In the event that $\tau_{01} \neq \tau_{02}$, then an additional term with the difference between them is required, which would reflect that one RFI type is expected more often than another. This was a heuristic technique we used in chapter 6 in order to reduce high false positive rates through varying noise levels (§7.2.2).

Within a single SSINS flagging iteration, there are several elements missing in the basic framework. First, the question of whether more than one signal was received simultaneously is not addressed, even though it does occur. Addressing this within one iteration only adds compute, since when two significant signals are received simultaneously, the signal of lower significance is always recovered in a later iteration than the one of higher significance. Next, the signals we are receiving are not necessarily normalized like the $\mathbf{x}_s$ of this framework, but instead have variable strength. Additionally, the z-score estimate for a given time and

band is corrupted by whatever else may have happened at other times. Thus, the idea that a non-signal should produce a signal drawn from a white standard normal distribution is not strictly obeyed if any RFI signal is received. In principle, this could be handled by relaxing the assumptions about the additive noise term, however this complicates the math significantly, and this is just another problem that is generally fixed with iteration so long as the RFI does not take up the majority of the observation. We recognize that persistent RFI that corrupts an entire observation is a reality that falls outside SSINS' preferred jurisdiction, and we strive to improve SSINS so that it can handle such circumstances. Efforts are underway to use the frequency statistics instead of time statistics to improve the z-scores, which will of course have the complementary problem with broadband RFI. We hope that by combining both approaches, we will have an extremely robust RFI flagger that will help produce clean and deep upper limits on the Epoch of Reionization power spectrum measurements.

Appendix C

NARROWBAND POINT SOURCE ENSEMBLES

The relative ease of power spectrum calculation for narrowband point sources allows one to derive some useful facts. The first is that the power spectrum of any ensemble is bounded from above by the coherent addition of their powers. The second is that a random ensemble of narrowband sources, uniformly distributed over the sky, is expected to sum powers incoherently. The incoherent sum of powers serves as an expected lower bound on the power of the ensemble. The power spectrum of an ensemble of N such sources is given by

$$\begin{aligned} P_{\text{ensemble}} &= \frac{1}{V_{\mathcal{M}}}\Psi(\nu_0)^2\kappa(\nu_0)^2 r_{\parallel}(\nu_0)^4\beta^2 \\ &\times \sum_{s=1}^{N}\sum_{s'=1}^{N} I_s^{\text{app}} I_{s'}^{\text{app}} \mathrm{e}^{-\mathrm{i}r_{\parallel}(\nu_0)\mathbf{k}_{\perp}\cdot(\boldsymbol{\theta_s}-\boldsymbol{\theta_{s'}})}. \end{aligned} \tag{C.1}$$

The double summation can be split into "diagonal" terms, where $s = s'$, and "cross" terms, where $s \neq s'$. Taking advantage of the symmetry in the indices, we can write this splitting in the following way:

$$\begin{aligned} P_{\text{ensemble}} &= \frac{1}{V_{\mathcal{M}}}\Psi(\nu_0)^2\kappa(\nu_0)^2 r_{\parallel}(\nu_0)^4\beta^2 \\ &\times \left(\sum_{s=1}^{N}[I_s^{\text{app}}]^2 + 2\sum_{s=1}^{N}\sum_{s'>s}^{N}[I_s^{\text{app}} I_{s'}^{\text{app}} \cos\left(r_{\parallel}(\nu_0)\mathbf{k}_{\perp}\cdot(\boldsymbol{\theta_s}-\boldsymbol{\theta_{s'}})\right)] \right). \end{aligned} \tag{C.2}$$

Since cosine is less than or equal to 1 everywhere, the term above in large parentheses satisfies the inequality

$$\begin{aligned} &\sum_{s=1}^{N}[I_s^{\text{app}}]^2 + 2\sum_{s=1}^{N}\sum_{s'>s}^{N}[I_s^{\text{app}} I_{s'}^{\text{app}} \cos\left(r_{\parallel}(\nu_0)\mathbf{k}_{\perp}\cdot(\boldsymbol{\theta_s}-\boldsymbol{\theta_{s'}})\right)] \\ \leq &\sum_{s=1}^{N}[I_s^{\text{app}}]^2 + 2\sum_{s=1}^{N}\sum_{s'>s}^{N}[I_s^{\text{app}} I_{s'}^{\text{app}}] = I_0^2, \end{aligned} \tag{C.3}$$

where

$$I_0 = \sum_{s=1}^{N} I_s^{\text{app}} \tag{C.4}$$

is the total apparent flux of the ensemble. This upper bound is a case of total constructive interference of all the sources in the ensemble. Depending on the source distribution and the wave mode of interest, the cross terms may interfere in a number of ways. For instance, two sources displaced from one another in the direction that is perpendicular to a given spatial mode will necessarily interfere totally constructively for that spatial mode, which can be seen by the dot product between $\mathbf{k}_\perp$ and $\boldsymbol{\theta_s} - \boldsymbol{\theta_{s'}}$ in the argument of the cosine. On the other hand, there exist other combinations of sources that exhibit total destructive interference for that same mode. The net effect for a power spectrum measurement depends crucially on the exact baseline distribution of the interferometer and the nature of the RFI environment. For instance, in [102], multiple instances of reflective aircraft flying North to South over the MWA were found, which can be thought of as an ensemble of sources with North-South displacements. These sources tend to produce stronger measurements on East-West baselines. On the other hand, a substantially less remote site might experience a distribution of RFI emitters that is more or less uniform over locations on the sky. Despite not resembling the physical circumstances of certain instruments and RFI environments, we calculate the expected power of a random ensemble of narrowband point-sources that are distributed uniformly and independently over a small patch of sky in order to gain intuition about the nature of possible coherence of RFI sources. We will also consider the true flux distribution as being independent of the location of the source on the sky.

