

A Next Generation Analysis Pipeline for 21cm Epoch of Reionization Studies

by

Joshua Kerrigan

B.S. in Physics, University of Massachusetts Amherst, 2014

Sc.M. in Physics, Brown University, 2016

*A dissertation submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy in the
Department of Physics at Brown University*

Providence, RI

May 2019

© Copyright 2019 by Joshua R. Kerrigan

This dissertation by Joshua R. Kerrigan is accepted in its present form by the Department of Physics as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date _____

Jonathan Pober, Advisor

Recommended to the Graduate Council

Date _____

Ian Dell'Antonio, Reader

Date _____

Gregory Tucker, Reader

Approved by the Graduate Council

Date _____

Andrew Campbell, Dean of the Graduate School

Curriculum Vitae

Education

MAY 2019 Doctor of Philosophy in PHYSICS, **Brown University**, Providence, RI
Advisor: Professor Jonathan Pober

MAY 2016 Master of Science in PHYSICS, **Brown University**, Providence, RI

MAY 2014 Bachelor of Science in PHYSICS, **University of Massachusetts**, Amherst, MA
Minor in Mathematics
Summa Cum Laude

Professional Experience

Graduate Research Assistant (DEC 2015 - MAY 2019)

ASTROPHYSICS/COSMOLOGY

Brown University

Physics Graduate Teaching Assistant (SEP 2014 - DEC 2015)

Brown University

Undergraduate Researcher (NOV 2012-MAY 2014)

LIGO Scientific Collaboration

UMASS Amherst

Awards & Grants

2019 Brown University Data Science Grant - \$13,740.86

‘Advancing Neural Network Analysis of Cosmological Data Sets’

2016 - 2017 NASA Rhode Island Space Grant Fellowship, *NASA RI Space Grant Consortium*

2015 Graduate TA Award, *Dept. of Physics, Brown University*

Publications

*Improved 21cm EoR Power Spectrum Measurements using a Hybrid Foreground
Subtraction and Avoidance Technique*

Joshua Kerrigan, Jonathan Pober, Zaki Ali, Carina Cheng, Aaron Parsons, James Aguirre,
Nichole Barry, Richard Bradley, Gianni Bernardi, Chris Carilli, David DeBoer, Joshua Dillon,
Daniel Jacobs, Saul Kohn, Matthew Kolopanis, Adam Lanman,
Wenyang Li, Adrian Liu, Ian Sullivan, **Published in ApJ**, Sep. 2018.

Optimizing Sparse Radio Frequency Interference Prediction using Deep Learning

Joshua Kerrigan, Paul La Plante, Saul Kohn, Jonathan Pober, James Aguirre,
HERA Collaboration, **Submitted to MNRAS**

*PAPER 64: Multi-Redshift Limits on the 21cm Power Spectrum from
the Epoch of Reionization*

Matthew Kolopanis, Daniel Jacobs, Carina Cheng, Aaron Parsons, Saul Kohn, Adrian Liu,
James Aguirre, Zaki Ali, Gianni Bernardi, Richard Bradley, Chris Carilli, David DeBoer,
Matthew Dexter, Joshua Dillon, Pat Kilma, David MacMahon, David Moore, Chuneeta Nunhokee,
Jonathan Pober, **Joshua Kerrigan**, William Walbrugh, Andre Walker, **Submitted to ApJ**

Collaboration Publications

First low frequency all-sky search for continuous gravitational wave signals

The LIGO Scientific Collaboration, the Virgo Collaboration: J. Aasi, B. P. Abbott, R. Abbott,
T. D. Abbott, M. R. Abernathy, [et al., including **J. Kerrigan**],
Published in Phys. Rev. D, Feb. 2016.

*A search of the Orion spur for continuous gravitational waves using a "loosely coherent"
algorithm on data from LIGO interferometers*

J. Aasi, B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, [et al., including **J. Kerrigan**],
Published in Phys. Rev. D, Feb. 2016.

Conferences

Deep Learning, Sky Calibration, and Foreground Subtraction for HERA

21cm Epoch of Reionization Studies

Dissertation Talk

American Astronomical Society 233rd Winter Meeting 2019, Seattle WA

Improved 21cm EoR Power Spectrum Measurements using a Hybrid Foreground

Subtraction and Avoidance Technique

URSI - National Radio Science Meeting 2018, University of Colorado-Boulder

Mitigating Spectral Leakage in Delay Filtered PAPER-64 Visibilities

Using Foreground Subtraction

URSI - National Radio Science Meeting 2017, University of Colorado-Boulder

Additional Experience

Astrobites.org Staff Writer (JAN 2017 - JAN 2019)

CHAMP Camp Mentor/Facilitator (MAY 2017 - AUG 2018)

Acknowledgements

This thesis is the culmination of a lot of hard work and determination set in motion many years ago. I would not have been as successful in completing this endeavor had it not been for the support from many wonderful people.

I would like to extend my sincere gratitude to my advisor Jonathan Pober who has been an amazing mentor. I have been told in the past that to become a great leader, mentor, or person one should borrow the best traits from a wide variety of people in those positions, but in this instance you'd get 99% of the way there by imitating Jonathan. In the Pober Lab, I'd like to say thank you to Wenyang Li, Adam Lanman, and Jacob Burba, for being there to break the dull periods and talk both astro- and non-astro related topics. In the Brown Physics community, both faculty and students alike have been some of the most welcoming people I've yet to meet and I feel it would be difficult to replicate in any place my future will lead me. For my time in the HERA collaboration, I'd like to say that it is full of some of the most exceptional scientists and I am proud to have had the opportunity to work alongside them.

My family have been an important source of love and support, without them this would have been beyond difficult. I'd like to thank my mother and father, Linda and Ralph, and of course my brother, Alex and sister, Samantha. I am fortunate to have the love, companionship, and support of Kalina and Newton; both of whom had to deal with the hardships and complications of my and her own graduate school experience.

Contents

Acknowledgements	vii
1 The Epoch of Reionization	1
1.1 Introduction	1
1.1.1 Modern Cosmology	1
1.1.2 CMB Foregrounds	4
1.2 The 21cm Signal	4
1.2.1 Global 21cm Signal	5
1.2.2 21cm Fluctuations	7
1.2.3 Evidence for Reionization	9
2 Principles of 21cm Cosmology	12
2.1 Measurement Equation	13
2.2 Imaging	15
2.3 Imaging Power Spectrum	17
2.4 Delay Power Spectrum	19
2.5 21cm Cosmology Problems Addressed in this Thesis	20
2.5.1 Foregrounds and the EoR Window	20
2.5.2 RFI Contamination	21
2.5.3 Calibration	22
3 PAPER	24
3.1 Background	24