Assuming the primary beam of the instrument to have some characteristic opening angle, θ_H, we can consider sources distributed over a small square patch of the sky of side length $2\theta_H$. Ignoring curvature of the sky, the probability density of a source can then be written

$$f(\boldsymbol{\theta}|\theta_H) = \frac{1}{(2\theta_H)^2}\Pi\left(\frac{\theta_x}{2\theta_H}\right)\Pi\left(\frac{\theta_y}{2\theta_H}\right). \tag{C.5}$$

Using the assumption of independent placement, we can write the joint probability distribution of two sources, labeled s and s' as the product of the above equation for the different

sources:

$$f(\boldsymbol{\theta_s}, \boldsymbol{\theta_{s'}}|\theta_H, \theta_H) = f(\boldsymbol{\theta_s}|\theta_H) f(\boldsymbol{\theta_{s'}}|\theta_H). \tag{C.6}$$

Recalling that $I_s^{\text{app}} = A(\boldsymbol{\theta_s}) I_s$, we integrate over this probability distribution and find that the expected power of the ensemble is given by

$$\langle P_{\text{ensemble}} \rangle = \zeta(f_0) \left(N \langle \tilde{A}(\boldsymbol{\theta})^2 \rangle \langle I^2 \rangle + N(N-1) \langle I \rangle^2 \frac{|\tilde{A}(\mathbf{u})|^2}{(2\theta_H)^4} \right), \tag{C.7}$$

where $\tilde{A}(\boldsymbol{\theta})$ is notation meant to indicate the beam clipped at the opening angle, $\tilde{A}(\mathbf{u})$ is its Fourier transform as a function of baseline separation vector, and $\zeta(f_0)$ is just the prefactor outside the sums in equation C.1:

$$\zeta(f_0) = \frac{1}{V_{\mathcal{M}}} \Psi(f_0)^2 \kappa(f_0)^2 r_{\parallel}(f_0)^4 \beta^2. \tag{C.8}$$

We have switched from $\mathbf{k}_\perp$ to $\mathbf{u}$ since quantities involving the primary beam are more readily understood in this frame.

We define the expected coherency factor, as a function of baseline, as

$$c(\mathbf{u}) = \frac{|\tilde{A}(\mathbf{u})|^2}{(2\theta_H)^4}. \tag{C.9}$$

For a non-negative beam on the sky that is peak-normalized to 1, one can show that this coherence function is bounded between 0 and 1, with its maximum attained at the origin in the uv-plane. As an example, for a beam that is equally sensitive out to the opening angle (a top-hat beam), this function takes the form

$$c(\mathbf{u}) = \text{sinc}^2(2\pi\theta_H u) \text{sinc}^2(2\pi\theta_H v). \tag{C.10}$$

This is equal to 1 at the origin and falls off as a power law in any given direction. As a result, baselines longer than the inverse width of the top-hat primary beam tend to experience dramatically less coherence than shorter baselines. We expect this intuition to transfer to the case of a primary beam with more realistic angular dependence by way of typical Fourier transform reasoning. What we see, then, is that uniformly distributed sources are expected to add incoherently for most baselines used in analysis, rather than conveniently adding

destructively, or much worse, constructively. Moreover, we can easily scale the calculated or simulated power spectrum of a single source such as in §4.3 by a linear factor if we instead want to consider an ensemble of emitters.

Note that in equation C.7, other than the assumption of independence of angular position, we have made no assumptions about the nature of the flux distribution of the ensemble. In principle, one could inform this distribution using studies such as [63] and [84]. This would allow one to specifically relate the average flux squared (incoherent term) to the average flux of the ensemble, and, in turn set an average allowed apparent flux of individual RFI sources within an integration.

Appendix D

POWER SPECTRA OF CHROMATIC SOURCES DISPLACED FROM PHASE CENTER

In this appendix, we consider cylindrical power spectra of point-sources that are displaced from phase center. We find that this creates a wedge. This wedge arises because sources that are fixed in angular coordinates as a function of frequency appear to move transversely in cosmological coordinates as a function of line-of-sight distance. This is essentially a redshift-space distortion. See Figure 8 of [30].

To simplify notation, let us define

$$\begin{aligned} \nu_c &= \frac{1}{2}(\nu_U + \nu_L) \\ B &= \nu_U - \nu_L \\ h(\nu) &= I_0 \phi(\nu) r_\parallel(\nu)^2 \Pi\left(\frac{\nu - \nu_c}{B}\right) \Psi(\nu) \end{aligned} \tag{D.1}$$

In principle, we could express this in terms of the variable $r_\parallel(\nu)$, which we do below without the need to explicitly evaluate the form of the function after the change of variables. We have

$$P(k_\perp, k_\parallel) = \frac{1}{V_{\mathcal{M}}} \int_{-\infty}^{\infty} dr_\parallel \int_{-\infty}^{\infty} dr'_\parallel \, h(r_\parallel) h^*(r'_\parallel) e^{ik_\parallel (r_\parallel - r'_\parallel)} \int_0^{2\pi} d\phi_{k_\perp} \, e^{ik_\perp |\theta_0| \cos(\phi_{k_\perp} - \phi_{\theta_0})(r_\parallel - r'_\parallel)} \tag{D.2}$$

Using the Anger-Jacobi expansion,[1] we see that this is none other than

$$P(k_\perp, k_\parallel) = \frac{1}{V_{\mathcal{M}}} \int_{-\infty}^{\infty} dr_\parallel \int_{-\infty}^{\infty} dr'_\parallel \, h(r_\parallel) h^*(r'_\parallel) J_0\big(k_\perp |\theta_0| (r_\parallel - r'_\parallel)\big) e^{ik_\parallel (r_\parallel - r'_\parallel)} \tag{D.3}$$

where J_0 is a cylindrical Bessel function. With appropriate changes of variables, these

[1] This expansion is actually taken as the definition of the Bessel function of the first kind in [85].

integrals may be written as

$$P(k_\perp, k_\parallel) = \frac{1}{V_\mathcal{M}} \int_{-\infty}^{\infty} dr_\parallel (h \star h)(r_\parallel) J_0(k_\perp |\theta_0| r_\parallel) e^{i k_\parallel r_\parallel} \tag{D.4}$$

where the $\star$ symbol indicates correlation. The autocorrelation of $h(r_\parallel)$ normalized by the cosmological volume factor is exactly the inverse Fourier transform of the quantity $P(0, 0, k_\parallel)$. Therefore, from the convolution theorem,

$$P(k_\perp, k_\parallel) = \frac{1}{k_\perp |\theta_0|} P(0, 0, k_\parallel) * \Upsilon\left(\frac{k_\parallel}{k_\perp |\theta_0|}\right) \tag{D.5}$$

where the $*$ operator represents convolution over the $k_\parallel$ direction and $\Upsilon(u)$ is the Fourier transform of $J_0(u)$, given explicitly under these Fourier conventions by

$$\Upsilon(u) = \Pi\left(\frac{u}{2}\right) \frac{2}{\sqrt{1 - u^2}} \tag{D.6}$$

which can also be shown using the Anger-Jacobi expansion. This convolution kernel has a larger width for sources far from phase center and at larger $k_\perp$ modes. For sources near phase center and typical regions of k-space used in analysis, this kernel is quite narrow.