3.2	The Donald C. Backer Precision Array for Probing the Epoch of Reionization	25
3.3	Wide-band Delay Filtering for Foreground Avoidance	27
3.4	Foreground Subtraction Methodology	30
3.4.1	Model Catalog	31
3.4.2	Model Visibility Parameters	31
3.4.3	Excess High Delay Power in FHD Models	32
3.5	Data and Preprocessing	34
3.5.1	PAPER	34
3.5.2	MWA	35
3.6	Implementation on PAPER Data	35
3.6.1	Effect of Foreground Subtraction	38
3.6.2	Performance Near Band Edges	41
3.6.3	Improved measurements within the horizon	44
3.7	Hybrid Foreground Removal Technique	
	Applied to MWA Phase I	47
3.8	Conclusions	48
	Acknowledgments	52
4	Machine Learning for Improving Sensitivity through RFI Mitigation	53
4.1	Interference in Radio Observations	53
4.2	Method	55
4.2.1	DFCN Architecture	55
4.2.2	Data Preparation	59
4.2.3	Training Dataset	60
4.3	Evaluation	64
4.4	Conclusions	69
	Acknowledgments	73
5	HERA	74
5.1	Hydrogen Epoch of Reionization Array	74

5.2	Sky Calibration	75
5.3	HERA Data	76
5.4	Fast Holographic Deconvolution	77
	HERA Primary Beam Model	77
	Source Catalog	78
	Parameter Script	78
	Observational Data	78
5.4.1	Capabilities	79
5.4.2	Under the hood	79
5.4.3	Caveats	80
5.5	Sky Calibration	82
5.5.1	Runtime	82
5.5.2	Calibration Case Studies	82
	Fornax A: Bright Extended Off Zenith Source	84
	GLEAM J0200-3053: Point Source	86
5.6	HERA-52 Calibration Stability	88
5.6.1	Per Day Calibration Solution Waterfalls	88
5.6.2	Per Antenna Correlation	88
5.7	Calibration Results	91
5.8	Conclusion	94
6	21cm EoR Parameter Estimation	95
6.1	21cm Simulations Dataset	95
6.2	Modified Convolutional Neural Network	100
6.3	Parameter Recovery with an FCN	100
6.4	EKF Formalism	101
6.5	EKF-FCN Future Work	106
	Acknowledgments	106

7 Conclusion	107
7.1 The Future of 21cm EoR Cosmology	107
A Foreground Avoidance Signal Loss	109
A.1 Wide-band Deconvolution Filter Signal Loss	109
B Remedying Model Visibility Issues with Fast Holographic Deconvolution	112
B.1 Creating Model Visibilities with Fast Holographic Deconvolution	112

List of Figures

1.1	Planck CMB Temperature Angular Power Spectrum	2
1.2	Fiducial Global 21cm Signal	8
1.3	Gunn-Peterson trough measurements over redshift	11
2.1	Correlation diagram for a single pair of radio antennas	15
2.2	Diagram of the EoR Window	20
2.3	RFI Flags Delay Spectra Example	22
3.1	PAPER Dipole	26
3.2	Toy Model Source on Celestial Sphere	29
3.3	PAPER FHD Imaging of Fornax A	33
3.4	PAPER Power Spectrum Pipeline	36
3.5	PAPER visibilities through processing stages	39
3.6	PAPER-64 Power Spectra	40
3.7	Hybrid Foreground Avoidance Power Spectrum Results	42
3.8	Power spectrum improvements within the horizon	45
3.9	Hybrid Foreground Avoidance applied to MWA Observations	46
3.10	MWA power spectrum Improvements at band edges	49
4.1	RFI amplitude and phase example in HERA data	55
4.2	Deep Fully Convolutional Neural Network Architecture	61
4.3	HERA RFI occupancy across the passband	64
4.4	DFCN Receiver Operating Characteristic curve	67

4.5	RFI class examples	68
4.6	RFI flagging algorithm comparison on real HERA data	70
4.7	RFI flagging algorithm comparison on simulated HERA data	71
5.1	HERA-52 Antenna Layout	76
5.2	FHD simulation errors in HERA model visibilities	81
5.3	Fornax A Dirty, Model, and Residual images	85
5.4	Fornax A extended source model	86
5.5	GLEAM J0200-3053 Point Source Dirty, Model, and Residual images	87
5.6	Waterfall gain calibration solutions	89
5.7	Gain amplitude solutions over LST	90
5.8	HERA antenna correlation matrix	91
5.9	Antenna mode systematics	92
5.10	HERA calibrated, model, and residual waterfall visibilities	93
6.1	Simulated 21cm comoving redshift slices	98
6.2	21cm cube preparation for parameter estimation	99
6.3	Parameter estimation results using minimally trained FCN	102
6.4	Parameter estimation results with 500 epochs of training	103
A.1	Signal loss from delay filtering in PAPER-64 visibilities	111
B.1	FHD model visibility uv plane sampling	113
B.2	Delay transformed models that have been filtered	114

List of Tables

4.1	RFI DFCN per layer architecture	58
4.2	Network hyperparameter gridsearch result	60
4.3	RFI algorithm results per RFI class	66
4.4	RFI algorithm results on real and simulated data	70
5.1	FHD calibration parameters for HERA	83

List of Abbreviations

HI	Neutral Hydrogen
EoR	Epoch of R eionization
WMAP	W ilkinson M icrowave A nisotropy P robe
PAPER	P recision A rray for P robing the E POCH of R eionization
HERA	H ydrogen E POCH of R eionization A rray
LOFAR	L Ow F requency A rray
SKA	S quare K ilometer A rray
RFI	R adio F requency I nterference
GSM	G lobal S ky M odel
ROC	R eciever O perating C haracteristics
LST	L ocal S idereal T ime
MWA	M urchison W idefield A rray
FHD	F ast H olographic D econvolution
GLEAM	Ga Lactic and E xtragalactic A ll-sky M WA
CNN	C onvolutional N eural N etwork
DFCN	D eep F ully C onvolutional neural N etwork
ReLU	R ectified L inear U nit
EKF	E xtended K alman F ilter
GPU	G raphics P rocessing U nit
MCMC	M arkov C hain M onte C arlo
PSF	P oint S pread F unction

Chapter 1

The Epoch of Reionization

1.1 Introduction

1.1.1 Modern Cosmology

In its earliest moments, the Universe existed as a hot dense electron baryon plasma. The constant scattering between the electrons and photons created a tightly coupled baryon-photon fluid. Due to cosmic expansion this dense plasma began to cool, and the interaction rate between the photons and electrons decreased. Eventually expansion overcame the rate of interactions, allowing electrons and nucleons to combine and form neutral atoms (Peacock, 1999; Dodelson and Efstathiou, 2004). This epoch of recombination was followed shortly thereafter by light decoupling from matter and the Cosmic Microwave Background (CMB) emission was produced at a redshift of $z \sim 1100$ or approximately 378,000 years after the Big Bang. The CMB encoded an important signature of conditions during the early universe. Due to cosmic expansion the nearly isotropic CMB was downshifted in frequency into the microwave spectrum. Initial power spectra at the largest scale was measured by the COsmic Background Explorer (COBE) which found the power spectrum to vary on the order of 10^{-5} but was unable to pin down the spatial extent of these anisotropies (Smoot et al., 1992). The initial measurement of the CMB angular power spectrum provided strong confirmation of inflationary cosmology. Inflation is supported by the isotropy and