Appendix E

UNCERTAINTY DISTRIBUTIONS IN THE POWER SPECTRUM

This appendix contains work in which the primary author mostly played an advisory role. Since it is relevant to the work in the main body of this thesis and is not contained elsewhere, we report it here. Jordan Brown, who is currently a PhD student in the UC Berkeley logic program, performed some of the original analytic calculations, wrote many of the original proofs, and spent many hours validating the results with simulations and other numerical methods.

E.1 Uncertainty Distribution in the εppsilon *Power Spectrum Estimator*

Following to equation 25 of [8],[1] the power spectrum estimator is given by

$$p \propto \Re(x_e x_o^*), \tag{E.1}$$

where x_e and x_o are the even and odd, weighted data cubes in the Fourier domain that result from the generalized Lomb-Scargle Fourier transform method. x_e and x_o are linear transformations of the visibility data. Each has a noise component, which is circular Gaussian. Breaking each term into a signal and noise component, we have

$$p \propto \Re((s_e + n_e)(s_o^* + n_o^*)). \tag{E.2}$$

The signal of the even and odd cubes is approximately identical. Thus we can expand this as

$$p \propto |s|^2 + \Re(s n_o^* + s^* n_e + n_e n_o^*). \tag{E.3}$$

[1] The equation in that reference is missing a "real component" operator, $\Re$, since the sum and difference cubes are actually modulus-squared.

For the power spectrum measurements in this work, we are vastly noise-dominated, meaning that we can ignore the cross terms between signal and noise when characterizing the uncertainty. Thus, we must consider the random variable

$$W = \Re(n_e n_o^*) \tag{E.4}$$

Note that the n_e and n_o are independent, but not necessarily identically distributed. Let us denote their standard deviations with σ_e and σ_o.

Consider these complex quantities in terms of the real and imaginary components:

$$\begin{aligned} n_e &= X_e + iY_e \\ n_o &= X_o + iY_o. \end{aligned} \tag{E.5}$$

Then

$$W = X_e X_o + Y_e Y_o. \tag{E.6}$$

Each of these iid terms is a product distribution of two independent Gaussian random variables. Let

$$\begin{aligned} W_X &= X_e X_o \\ W_Y &= Y_e Y_o \end{aligned} \tag{E.7}$$

Then the distribution of W_X is given by

$$f_{W_X}(w) = \frac{1}{2\pi\sigma_e\sigma_o} \int_{-\infty}^{\infty} \frac{dx}{|x|} e^{-x^2/2\sigma_e^2} e^{-w^2/2x^2\sigma_o^2} \tag{E.8}$$

We are free to rewrite this like so:

$$f_{W_X}(w) = \frac{1}{\pi\sigma_e\sigma_o} \int_0^{\infty} \frac{dx}{x} e^{-\frac{|w|}{2\sigma_e\sigma_o}(e^{\ln(x^2\sigma_o/|w|\sigma_e)} + e^{\ln(|w|\sigma_e/x^2\sigma_o)})} \tag{E.9}$$

Using the substitution $t = \ln(x^2\sigma_o/|w|\sigma_e)$, we can write

$$f_{W_X}(w) = \frac{1}{\pi\sigma_e\sigma_o} \int_0^{\infty} dt\, e^{-|w|\cosh t/\sigma_e\sigma_o} = \frac{K_0(|w|/\sigma_e\sigma_o)}{\pi\sigma_e\sigma_o} \tag{E.10}$$

where K_0 is a modified Bessel function of the second kind, and the final identity can be found using contour integration methods [100].

W is a sum of two independent terms with exactly this distribution. The resulting sum is distributed according to the autoconvolution of the above distribution. Rather than compute this convolution directly, we instead evaluate the square of the characteristic function, which is the characteristic function of the autoconvolution by the convolution theorem. From [2], we have

$$K_0(x) = \frac{1}{2}\int_{-\infty}^{\infty} dk\, \frac{\cos(kx)}{\sqrt{k^2+1}} \tag{E.11}$$

Since $\sin(kx)$ is odd and $\sqrt{k^2+1}$ is even, we may write this as

$$\frac{K_0(|x|)}{\pi} = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, \frac{e^{-ikx}}{\sqrt{k^2+1}} \tag{E.12}$$

where we have taken advantage of the fact that only the even part of e^{-ikx} contributes to the integral in order to introduce the $|x|$ on the left-hand-side. Fourier inversion and the convolution theorem implies then that the characteristic function of W is

$$E[e^{ikw}] = \frac{1}{\sigma_e^2\sigma_o^2k^2+1} \tag{E.13}$$

which is precisely the characteristic function of a biexponential random variable centered on the origin with a rate constant of $1/\sigma_e\sigma_o$. Due to a statement of the central limit theorem by Lyapunov, the fact that a biexponential random variable has finite mean and variance, as well as zero skew, implies that the sample means converge to a Gaussian distribution in the limit of large sample size. Simulation methods show that this convergence is quite rapid, meaning that many of our spherical power spectrum bins can be characterized as having Gaussian uncertainties.

E.2 Distribution of the Difference Power Spectrum

Another noise estimate used in εPPSILON is to take the modulus square of the difference of the even and odd cubes, hereon the difference power spectrum. The difference of two independent circular gaussian random variables is yet antoher circular gaussian random variable. The variance of each component of the differenced quantity is the sum of the variances of each component. The modulus square of a circular Gaussian random variable is an exponential

random variable with rate constant equal to the reciprocal of twice the variance of the gaussian components. Thus, each voxel of the difference power spectrum is exponentially distributed so long as the difference is not corrupted by non-thermal contributions e.g. RFI or residual sky signal.

Recall from the previous section that the thermal distribution in the voxels of the power spectrum estimator are bixeponentially distributed with rate constant $1/\sigma_e\sigma_o$. Such a distribution has a variance of $2\sigma_e^2\sigma_o^2$. The corresponding voxel in the difference power spectrum has a mean value of $2(\sigma_e^2 + \sigma_o^2)$. If $\sigma_e \approx \sigma_o$, then the standard deviation of the biexponential distribution is approximately equal to the expected value of the difference power spectrum up to a constant factor. Since means and (co)variances add in predictable ways, we can compare these two different noise metrics in a straight-forward way as a form of validation in the pipeline at different levels of $\boldsymbol{k}$-space averaging.