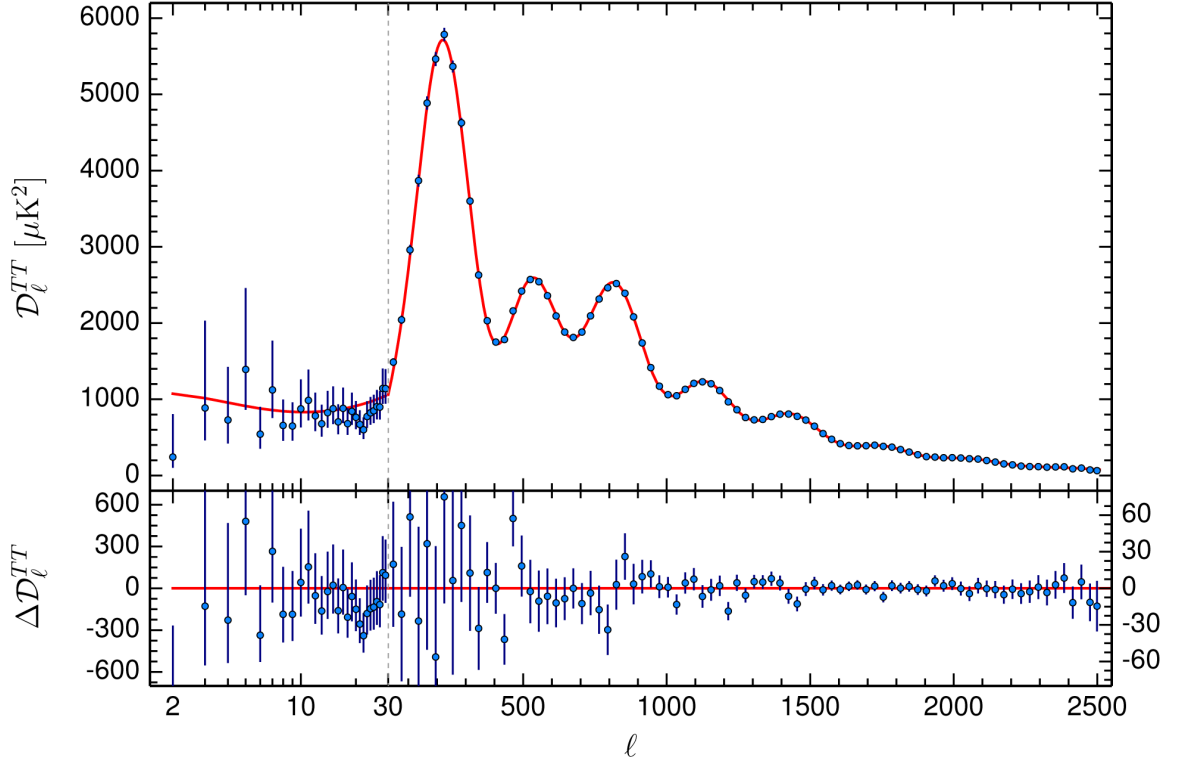


FIGURE 1.1: CMB temperature angular power spectra as measured by the Planck space telescope. The top plot shows the Planck measurements with $\pm 1\sigma$ uncertainties (blue) as compared to the best fit model (red). Uncertainties at the largest scales (low multipole modes ℓ) are dominated by cosmic variance. Oscillations at small angular scales, modes $\ell > 300$ are the signature of baryon acoustic oscillations. The bottom plot are the residuals as compared to the model. From Planck Collaboration et al. (2016).

homogeneity of the CMB temperature fluctuations which were amplified from quantum fluctuations present in the early universe. The inflation of space-time is consistent with large cosmological scales that contain the same characteristic temperature fluctuations, meaning that the initial conditions of the universe requires matter to have been in local proximity and therefore space-time must have undergone a rapid expansion very early on. Beyond these initial measurements, modern CMB angular power spectra have been performed with improved angular resolution allowing them to probe much smaller angular scales. These recent more sensitive space-borne instruments include the Wilkinson Microwave Anisotropy Probe (WMAP) (Bennett et al., 2013) and Planck (Planck Collaboration et al., 2018). An example of the CMB angular power spectrum as measured by Planck is shown in Figure 1.1.

Measurement of the CMB angular power spectrum shape and structure can then be used to constrain 6 important parameters of the Λ CDM model of the universe (Spergel et al., 2007). These parameters are, $\Omega_b h^2$ the physical baryon density, $\Omega_m h^2$ which is the physical matter density accounting for the combined baryonic and dark matter contributions, the Hubble constant H_0 (or its dimensionless version h), σ_8 the amplitude of fluctuations, n_s the scalar spectral index, and finally the optical depth of reionization, τ . The unitless parameter τ is a measure of the line-of-sight opacity to the CMB radiation of regions of intervening free-electrons. This can be expressed as

$$\tau = \sigma_T \int \bar{n}_e(z) \frac{d\ell}{dz} dz, \quad (1.1)$$

where $\bar{n}_e(z)$ is the average free-electron number density, σ_T the Thomson cross-section, z is the redshift, and $d\ell/dz$ is the line-of-sight proper distance per redshift (Griffiths, Barbosa, and Liddle, 1999). Observations today indicate that the Universe is mostly ionized, implying that at some point from recombination to now, there was a period of reionization. The occurrence of reionization implies an increase in $\bar{n}_e(z)$ and therefore more Thomson scattering. When following through with the integration of Equation 1.1 (See Liu et al. (2016)) and using some cosmological approximations such as a flat universe and instantaneous reionization¹ leads to

$$\tau \propto \frac{h\Omega_b}{\Omega_m} \left(\sqrt{\Omega_\Lambda + \Omega_m(1 + z_{\text{ion}})^3} - 1 \right), \quad (1.2)$$

with Ω_m being the matter density, Ω_Λ the dark energy density, Ω_b the baryon density, and z_{ion} the presumed redshift of reionization. What Equations 1.1 & 1.2 describe is that while z_{ion} can be constrained by the CMB power spectrum through measurement of cosmological parameters it still leaves a lot to the imagination. Fundamental limitations of what the CMB can probe in the large cosmological time frame between recombination and the present will prompt an exploration of alternative methods.

¹Which has been shown to be an incorrect assumption based on initial constraints from Bowman, Rogers, and Hewitt, 2008 which sets the duration of reionization to $\Delta z > 0.06$ and even more recent measurements in Monsalve et al., 2017, which rejects $\Delta z \leq 0.4$ at a significance of 2σ .

1.1.2 CMB Foregrounds

Getting to the sensitivity level for the results shown in Figure 1.1 required significant attention to foregrounds. I use the term foregrounds here to mean anything between the instrument and recombination that obscures the CMB observation, but foregrounds primarily consist of diffuse galactic emission from the Milky Way. Separating these diffuse galactic emissions employed foreground subtraction techniques such as frequency map differencing to create templates for wavelet decomposition (Fernández-Cobos et al., 2012), internal linear combination (ILC) in both a harmonic (Cardoso et al., 2008) and wavelet (Delabrouille et al., 2009) domain, and a Bayesian component separation technique using Gibbs sampling (Eriksen et al., 2008). Without these techniques the CMB power spectrum would be contaminated at some of the smallest scales that allow the extraction of cosmologically important parameters. With the help of these subtraction techniques and of course handling of telescope systematics, the CMB angular power spectrum is able to provide several key cosmological observations, namely the evidence for inflation, a spatial curvature measurement, acoustic oscillations, and the thermal history. It is in this context that I would like to motivate the need for the foreground avoidance techniques and strategies I introduce in Chapter 3.