We note here that calculation of the probability density function of a sum of exponential random variables is analytically tractable. The calculation can be performed by making use of the convolution theorem, i.e. transforming to the Fourier domain, multiplying the characteristic functions, and inverting the Fourier transform. The inversion requires a contour integral. All the residues lie in one half of the complex plane, and Jordan's Lemma show that a pair of semicircular contours yields the answer. If all the rate constants of the set of exponentials are distinct, the poles are all simple. If any of them have multiplicty greater than 1, those poles are not simple. After a rather gory calculation with a dizzying number of factors of i, the final expression is

$$f_Y(y) = \theta(y) \sum_{j=1}^{n} \sum_{m=0}^{m_j-1} \frac{\lambda_j^{m_j}}{(m_j - m - 1)!m!} y^{m_j-m-1} e^{-\lambda_j y} \Psi_{mj}(-\lambda_j), \tag{E.14}$$

where Y is the summed random variable, $\theta(y)$ is the Heaviside step function (i.e. the variable is positive definite), λ_j are the unique rate constants in the set, m_j are their multiplicities, and Ψ_{mj} is given by:

$$\Psi_{mj}(t) = \frac{d^m}{dt^m}\left[\prod_{l\neq j}\left(\frac{\lambda_l}{\lambda_l + t}\right)^{m_l}\right] \tag{E.15}$$

E.3 The Phase-type Distribution

It turns out that arbitrary sums of erlang mixtures (a set of which the above distribution is an element) can actually be described rather neatly using what is called a Phase-type Distribution, which commonly occurs in queuing theory (need a reference). A phase-type random variable has probability density function

$$f_Z(z) = \boldsymbol{\alpha} \exp\left(z\boldsymbol{S}\right)\boldsymbol{S}^0, \tag{E.16}$$

where $\boldsymbol{S}$ is expressed in terms of a transition rate matrix $\boldsymbol{Q}$ for a continuous-time markov process of order $m+1$, where the 0th state is the absorbing state:

$$\boldsymbol{Q} = \begin{pmatrix} 0 & \mathbf{0} \\ \boldsymbol{S}^0 & \boldsymbol{S} \end{pmatrix}. \tag{E.17}$$

α represents the initial state of the process in the nonabsorbing states,[2] $\boldsymbol{S}^0 = -\boldsymbol{S}\mathbf{1}$, $\mathbf{1}$ is a vector of all 1's, and exp denotes a matrix exponential.

For a sum of exponential random variables with two distinct rate constants of multiplicty 2 and 3, we would use the parameters

$$\alpha = (1, 0, 0, 0, 0) \tag{E.18}$$

$$\boldsymbol{S} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & 0 \\ 0 & -\lambda_1 & \lambda_1 & 0 & 0 \\ 0 & 0 & -\lambda_2 & \lambda_2 & 0 \\ 0 & 0 & 0 & -\lambda_2 & \lambda_2 \\ 0 & 0 & 0 & 0 & -\lambda_2 \end{pmatrix} \tag{E.19}$$

. This special case is known as a hypoexponential random variable.

If the rate constants are known, as in our case, then this allows for an efficient numerical computation of the probability density function that describes the difference power spectrum. Moreoever, a biexpoential random variable can be written as a difference of two exponential

[2]Usually the probability of starting in the absorbing state is assumed to be 0.

random variables with identical rate constants. A sum of independent biexponential random variables can thusly be written as a difference of two phase-type random variables. Therefore, one could also use the phase-type distribution to numerically calculate the error bars for the non-Gaussian bins in the power spectrum estimator as well. This is not yet implemented in εPPSILON. There is one final caveat to mention, which is that the voxels in the power spectrum are not strictly independent, but are lightly correlated. We find empirically that bins used for limit estimation are still largely Gaussian, and since covariances neatly add, the error bars are still properly estimated for those bins. Before this formalism can be properly implemented in εPPSILON, we would need to assess the effect of such correlations on the resulting probability density function to determine the relationship between the summed variances and the actual confidence intervals that the error bars express.

Appendix F

POWER SPECTRUM WALL OF SHAME

We show all power spectra that were used to remove observations from the 21-cm power spectrum upper limit in Chapter 7 that were not determined to contain RFI using only SSINS. We show them in no particular order, along with the number of observations present in each integration. Since the excess window power is almost exclusively observed in the East-West polarization, we show only power spectra in that polarization. There is some overlap between the lists, resulting in an observation count higher than 119, which is the number of observations finally removed from the limit calculation.

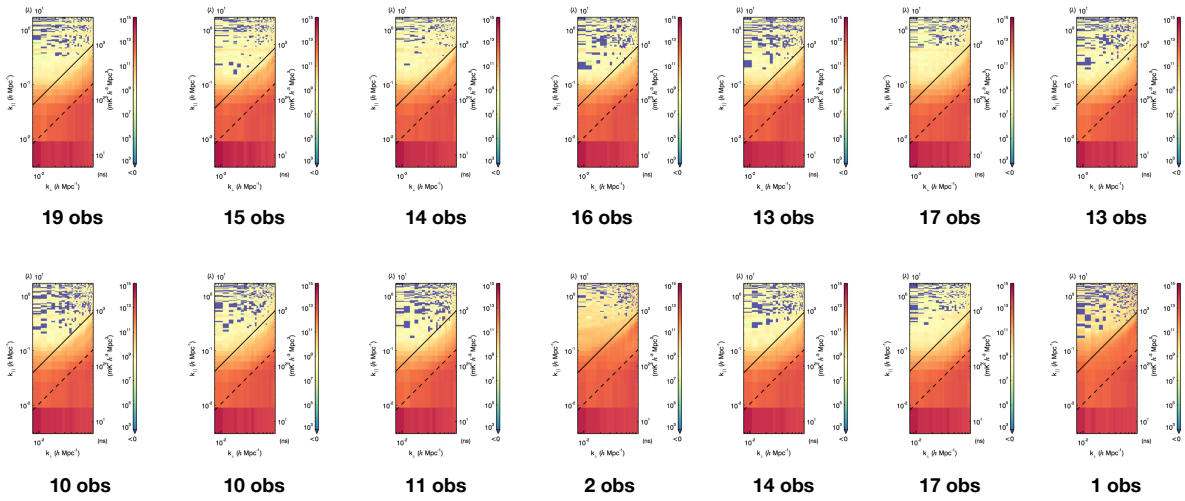


Figure F.1: All subintegrations that were determined to possess clear indications of RFI contamination in the EoR window that were subsequently removed from the limit integration.

BIBLIOGRAPHY

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

[2] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition, 1964.

[3] Zaki S. Ali, Aaron R. Parsons, Haoxuan Zheng, Jonathan C. Pober, Adrian Liu, James E. Aguirre, Richard F. Bradley, Gianni Bernardi, Chris L. Carilli, Carina Cheng, David R. DeBoer, Matthew R. Dexter, Jasper Grobbelaar, Jasper Horrell, Daniel C. Jacobs, Pat Klima, David H. E. MacMahon, Matthys Maree, David F. Moore, Nima Razavi, Irina I. Stefan, William P. Walbrugh, and Andre Walker. PAPER-64 Constraints on Reionization: The 21 cm Power Spectrum at z = 8.4. *The Astrophysical Journal*, 809(1):61, August 2015.