1.2 The 21cm Signal

The most abundant element in the universe is hydrogen which dutifully places itself in corners of the universe you'd find most interesting for understanding cosmology and astrophysics. That means if it were possible to somehow harness an emission from the ubiquitous hydrogen it would be a very strong probe for peering into much of the universe. Luckily just that signal exists in the form of the 21cm hyperfine transition. The hyperfine transition is the occurrence of a spontaneous spin-flip of the electron spin relative to the proton spin from being parallel (higher energy) to anti-parallel (lower energy) or vice versa. The change in energy between these two states is $\Delta E_{10} = 5.9 \times 10^{-6} \text{ eV}$ which corresponds to an emission of a photon with a wavelength of $\lambda = 21 \text{ cm}$ (or frequency $f = 1420 \text{ MHz}$)

in its rest frame. The spontaneous transition rate of the 21cm spin flip is governed by the Einstein emission coefficient of $A_{10} \approx 2.8688 \times 10^{-15} \text{ s}^{-1}$ which gives a characteristic time of $t_{1/2} \approx 11 \text{ Myr}$. A single atom of neutral hydrogen (HI) is therefore unlikely to emit a 21cm signal in any reasonable timeframe, thus there must be a reliance on the probability of large quantities of HI to produce this signal.

1.2.1 Global 21cm Signal

The distribution of dark matter put in place by initial density perturbations from quantum fluctuations in the early universe set the stage for baryonic matter to begin to gravitationally collect itself. This clumping of normal matter into dark matter ‘pockets’ began roughly at $z > 30$ and initiated the formation of large scale structures as the primordial gas was too hot to virialize by itself (Tegmark et al., 1997; Barkana and Loeb, 2001). These first collapsed structures allowed Population III stars to form which were almost entirely devoid of metals. Soon after the first primordial galaxies formed and the universe consisted of the dark matter substrate which collected these galaxies and stars, situating them in dark matter halos interconnected by clouds of mostly HI that is called the Intergalactic Medium (IGM). These early stars and galaxies began emitting ionizing radiation which heated and reionized the surrounding HI in the IGM.

The spin temperature T_S , is defined by the population of HI atoms in either of the triplet (n_1) or singlet (n_0) energy states is related by

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} e^{-\Delta E_{10}/k_b T_S}, \quad (1.3)$$

which relies on the statistical weights of each state g_i ($g_1 = 3$ and $g_0 = 1$) and k_b is Boltzmann’s constant (Furlanetto, Oh, and Briggs, 2006). The brightness temperature defined in the Rayleigh-Jeans limit for this excited state along the line of sight through a cloud of neutral hydrogen is

$$T_b(\nu) = T_S(1 - e^{-\tau_\nu}) + T_{\text{CMB}}e^{-\tau_\nu} \approx T_S\tau_\nu + T_{\text{CMB}}(1 - \tau_\nu) \quad (1.4)$$

and depends on the optical depth of hydrogen (where $\tau_\nu \ll 1$) at a frequency ν and the background CMB brightness temperature (Pritchard and Loeb, 2012). By calculating the optical depth of 21cm photons through a cloud of hydrogen and taking the difference with the CMB temperature, the differential brightness temperature

$$\delta T_b \approx 9x_{\text{HI}}(1+\delta)(1+z)^{1/2} \left[1 - \frac{T_{\text{CMB}}}{T_S} \right] \left[\frac{H(z)/(1+z)}{dv_{\parallel}/dr_{\parallel}} \right] \text{mK}, \quad (1.5)$$

gives a relationship between several astrophysical and cosmological parameters such as the neutral fraction of hydrogen x_{HI} , the fractional overdensity of baryons $(1+\delta)$, contribution from the Hubble expansion at high redshift $H(z)/(1+z)$, and the gradient of the proper velocity along the line of sight $dv_{\parallel}/dr_{\parallel}$ (Furlanetto, Oh, and Briggs, 2006). From Equation 1.5 a dependence relative to the background CMB temperature T_{CMB} is readily apparent. This implies that when $T_S > T_{\text{CMB}}$ the 21cm signal is in emission, $T_S < T_{\text{CMB}}$ is an absorption, and $T_S = T_{\text{CMB}}$ the signal is unobservable. Thus the CMB provides a contrast between the 21cm signal from clouds of neutral hydrogen but this is entirely dependent on the nature of the spin temperature. The spin temperature behavior is driven by several external mechanisms with their relationship presented as

$$T_S^{-1} = \frac{T_{\text{CMB}}^{-1} + x_c T_K^{-1} + x_\alpha T_c^{-1}}{1 + x_c + x_\alpha}, \quad (1.6)$$

where x_c is the collisional coupling coefficient to surrounding gas, T_K the kinetic gas temperature, x_α the coupling coefficient for UV scattering from the Wouthuysen-Field Effect (Wouthuysen, 1952; Field, 1959), and T_c is the temperature of the UV radiation field. The coupling to the CMB, gas temperature, and through the Wouthuysen-Field effect to Ly- α radiation can provide a unique signal for observing and understanding the physics during these intermediate stages of the universe.

Initially at high redshift ($z > 200$), the average gas temperature $\bar{T}_K$ is thermally coupled to the background CMB temperature T_{CMB} through Compton scattering, this in turn couples T_S to T_{CMB} . During this period the global 21cm signal is unobservable ($\delta \bar{T}_b = 0$) and only becomes observable again when T_S is sufficiently decoupled by the gas becoming less

dense and cooling. As the gas adiabatically cools, $T_S = T_K$, through collisional coupling and produces an observable absorption feature with $\delta\bar{T}_b < 0$. This feature presents itself during the epoch known as the Dark Ages seen in Figure 1.2. Eventually the gas becomes too thin making collisions ineffective and T_S settles back into T_{CMB} . The first luminous sources then appear and begin producing $Ly\alpha$ photons leading to $Ly\alpha$ pumping in the surrounding gas. $Ly\alpha$ pumping also known as the Wouthuysen-Field Effect occurs when the UV radiation from the early stars and galaxies is absorbed and re-emitted by the neutral hydrogen. This allows the neutral hydrogen to become excited from a hyperfine state into a higher electronic state, and then decay back down into the opposite hyperfine state. Therefore the Wouthuysen-Field Effect allows for the bypassing of the normally forbidden hyperfine transition which arises from the dipole selection rules. The coupling to the $Ly\alpha$ background eventually saturates where $T_K = T_C = T_S$, producing a strong global absorption feature. At this point the earliest sources start producing X-rays which heats nearby gas, driving the kinetic temperature higher (Pritchard and Furlanetto, 2007). The heating process takes T_S with it, pushing the 21cm signal into an emission relative to the background CMB temperature ($\delta T_b > 0$). The global 21cm reionization signal is therefore highly dependent on the background radiation fields i.e. $Ly\alpha$, X-ray, and ionizing UV radiation. The relative importance of each of these fields is unknown, and depend on undetermined astrophysical parameters from the star formation history such as the spectral energy distribution and the escape fraction of ionizing photons (Barkana and Loeb, 2001; Barkana and Loeb, 2004; Pritchard and Furlanetto, 2007; Furlanetto, Oh, and Briggs, 2006). Measurements of the 21cm signal at these redshifts are thus required to constrain both the reionization epoch and the conditions during the formation of the early stars and galaxies.

1.2.2 21cm Fluctuations

The global 21cm signal can help constrain broad properties of the early Universe but to really unlock the underlying astrophysical processes it is required to look at the spatial fluctuations in the 21cm signal. The reliance on position for the spin temperature becomes apparent when looking back to Equation 1.5 with X_{HI} , δ , T_S , and $dv_{\parallel}/dr_{\parallel}$ all varying

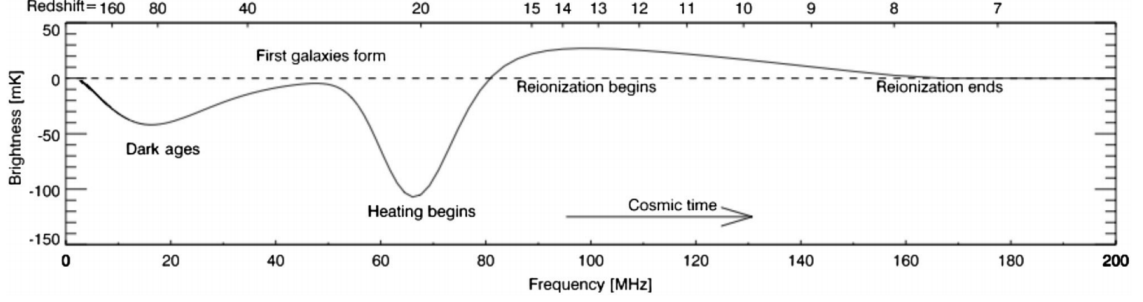


FIGURE 1.2: Differential brightness temperature $\delta\bar{T}_b$ of the global 21cm signal. During the dark ages the 21cm signal decouples from the CMB due to adiabatic cooling ($T_K < T_{\text{CMB}}$). Heating of the IGM after the first galaxies form drive $\delta\bar{T}_b$ from absorption to emission ($T_K > T_{\text{CMB}}$). The 21cm signal then remains an emission until the end of reionization. Figure taken from Liu et al., 2013

from one region to the next. From this it can be seen that the spatially inhomogeneous 21cm signal thus describes the variation between the relevant radiation fields and the state of the intergalactic medium (IGM). By measuring this signal, important characteristics of reionization can be teased out at many scales, as opposed to the average offered by the global 21cm signal. For example, this can be seen in the difference in scales probed by the fluctuations in the density field in contrast to the discrete placement of ionizing sources. To quantify the 21cm fluctuations a power spectrum can be defined for a cosmological volume at a specified redshift. The two-point correlation function for 21cm brightness temperature perturbations $\delta_{21} \equiv (\delta T_b(\mathbf{x}) - \delta\bar{T}_b)/\delta\bar{T}_b$ leads to the 21cm power spectrum

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

where $\sim$ is the Fourier transform, δ_D is the Dirac delta function, and $\mathbf{k}_1, \mathbf{k}_2$ are Fourier modes of the cosmological volume (Furlanetto, Oh, and Briggs, 2006). Luckily the 21cm fluctuation power spectrum can be naturally probed by radio interferometers, which I present as an introduction to in Chapter 2.

A more immediate benefit to probing the 21cm fluctuations compared to the global average is directly related to the non-gaussian nature during reionization (Furlanetto, Oh, and Briggs, 2006). The average 21cm signal during reionization progresses slowly and has a

smooth structure very similar to that of the bright Galactic and extragalactic foregrounds. The opposite is true for the fluctuations, which offer a stark contrast to these astronomical foregrounds and allow the 21cm small scale fluctuation signal to be disentangled. While this is extremely promising, there are issues with both removing these foregrounds and interferometer systematics, which I explore in Chapters 3 & 5. Regardless of these issues, observing the 21cm fluctuations will open a new astrophysically rich window into the reionization epoch with the advent of 21cm tomography.

1.2.3 Evidence for Reionization

The CMB angular power spectrum provides a measurement of the optical depth of reionization, which according to results from Planck Collaboration et al., 2018 places the midpoint of reionization at $z_{50\%} = 7.68 \pm 0.79$. But as stated in §1.1.1 this makes an assumption about reionization being an instantaneous process which is certainly not true (Bowman, Rogers, and Hewitt, 2008). Beyond the CMB, observational cosmologists have been able to use quasars as a limited probe into the end of the epoch of reionization.

Some of the strongest evidence of reionization comes from the spectroscopic measurements of high-redshift ($z > 5$) quasars. These quasars emit UV radiation that becomes redshifted as it travels through space and thus shifts into the Ly α resonance. As these Ly α photons are travelling they come into contact with clouds of neutral hydrogen. These intergalactic clouds of HI, also referred to as the IGM, strongly absorb Ly α . This means that a spectroscopic measurement of a high-redshift quasar should be able to detail the intervening space as Ly α photons get redshifted into resonance and absorbed by HI. Stemming from this effect there are several important features to be made aware of, specifically the Ly α forest (Rauch, 1998) and the Gunn-Peterson trough (Gunn and Peterson, 1965). The Ly α forest represents the limit where neutral hydrogen has become mostly reionized and absorption of Ly α photons is sporadic giving the appearance of a ‘forest’ of absorption lines. The Gunn-Peterson trough is the other extreme, where HI is still present and all spectroscopic measurements blueward of the Ly α resonance have nearly zero flux due to absorption. Evidence for a Gunn-Peterson trough is seen in Figure 1.3 where quasar

spectra beginning at $z = 6.42$ show nearly zero flux at low λ and a steep drop off at the $\text{Ly}\alpha$ resonance. Looking to lower redshifts ($z < 6$), the Gunn-Peterson effect begins to disappear and the $\text{Ly}\alpha$ forest becomes apparent implying that reionization has mostly ended and HI is not as predominant as it once was. This type of deep pencil beam survey however does not set a strong constraint on the end of reionization. While several spectroscopic measurements (Becker et al., 2001; Fan et al., 2006) have found additional evidence of the Gunn-Peterson trough, extrapolating the average HI fraction ($\bar{x}_{\text{HI}}$) from this to get the redshift for the end of the epoch of reionization would be highly model dependent. Recent surveys of this type have given the statistics of the $\text{Ly}\alpha$ opacity (Bosman et al., 2018) during the $4.7 < z < 6.2$ period and an attempt at understanding the Gunn-Peterson effect when it's found at lower redshifts $z < 5$ (Davies, Becker, and Furlanetto, 2018), all of which are important for furthering the constraint on reionization using quasar spectra.