[4] Zaki S. Ali, Aaron R. Parsons, Haoxuan Zheng, Jonathan C. Pober, Adrian Liu, James E. Aguirre, Richard F. Bradley, Gianni Bernardi, Chris L. Carilli, Carina Cheng, David R. DeBoer, Matthew R. Dexter, Jasper Grobbelaar, Jasper Horrell, Daniel C. Jacobs, Pat Klima, David H. E. MacMahon, Matthys Maree, David F. Moore, Nima Razavi, Irina I. Stefan, William P. Walbrugh, and Andre Walker. Erratum: "PAPER-64 Constraints on Reionization: The 21 cm Power Spectrum at z = 8.4" (¡A href="http://doi.org/10.1088/0004-637x/809/1/61"¿2015, ApJ, 809, 61¡/A¿). *The Astrophysical Journal*, 863(2):201, August 2018.

[5] T. An, X. Chen, P. Mohan, and B. Q. Lao. Radio Frequency Interference Mitigation. *Acta Astronomica Sinica*, 58(5):43, September 2017.

[6] Ramana Athreya. A New Approach to Mitigation of Radio Frequency Interference in Interferometric Data. *The Astrophysical Journal*, 696(1):885–890, May 2009.

[7] Willem A. Baan. Implementing RFI Mitigation in Radio Science. *Journal of Astronomical Instrumentation*, 8(1):1940010, January 2019.

[8] N. Barry, A. P. Beardsley, R. Byrne, B. Hazelton, M. F. Morales, J. C. Pober, and I. Sullivan. The FHD/∈ppsilon Epoch of Reionisation power spectrum pipeline. *Publications of the Astronomical Society of Australia*, 36:e026, July 2019.

[9] 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.

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

[11] Nichole Barry. *Enhancing EoR Limits through Improved Instrumental Calibration of the MWA*. PhD thesis, University of Washington, January 2018.

[12] 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, S. 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*, 429:L5–L9, February 2013.

[13] 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.

[14] 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.

[15] L. V. Benkevitch, A. E. E. Rogers, C. J. Lonsdale, R. J. Cappallo, D. Oberoi, P. J. Erickson, and K. A. V. Baker. Van Vleck correction generalization for complex correlators with multilevel quantization. *arXiv e-prints*, page arXiv:1608.04367, August 2016.

[16] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley, 2012.

[17] S. Bretteil and R. Weber. Comparison of two cyclostationary detectors for radio frequency interference mitigation in radio astronomy. *Radio Science*, 40(5):RS5S15, July 2005.

[18] P. R. Burd, K. Mannheim, T. März, J. Ringholz, A. Kappes, and M. Kadler. Detecting radio frequency interference in radio-antenna arrays with the recurrent neural network algorithm. *Astronomische Nachrichten*, 339(5):358–362, June 2018.

[19] Ruby Byrne, Miguel F. Morales, Bryna Hazelton, Wenyang Li, Nichole Barry, Adam P. Beardsley, Ronniy Joseph, Jonathan Pober, Ian Sullivan, and Cathryn Trott. Fundamental Limitations on the Calibration of Redundant 21 cm Cosmology Instruments and Implications for HERA and the SKA. *The Astrophysical Journal*, 875(1):70, April 2019.

[20] Ruby Byrne, Miguel F. Morales, Bryna J. Hazelton, and Michael Wilensky. A unified calibration framework for 21 cm cosmology. , 503(2):2457–2477, May 2021.

[21] Sean M. Carroll. *Spacetime and geometry. An introduction to general relativity*. 2004.

[22] Carina Cheng, Aaron R. Parsons, Matthew Kolopanis, Daniel C. Jacobs, Adrian Liu, Saul A. Kohn, James E. Aguirre, Jonathan C. Pober, Zaki S. Ali, Gianni Bernardi, Richard F. Bradley, Chris L. Carilli, David R. DeBoer, Matthew R. Dexter, Joshua S.

Dillon, Pat Klima, David H. E. MacMahon, David F. Moore, Chuneeta D. Nunhokee, William P. Walbrugh, and Andre Walker. Characterizing Signal Loss in the 21 cm Reionization Power Spectrum: A Revised Study of PAPER-64. *The Astrophysical Journal*, 868(1):26, November 2018.

[23] 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.

[24] David R. DeBoer, Aaron R. Parsons, James E. Aguirre, Paul Alexander, Zaki S. Ali, Adam P. Beardsley, Gianni Bernardi, Judd D. Bowman, Richard F. Bradley, Chris L. Carilli, Carina Cheng, Eloy de Lera Acedo, Joshua S. Dillon, Aaron Ewall-Wice, Gcobisa Fadana, Nicolas Fagnoni, Randall Fritz, Steve R. Furlanetto, Brian Glendenning, Bradley Greig, Jasper Grobbelaar, Bryna J. Hazelton, Jacqueline N. Hewitt, Jack Hickish, Daniel C. Jacobs, Austin Julius, MacCalvin Kariseb, Saul A. Kohn, 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, Nipanjana Patra, Samantha Pieterse, Jonathan C. Pober, Nima Razavi-Ghods, Jon Ringuette, James Robnett, Kathryn Rosie, Raddwine Sell, Craig Smith, Angelo Syce, Max Tegmark, Nithyanandan Thyagarajan, Peter K. G. Williams, and Haoxuan Zheng. Hydrogen Epoch of Reionization Array (HERA). *Publications of the Astronomical Society of the Pacific*, 129(974):045001, April 2017.

[25] Aaron Ewall-Wice, Joshua S. Dillon, Adrian Liu, and Jacqueline Hewitt. The impact of modelling errors on interferometer calibration for 21 cm power spectra. *Monthly Notices of the Royal Astronomical Society*, 470(2):1849–1870, September 2017.

[26] Xiaohui Fan, C. L. Carilli, and B. Keating. Observational Constraints on Cosmic Reionization. *Annual Review of Astronomy and Astrophysics*, 44(1):415–462, September 2006.

[27] 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(4-6):181–301, October 2006.

[28] Robert Gallager. *Stochastic Processes: Theory for Applications*. 12 2013.

[29] 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(2):759–771, April 2005.

[30] A. J. S. Hamilton. *Linear Redshift Distortions: a Review*, volume 231, page 185. 1998.

[31] F. J. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978.