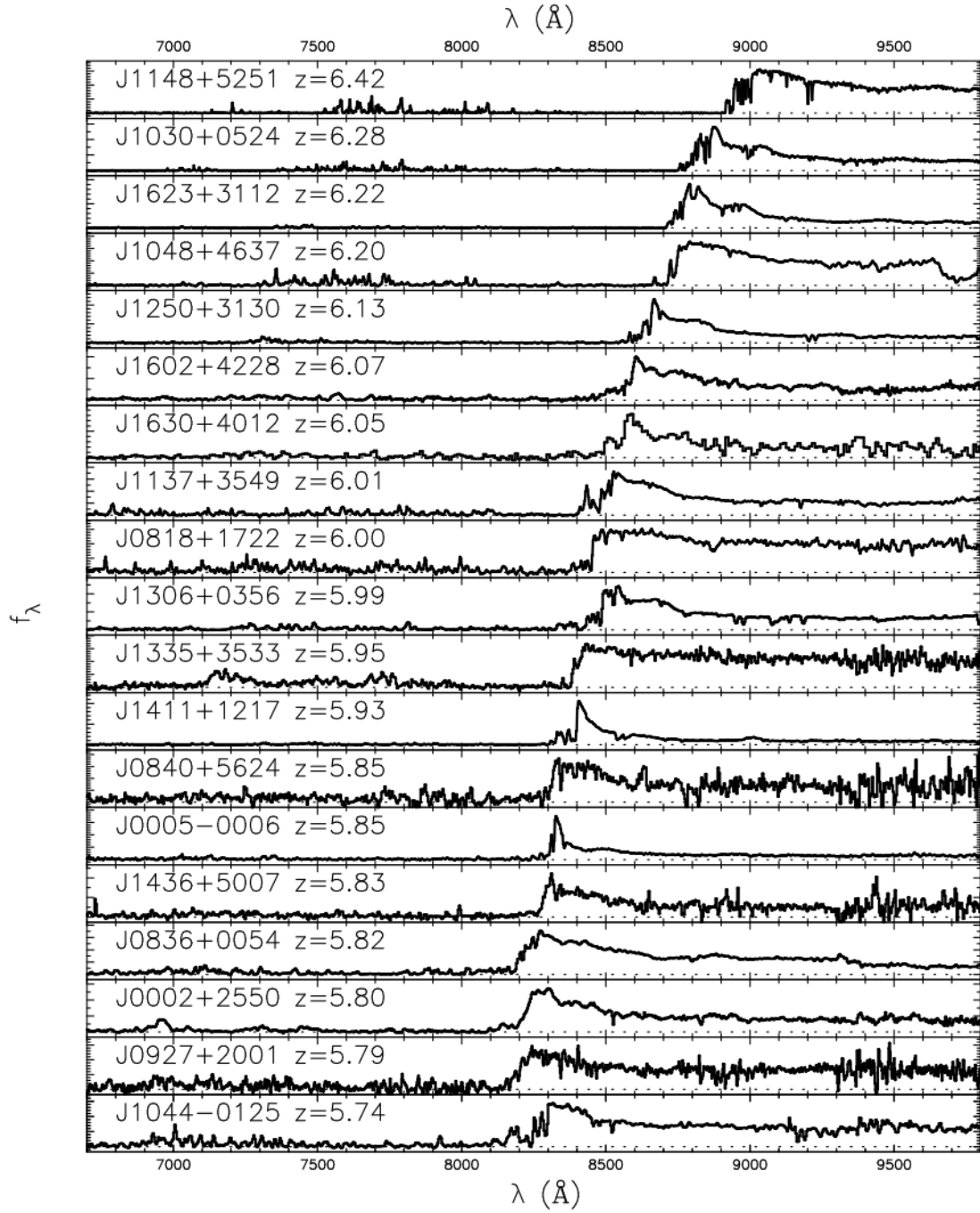


FIGURE 1.3: Flux f_λ of spectroscopic measurements of high-redshift quasars showing the absorption of UV radiation by neutral hydrogen from the intergalactic medium. The trough feature demonstrates that the amount of neutral hydrogen in the universe slowly disappears as you approach lower redshifts. Therefore the phase change of the universe from neutral to ionized has mostly completed by a redshift of ~ 6 . From Fan et al. (2006).

Chapter 2

Principles of 21cm Cosmology

Measurement of the 21cm signal from the Epoch of Reionization requires low-frequency radio telescopes. This is because the 21cm emission gets redshifted due to cosmological expansion which lengthens it to the order of meters. For the purpose of measuring the EoR, which is expected to have occurred at some point in the redshift range of $6 < z < 12$ (Zaroubi, 2013), $f_{\text{obs}} = f_{\text{rest}}/(1+z)$ can be used which gives the observation frequency f_{obs} at redshift z of an emission in its restframe frequency f_{rest} . This gives an approximate observing band of 100 - 200 MHz for a conservative estimate of the beginning and end of reionization. This places measurements solidly in the low frequency radio spectrum which comes with some benefits such as the ability to perform measurements from the ground, because Earth's atmosphere is transparent at these frequencies, but also with some pitfalls. A non-exhaustive list of these pitfalls¹ include Galactic and extragalactic foregrounds which obscure measurement of the 21cm EoR signal (See Chapter 3), radio frequency interference (RFI) (See Chapter 4), and precision calibration (See Chapter 5).

A radio interferometer is composed of an array of individual radio antennas (or dishes) that are designed to receive radio emissions within a finite band of frequencies. While interferometers are typically chosen for their ability to push the boundaries of angular resolution, this is not the case for EoR studies. For reference, a two element interferometer

¹Ionospheric distortions are also an ever-present concern and potentially rank near the same level of these issues but are not covered in this thesis (Datta et al., 2014).

has the angular resolution that is equivalent to $\delta\Theta = \lambda/|\vec{b}|$ in units of radians where $\vec{b} = \vec{x}_A - \vec{x}_B$ is the baseline vector from antenna A to antenna B, therefore giving the effective angular resolution of a single radio telescope with a diameter of $|\vec{b}|$. The main draw for using radio interferometry in cosmological studies is to achieve the most sensitivity to wide-field (large scale) emissions while remaining feasible cost and engineering wise. The sensitivity of a single radio antenna element to a source can be approximated using the radiometer equation

$$T_{\text{rms}} = \frac{T_{\text{sys}}}{\sqrt{\Delta\nu\tau}}, \quad (2.1)$$

where the system temperature T_{sys} is the combination of the receiver noise T_{rcvr} and the sky noise T_{sky} (Wrobel and Walker, 1999; Thompson, Moran, and Swenson, 2017). Depending on the observing frequency there is a trade-off between either sky noise or receiver noise dominating but for most cases at low frequencies $T_{\text{sys}} > T_{\text{rcvr}}$. The lower limit of the global 21cm EoR signal should be on the order of 10 mK, which with an effective sub-band of $\nu = 5$ MHz and a T_{sys} in the range of ~ 1500 K gives a total required observing time of about 75 minutes. So at first glance it appears that an EoR detection should be achievable with minimal effort, but in later chapters I introduce why this isn't true.

In the next section I will introduce the fundamental measurement equation which is crucial for understanding how radio interferometers observe the sky. In §2.2 I will provide an overview of transforming measurements from an interferometer with the understanding learned from §2.1 on how the radio sky is reconstructed into an image. §2.3 will explain how cosmological power spectra can be estimated from imaged visibilities. I will then present the formalism for estimating a cosmological power spectrum directly from visibilities without imaging in 2.4.

2.1 Measurement Equation

For every antenna j in an interferometer an incident plane EM wave induces a voltage $\mathcal{V}_j = \mathcal{V}_0 e^{-2\pi i \nu t}$ at time t and frequency ν . This signal is correlated with another antenna i and its voltage $\mathcal{V}_i = \mathcal{V}_0 e^{-2\pi i \nu (t + \tau_g)}$ which introduces a geometric time delay τ_g , due to the increased path the signal must traverse; this forms a baseline measurement of the radio

sky. A diagram of an antenna pair baseline is shown in Figure 2.1. Every measurement at some point on the celestial sphere corresponds to a measurement of a Fourier mode in a plane normal to this direction. The relationship between the delay measurements and the instrument relative to a source can be seen through $\tau\nu = \vec{b} \cdot \hat{s} / \lambda = \vec{u} \cdot \hat{s}$, where $\vec{u} = (u, v, w)$ is the Fourier conjugate of a position on the sky and $\hat{s} = (\ell, m, \sqrt{1 - \ell^2 - m^2})$ the direction to a source at that position. Therefore measurements of the sky exist as projections in the uv plane. The time-averaged correlation $\langle \cdot \rangle$ between the output voltages as measuring an extended source in a region of the sky Ω for antenna pair i, j can be described with

$$\begin{aligned}
V &= \langle \mathcal{V}_i(t + \tau) \mathcal{V}_j^*(t) \rangle \\
&= \left\langle \int \mathcal{V}_0 e^{-2\pi i \nu(t + \tau)} d\Omega \int \mathcal{V}_0^* e^{2\pi i \nu t} d\Omega \right\rangle \\
&= \int \int \langle |\mathcal{V}_0|^2 \rangle e^{-2\pi i \nu \tau} d\ell dm \\
&= \int \int A(\ell, m) \cdot I(\ell, m) \cdot e^{-2\pi i(u\ell + vm + w\sqrt{1 - \ell^2 - m^2})} d\ell dm,
\end{aligned} \tag{2.2}$$

where u, v are measurements in the uv -plane, ℓ, m direction cosines on the sky, $A(\ell, m)$ is the direction dependent response of the antenna primary beam, $I(\ell, m)$ is the source intensity distribution on the sky (Thompson, Moran, and Swenson, 2017). For a more complete derivation of the visibility equation's relation to the antenna pair correlation of a intensity distribution on the sky see Taylor, Carilli, and Perley, 1999. In general Equation 2.2 is simplified using the flat sky approximation where either $\ell, m \ll 1$ giving $\sqrt{1 - \ell^2 - m^2} = 1$ or $w = 0$ is assumed.

$$V(u, v) \approx \int \int A(\ell, m) \cdot I(\ell, m) \cdot e^{-2\pi i(u\ell + vm)} d\ell dm. \tag{2.3}$$

Equation 2.3 therefore gives a direct method for relating astronomical radio sources on the sky to the measured visibilities from a radio interferometer. For every one visibility measurement by an antenna pair baseline this directly measures the Fourier domain of uv -space where large angular scales are related to short baseline distances and vice versa. By taking many of these measurements in the Fourier domain a representation of the radio sky can be recreated.

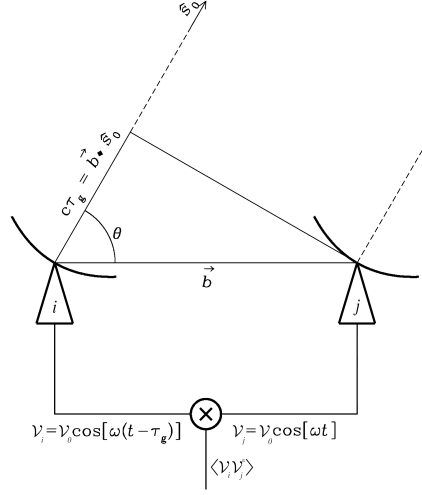


FIGURE 2.1: Diagram of the correlation $\langle \mathcal{V}_i(t + \tau_g) \mathcal{V}_j^*(t) \rangle$ between the measured voltages $\mathcal{V}_i, \mathcal{V}_j$ of the i, j antenna pair. The geometric delay τ_g is equal to the speed of light travel time over the baseline $\vec{b}$ relative to the direction $\hat{S}_0$ of a source on the sky. Modified from Condon and Ransom, 2016

2.2 Imaging

An array of N antennae have the ability to instantaneously measure the complex uv plane with $N(N-1)/2$ unique measurements. Each one of these measurements of $V(u, v)$ has a complex conjugate of $V^*(-u, -v)$ meaning that for every visibility measurement you're actually getting two because the sky $I(\ell, m)$, is real valued. Understanding how these measurements in uv space relate to the reconstruction of the sky leads to an important concept in radio interferometry imaging which is uv coverage. The amount of sampling that is performed of the uv -plane and the more completely sampled it is the closer the resulting image is to the true radio sky. Of course a radio interferometer is never capable of measuring the entire uv -plane because interferometers are insensitive to the global radio average which would be the origin of the uv -plane and would require an antenna pair with a baseline length of 0m. The 'sampling' function of the uv -plane is typically described as $S(u, v) = \sum_k \delta(u - u_k) \delta(v - v_k)$, where every u_k, v_k is the corresponding visibility measurement from a baseline antenna pair. This means a sampled uv -plane can now be turned into a very

simple reconstruction of the sky with

$$I^D(\ell, m) = \mathcal{F}^{-1}(S(u, v)V(u, v)) = b(\ell, m) * I(\ell, m), \quad (2.4)$$

where $I^D(\ell, m)$ is the ‘dirty’ sky image and $\mathcal{F}^{-1}$ is the inverse Fourier transform. Using the Convolution theorem this sampling can also be described as the true sky brightness distribution $I(\ell, m)$ convolved with the synthesized beam $b(\ell, m)$ which is determined by the layout of the array. Because it’s impossible to completely measure the uv -plane, the sampling of these points must be gridded to a regular grid so an inverse Fourier transform can be performed using the Fast Fourier Transform (FFT) algorithm. This resultant image is unfortunately not of the best quality but several techniques have been developed to improve upon the image from a radio interferometer. More specifically, weighting of the measurements in the uv -plane has offered the ability to express certain characteristic scales on the sky and is introduced simply with $S(u, v) \rightarrow W(u, v)S(u, v)$ where $W(u, v)$ is the weighting function. Some weighting schemes include uniform weighting which is simply unit weighting applied to all gridded uv samples, this gives great angular resolution however poor signal to noise. Another is natural weighting $W(u, v) = \frac{1}{N(u, v)}$ which weights every uv grid by the number of samples $N(u, v)$ contained within it. In any interferometer short baselines will always be more numerous, resulting in great signal to noise and poor angular resolution due to the emphasis on measurements closest to the uv -plane origin. For all imaging performed in this thesis I use uniform weighting as most of the arrays used for these analyses sparsely measure small angular scales (long baselines) due to their compactness.