[32] Bryna J. Hazelton, Daniel C. Jacobs, Jonathan C. Pober, and Adam P. Beardsley. pyuvdata: an interface for astronomical interferometeric datasets in python. *The Journal of Open Source Software*, 2(10):140, February 2017.

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

[34] Grégory Hellbourg, Rodolphe Weber, Cécile Capdessus, and Albert-Jan Boonstra. Cyclostationary approaches for spatial RFI mitigation in radio astronomy. *Comptes Rendus Physique*, 13(1):71–79, January 2012.

[35] David W. Hogg. Distance measures in cosmology. *arXiv e-prints*, pages astro–ph/9905116, May 1999.

[36] 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, January 2017.

[37] 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.

[38] Nicholas S. Kern, Joshua S. Dillon, Aaron R. Parsons, Christopher L. Carilli, Gianni Bernardi, Zara Abdurashidova, James E. Aguirre, Paul Alexander, Zaki S. Ali, Yanga Balfour, Adam P. Beardsley, Tashalee S. Billings, Judd D. Bowman, Richard F. Bradley, Philip Bull, Jacob Burba, Steven Carey, Carina Cheng, David R. DeBoer, Matt Dexter, Eloy de Lera Acedo, John Ely, Aaron Ewall-Wice, Nicolas Fagnoni,

Randall Fritz, Steve R. Furlanetto, Kingsley Gale-Sides, Brian Glendenning, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Bryna J. Hazelton, Jacqueline N. Hewitt, Jack Hickish, Daniel C. Jacobs, Austin Julius, Joshua Kerrigan, Piyanat Kittiwisit, Saul A. Kohn, Matthew Kolopanis, Adam Lanman, Paul La Plante, Telalo Lekalake, Adrian Liu, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary E. Martinot, Eunice Matsetela, Andrei Mesinger, Mathakane Molewa, Miguel F. Morales, Tshegofalang Mosiane, Steven G. Murray, Abraham R. Neben, Bojan Nikolic, Chuneeta D. Nunhokee, Nipanjana Patra, Samantha Pieterse, Jonathan C. Pober, Nima Razavi-Ghods, Jon Ringuette, James Robnett, Kathryn Rosie, Peter Sims, Craig Smith, Angelo Syce, Nithyanandan Thyagarajan, Peter K. G. Williams, and Haoxuan Zheng. Absolute Calibration Strategies for the Hydrogen Epoch of Reionization Array and Their Impact on the 21 cm Power Spectrum. *The Astrophysical Journal*, 890(2):122, February 2020.

[39] Nicholas S. Kern and Adrian Liu. Gaussian process foreground subtraction and power spectrum estimation for 21 cm cosmology. , 501(1):1463–1480, February 2021.

[40] Nicholas S. Kern, Aaron R. Parsons, Joshua S. Dillon, Adam E. Lanman, Nicolas Fagnoni, and Eloy de Lera Acedo. Mitigating Internal Instrument Coupling for 21 cm Cosmology. I. Temporal and Spectral Modeling in Simulations. *The Astrophysical Journal*, 884(2):105, October 2019.

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

[42] Joshua Kerrigan, Paul La Plante, Saul Kohn, Jonathan C. Pober, James Aguirre, Zara Abdurashidova, Paul Alexander, Zaki S. Ali, Yanga Balfour, Adam P. Beards-

ley, Gianni Bernardi, Judd D. Bowman, Richard F. Bradley, Jacob Burba, Chris L. Carilli, Carina Cheng, David R. DeBoer, Matt Dexter, Eloy de Lera Acedo, Joshua S. Dillon, Julia Estrada, Aaron Ewall-Wice, Nicolas Fagnoni, Randall Fritz, Steve R. Furlanetto, Brian Glendenning, Bradley Greig, Jasper Grobbelaar, Deepthi Gorthi, Ziyaad Halday, Bryna J. Hazelton, Jack Hickish, Daniel C. Jacobs, Austin Julius, Nicholas S. Kern, Piyanat Kittiwisit, Matthew Kolopanis, Adam Lanman, Telalo Lekalake, Adrian Liu, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary E. Martinot, Eunice Matsetela, Andrei Mesinger, Mathakane Molewa, Miguel F. Morales, Tshegofalang Mosiane, Abraham R. Neben, Aaron R. Parsons, Nipanjana Patra, Samantha Pieterse, Nima Razavi-Ghods, Jon Ringuette, James Robnett, Kathryn Rosie, Peter Sims, Craig Smith, Angelo Syce, Nithyanandan Thyagarajan, Peter K. G. Williams, and Haoxuan Zheng. Optimizing sparse RFI prediction using deep learning. *Monthly Notices of the Royal Astronomical Society*, 488(2):2605–2615, September 2019.

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

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

[45] Adrian Liu, Aaron R. Parsons, and Cathryn M. Trott. Epoch of reionization window. II. Statistical methods for foreground wedge reduction. *Physical Review D*, 90(2):023019, July 2014.

[46] Adrian Liu and J. Richard Shaw. Data Analysis for Precision 21 cm Cosmology. *Publications of the Astronomical Society of the Pacific*, 132(1012):062001, June 2020.

[47] Shyeh Tjing Loi, Iver H. Cairns, Tara Murphy, Philip J. Erickson, Martin E. Bell, Antonia Rowlinson, Balwinder Singh Arora, John Morgan, Ronald D. Ekers, Natasha Hurley-Walker, and David L. Kaplan. Density duct formation in the wake of a travelling ionospheric disturbance: Murchison Widefield Array observations. *Journal of Geophysical Research (Space Physics)*, 121(2):1569–1586, February 2016.

[48] Alejandro Luque. Radio frequency electromagnetic radiation from streamer collisions. *Journal of Geophysical Research: Atmospheres*, 122(19):10,497–10,509, 2017.

[49] David J. C. MacKay. *Information Theory, Inference Learning Algorithms*. Cambridge University Press, USA, 2002.

[50] S. J. McSweeney, S. M. Ord, D. Kaur, N. D. R. Bhat, B. W. Meyers, S. E. Tremblay, J. Jones, B. Crosse, and K. R. Smith. MWA tied-array processing III: Microsecond time resolution via a polyphase synthesis filter. , 37:e034, August 2020.

[51] F. G. Mertens, A. Ghosh, and L. V. E. Koopmans. Statistical 21-cm signal separation via Gaussian Process Regression analysis. , 478(3):3640–3652, August 2018.

[52] F. G. Mertens, M. Mevius, L. V. E. Koopmans, A. R. Offringa, G. Mellema, S. Zaroubi, M. A. Brentjens, H. Gan, B. K. Gehlot, V. N. Pandey, A. M. Sardarabadi, H. K. Vedantham, S. Yatawatta, K. M. B. Asad, B. Ciardi, E. Chapman, S. Gazagnes, R. Ghara, A. Ghosh, S. K. Giri, I. T. Iliev, V. Jelić, R. Kooistra, R. Mondal, J. Schaye, and M. B. Silva. Improved upper limits on the 21 cm signal power spectrum of neutral hydrogen at z $\approx$ 9.1 from LOFAR. *Monthly Notices of the Royal Astronomical Society*, 493(2):1662–1685, April 2020.

[53] Andrei Mesinger, Steven Furlanetto, and Renyue Cen. 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal. *Monthly Notices of the Royal Astronomical Society*, 411(2):955–972, February 2011.

[54] 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, November 2008.

[55] Miguel F. Morales, Adam Beardsley, Jonathan Pober, Nichole Barry, Bryna Hazelton, Daniel Jacobs, and Ian Sullivan. Understanding the diversity of 21 cm cosmology analyses. *Monthly Notices of the Royal Astronomical Society*, 483(2):2207–2216, February 2019.

[56] 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.

[57] Miguel F. Morales and Jacqueline Hewitt. Toward Epoch of Reionization Measurements with Wide-Field Radio Observations. *The Astrophysical Journal*, 615(1):7–18, November 2004.

[58] 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.

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

[60] Steven G. Murray and C. M. Trott. The Effect of Baseline Layouts on the Epoch of Reionization Foreground Wedge: A Semianalytical Approach. , 869(1):25, December 2018.

[61] A. R. Offringa, A. G. de Bruyn, M. Biehl, S. Zaroubi, G. Bernardi, and V. N. Pandey. Post-correlation radio frequency interference classification methods. *Monthly Notices of the Royal Astronomical Society*, 405(1):155–167, June 2010.

[62] A. R. Offringa, A. G. de Bruyn, and S. Zaroubi. Post-correlation filtering techniques for off-axis source and RFI removal. *Monthly Notices of the Royal Astronomical Society*, 422(1):563–580, May 2012.

[63] A. R. Offringa, A. G. de Bruyn, S. Zaroubi, L. V. E. Koopmans, S. J. Wijnholds, F. B. Abdalla, W. N. Brouw, B. Ciardi, I. T. Iliev, G. J. A. Harker, G. Mellema, G. Bernardi, P. Zarka, A. Ghosh, A. Alexov, J. Anderson, A. Asgekar, I. M. Avruch, R. Beck, M. E. Bell, M. R. Bell, M. J. Bentum, P. Best, L. Bîrzan, F. Breitling, J. Broderick, M. Brüggen, H. R. Butcher, F. de Gasperin, E. de Geus, M. de Vos, S. Duscha, J. Eislöffel, R. A. Fallows, C. Ferrari, W. Frieswijk, M. A. Garrett, J. Grießmeier, T. E. Hassall, A. Horneffer, M. Iacobelli, E. Juette, A. Karastergiou, W. Klijn, V. I. Kondratiev, M. Kuniyoshi, G. Kuper, J. van Leeuwen, M. Loose, P. Maat, G. Macario, G. Mann, J. P. McKean, H. Meulman, M. J. Norden, E. Orru, H. Paas, M. Pandey-Pommier, R. Pizzo, A. G. Polatidis, D. Rafferty, W. Reich, R. van Nieuwpoort, H. Röttgering, A. M. M. Scaife, J. Sluman, O. Smirnov, C. Sobey, M. Tagger, Y. Tang, C. Tasse, S. ter Veen, C. Toribio, R. Vermeulen, C. Vocks, R. J. van Weeren, M. W. Wise, and O. Wucknitz. The brightness and spatial distributions of terrestrial radio sources. *Monthly Notices of the Royal Astronomical Society*, 435(1):584–596, October 2013.

[64] 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 and Astrophysics*, 549:A11, January 2013.

[65] A. R. Offringa, F. Mertens, and L. V. E. Koopmans. The impact of interference excision on 21-cm epoch of reionization power spectrum analyses. *Monthly Notices of the Royal Astronomical Society*, 484(2):2866–2875, April 2019.

[66] A. R. Offringa, F. Mertens, S. van der Tol, B. Veenboer, B. K. Gehlot, L. V. E. Koopmans, and M. Mevius. Precision requirements for interferometric gridding in the analysis of a 21 cm power spectrum. *Astronomy and Astrophysics*, 631:A12, November 2019.

[67] A. R. Offringa, J. J. van de Gronde, and J. B. T. M. Roerdink. A morphological algorithm for improving radio-frequency interference detection. *Astronomy and Astrophysics*, 539:A95, March 2012.

[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] Gregory Paciga, Joshua G. Albert, Kevin Bandura, Tzu-Ching Chang, Yashwant Gupta, Christopher Hirata, Julia Odegova, Ue-Li Pen, Jeffrey B. Peterson, Jayanta Roy, J. Richard Shaw, Kris Sigurdson, and Tabitha Voytek. A simulation-calibrated

limit on the H I power spectrum from the GMRT Epoch of Reionization experiment. , 433(1):639–647, July 2013.

[70] Jaehong Park, Andrei Mesinger, Bradley Greig, and Nicolas Gillet. Inferring the astrophysics of reionization and cosmic dawn from galaxy luminosity functions and the 21-cm signal. *Monthly Notices of the Royal Astronomical Society*, 484(1):933–949, March 2019.

[71] 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, September 2012.

[72] Ajinkya H. Patil, Sarod Yatawatta, Saleem Zaroubi, Léon V. E. Koopmans, A. G. de Bruyn, 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.

[73] Luke W. Peck and Danielle M. Fenech. SERPent: Scripted E-merlin Rfi-mitigation PipelinE for iNTerferometry, December 2013.

[74] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio,

M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck 2018 results. VI. Cosmological parameters. *Astronomy and Astrophysics*, 641:A6, September 2020.

[75] Jonathan C. Pober, Zaki S. Ali, Aaron R. Parsons, Matthew McQuinn, James E. Aguirre, Gianni Bernardi, Richard F. Bradley, Chris L. Carilli, Carina Cheng, David R. DeBoer, Matthew R. Dexter, Steven R. Furlanetto, Jasper Grobbelaar, Jasper Horrell, Daniel C. Jacobs, Patricia J. Klima, Saul A. Kohn, Adrian Liu, David H. E. MacMahon, Matthys Maree, Andrei Mesinger, David F. Moore, Nima Razavi-Ghods, Irina I. Stefan, William P. Walbrugh, Andre Walker, and Haoxuan Zheng. PAPER-64 Constraints On Reionization. II. The Temperature of the z =8.4 Intergalactic Medium. *The Astrophysical Journal*, 809(1):62, August 2015.

[76] Jonathan C. Pober, Adrian Liu, Joshua S. Dillon, James E. Aguirre, Judd D. Bowman, Richard F. Bradley, Chris L. Carilli, David R. DeBoer, Jacqueline N. Hewitt, Daniel C. Jacobs, Matthew McQuinn, Miguel F. Morales, Aaron R. Parsons, Max Tegmark, and Dan J. Werthimer. What Next-generation 21 cm Power Spectrum Measurements can Teach us About the Epoch of Reionization. *The Astrophysical Journal*, 782(2):66, February 2014.

[77] Jonathan C. Pober, Aaron R. Parsons, David R. DeBoer, Patrick McDonald, Matthew McQuinn, James E. Aguirre, Zaki Ali, Richard F. Bradley, Tzu-Ching Chang, and Miguel F. Morales. The Baryon Acoustic Oscillation Broadband and Broad-beam Array: Design Overview and Sensitivity Forecasts. *The Astronomical Journal*, 145(3):65, March 2013.

[78] S. Prabu, P. Hancock, X. Zhang, and S. J. Tingay. A low-frequency blind survey of the low Earth orbit environment using non-coherent passive radar with the Murchison widefield array. , 37:e052, December 2020.

[79] 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 MurchisonWidefield Array. *Experimental Astronomy*, 39(1):73–93, March 2015.

[80] M. J. Reid, D. W. Pesce, and A. G. Riess. An Improved Distance to NGC 4258 and Its Implications for the Hubble Constant. *The Astrophysical Journal Letters*, 886(2):L27, December 2019.

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

[82] Srikrishna Sekhar and Ramana Athreya. Two Procedures to Flag Radio Frequency Interference in the UV Plane. , 156(1):9, July 2018.

[83] Feng Shi, Ningyu Liu, Joseph R. Dwyer, and Kevin M. A. Ihaddadene. Vhf and uhf electromagnetic radiation produced by streamers in lightning. *Geophysical Research Letters*, 46(1):443–451, 2019.

[84] Marcin Sokolowski, Randall B. Wayth, and Morgan Lewis. The statistics of low frequency radio interference at the Murchison Radio-astronomy Observatory. *arXiv e-prints*, page arXiv:1610.04696, October 2016.

[85] E.M. Stein and R. Shakarchi. *Fourier Analysis: An Introduction*. Princeton University Press, 2003.

[86] 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(1):17, November 2012.

[87] 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):52–65, January 2015.

[88] G. B. Taylor, S. W. Ellingson, N. E. Kassim, J. Craig, J. Dowell, C. N. Wolfe, J. Hartman, G. Bernardi, T. Clarke, A. Cohen, N. P. Dalal, W. C. Erickson, B. Hicks, L. J. Greenhill, B. Jacoby, W. Lane, J. Lazio, D. Mitchell, R. Navarro, S. M. Ord, Y. Pihlström, E. Polisensky, P. S. Ray, L. J. Rickard, F. K. Schinzel, H. Schmitt, E. Sigman, M. Soriano, K. P. Stewart, K. Stovall, S. Tremblay, D. Wang, K. W. Weiler, S. White, and D. L. Wood. First Light for the First Station of the Long Wavelength Array. *Journal of Astronomical Instrumentation*, 1(1):1250004–284, December 2012.

[89] Max Tegmark. How to Make Maps from Cosmic Microwave Background Data without Losing Information. , 480(2):L87–L90, May 1997.

[90] A. Richard Thompson, James M. Moran, and Jr. Swenson, George W. *Interferometry and Synthesis in Radio Astronomy, 3rd Edition*. 2017.

[91] 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, October 2013.

[92] 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. , 30:e007, January 2013.

[93] Cathryn M. Trott, C. H. Jordan, S. Midgley, N. Barry, B. Greig, B. Pindor, J. H. Cook, G. Sleap, S. J. Tingay, D. Ung, P. Hancock, A. Williams, J. Bowman, R. Byrne, A. Chokshi, B. J. Hazelton, K. Hasegawa, D. Jacobs, R. C. Joseph, W. Li, J. L. B. Line, C. Lynch, B. McKinley, D. A. Mitchell, M. F. Morales, M. Ouchi, J. C. Pober, M. Rahimi, K. Takahashi, R. B. Wayth, R. L. Webster, M. Wilensky, J. S. B. Wyithe, S. Yoshiura, Z. Zhang, and Q. Zheng. Deep multiredshift limits on Epoch of Reionization 21 cm power spectra from four seasons of Murchison Widefield Array observations. , 493(4):4711–4727, April 2020.

[94] 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.

[95] 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.

[96] Alireza Vafaei Sadr, Bruce A. Bassett, Nadeem Oozeer, Yabebal Fantaye, and Chris Finlay. Deep learning improves identification of Radio Frequency Interference. *Monthly Notices of the Royal Astronomical Society*, 499(1):379–390, October 2020.

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

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

[98] Jacob T. VanderPlas. Understanding the Lomb-Scargle Periodogram. *The Astrophysical Journal Supplement Series*, 236(1):16, May 2018.

[99] D. M. Vine. Review of measurements of the RF spectrum of radiation from lightning. *Meteorology and Atmospheric Physics*, 37(3):195–204, January 1987.

[100] Watson, G.N. *A Treatise on the Theory of Bessel Functions*. 1966.

[101] Michael J. Wilensky, Nichole Barry, Miguel F. Morales, Bryna J. Hazelton, and Ruby Byrne. Quantifying excess power from radio frequency interference in Epoch of Reionization measurements. *Monthly Notices of the Royal Astronomical Society*, 498(1):265–275, August 2020.

[102] Michael J. Wilensky, Miguel F. Morales, Bryna J. Hazelton, Nichole Barry, Ruby Byrne, and Sumit Roy. Absolving the SSINS of Precision Interferometric Radio Data: A New Technique for Mitigating Faint Radio Frequency Interference. *Publications of the Astronomical Society of the Pacific*, 131(1005):114507, November 2019.

[103] Zhicheng Yang, Ce Yu, Jian Xiao, and Bo Zhang. Deep residual detection of radio frequency interference for FAST. *Monthly Notices of the Royal Astronomical Society*, 492(1):1421–1431, February 2020.

[104] D. Zwillinger and S. Kokoska. *CRC Standard Probability and Statistics Formulae*. Boca Raton: Chapman Hall/CRC, 2000.