Beyond weighting schemes there exist deconvolution algorithms for improving imaging. Deconvolving point or extended sources in a reconstructed sky image however can be quite complicated and because the original convolution is performed with missing data, the deconvolution ends up being an extrapolation over the uv -plane. Of these deconvolution algorithms is most notably the Högbom CLEAN algorithm (Högbom, 1974) which can be described as

$$\gamma \text{Max} \left[F^{-1} \left(S(u, v) \cdot (v_{cal} - \sum_m v_m) \right) \right], \quad (2.5)$$

where the calibrated visibility v_{cal} and model visibility v_m is subtracted, and gridded using $S(u, v)$ aka ‘the dirty beam’. The inverse Fourier transform is then performed bringing the residual into the image domain where the max pixel value is identified, multiplied by some gain $\gamma < 1$, and is added to the model visibility. This is performed iteratively until all point sources have been deconvolved from the image. The result is then ideally a good approximation to the sky sans the dirty beam. This algorithm has been improved upon since the original implementation of the CLEAN algorithm through the Clark CLEAN algorithm (Clark, 1980), A-projection (Bhatnagar et al., 2008), and Forward modeling (Bernardi et al., 2011) for example.

2.3 Imaging Power Spectrum

Using the understanding of the visibility equation from Equation 2.2 the interferometer’s Fourier relationship with the sky ($u, v, \eta \leftrightarrow \ell, m, f$) (Born and Wolf, 1999) can be translated to cosmological coordinates x, y, z and their Fourier conjugates of k_x, k_y, k_z which represent cosmological scales. This is realized through the relationship

$$2\pi(u\ell + vm + \eta f) = \frac{xk_x}{X} + \frac{yk_y}{Y} + \frac{zk_z}{Y}, \quad (2.6)$$

where f is the measured frequency, η is the Fourier conjugate of frequency, and X,Y are cosmological conversions between angular to comoving transverse scales and frequency to comoving line-of-sight scale respectively, with both being defined in Equations 2.7 and 2.8 (Furlanetto, Oh, and Briggs, 2006; Parsons et al., 2012b). It should be noted that the mapping between $\nu \leftrightarrow k_z$ is a unique result of the 21cm line and is not applicable to radio interferometry in general. The transverse cosmological modes k_x, k_y are typically referred to as $k_\perp = \sqrt{k_x^2 + k_y^2}$ and k_z is interchangeable with the cosmological line-of-sight mode $k_\parallel$.

$$X = 6.5 \times 10^3 \sqrt{\frac{0.27}{\Omega_m}} \left(\frac{1+z}{10} \right)^{0.2} h^{-1} \text{Mpc rad}^{-1}. \quad (2.7)$$

$$Y = 1.7 \times 10^{-2} \sqrt{\frac{0.27}{\Omega_m}} \sqrt{\frac{1+z}{10}} h^{-1} \text{Mpc GHz}^{-1}. \quad (2.8)$$

With these equations in mind, there is now a way to naturally relate interferometer space to a cosmological space. This implies that observations² through a radio interferometer can be translated into 3D cosmological cubes that can be used to measure the EoR power spectrum (Zaldarriaga, Furlanetto, and Hernquist, 2004; Morales and Hewitt, 2004). An imaging power spectrum estimate is performed quite simply by taking the Fourier transform (represented by $\sim$) in both the transverse direction of the reconstructed sky (ℓ, m) and along the frequency direction (f), going from $I^D(\ell, m, f) \rightarrow \tilde{I}^D(u, v, \eta)$. The image based power spectrum estimate $\hat{P}_{21}$ as derived from the transformed dirty sky image $\tilde{I}^D$ is

$$\langle |\tilde{I}^D(u, v, \eta)|^2 \rangle = \int \hat{P}_{21}(u, v, \eta) |\widetilde{W}(u - u', v - v', \eta - \eta')|^2 du' dv' d\eta', \quad (2.9)$$

where $\widetilde{W}(u, v, \eta)$ is the Fourier transform of the sky based window function $W(\ell, m, f)$, which is determined by the field of view and chosen bandwidth (Morales and Hewitt, 2004). The power spectrum estimate can then be related to the cosmological spatial modes with the result being a 3D power spectrum estimate of $\hat{P}_{21}(\vec{k})$.

While an image based power spectrum is fairly straightforward in its implementation and is currently in use by the MWA (Beardsley et al., 2016) and the LOw Frequency ARray (LOFAR) (Patil et al., 2017), there are some drawbacks. The sensitivity of the resultant power spectrum is dependent on the quality of the sky calibration. Then there is the issue of taking the Fourier transform over the uv plane, which mixes all of the modes measured by the interferometer allowing for the possibility of including bad data. Additionally, there is the issue of frequency dependence in the uv plane which adds additional spectral components when taking the Fourier transform along f . In recent years an alternative method for estimating the EoR power spectrum has been developed which directly measures the power spectrum from individual baseline visibilities (Parsons et al., 2010; Parsons et al., 2012a; Ali et al., 2015). The formalism for this power spectrum estimation is described in the next section.

²Neglecting issues such as the light cone effect along $k_{||}$ (Datta et al., 2012), and redshift space distortion (Mao et al., 2012).

2.4 Delay Power Spectrum

Current interferometers lack the ability to detect the EoR in the image domain due to mode-mixing and being more sensitive to calibration errors; altogether reducing the sensitivity to the EoR signal (Morales et al., 2019). Other instruments such as the Precision Array for Probing the Epoch of Reionization (PAPER) and the Hydrogen Epoch of Reionization Array (HERA) have opted for constraining the EoR signal through a statistically measured power spectrum. Beginning with Equation 2.3 which is how an interferometer sees the radio sky, a relationship can be built for probing the power spectrum of cosmological Fourier modes. The power spectrum in non-cosmological units is simply

$$\begin{aligned} \tilde{V}(u, v, \tau) \cdot \tilde{V}^*(u, v, \tau) \approx & \int \int A(\ell, m, \nu) A(\ell, m, \nu') I(\ell, m, \nu) I(\ell, m, \nu') \\ & \times e^{-2\pi i(u(\ell-\ell') + v(m-m') + \nu(\tau-\tau'))} d\ell d\ell' dm dm' d\nu d\nu', \end{aligned} \quad (2.10)$$

which can be used to convert the brightness distribution on the sky $I(\ell, m, \nu)$ to a brightness temperature using the Rayleigh-Jeans low frequency limit of $I(\ell, m, \nu) = 2k_B T(\ell, m, \nu)/\lambda^2$. This leads to

$$\begin{aligned} |\tilde{V}(u, v, \tau)|^2 \approx & \left(\frac{2k_B}{\lambda^2} \right)^2 \int \int A(\ell, m, \nu) A(\ell, m, \nu') T(\ell, m, \nu) T(\ell, m, \nu') \\ & \times e^{-2\pi i(u(\ell-\ell') + v(m-m') + \nu(\tau-\tau'))} d\ell d\ell' dm dm' d\nu d\nu', \end{aligned} \quad (2.11)$$

where it can be noted that according to Parsons et al., 2012a an estimate of the power spectrum is $\hat{P}(\vec{k}) = \langle |\tilde{T}(\vec{k})|^2 \rangle = \int \hat{\zeta}(\vec{r}) e^{-i\vec{k} \cdot \vec{r}} d^3r$, where $\hat{\zeta}$ is an estimate of the two-point correlation function of the form $\hat{\zeta}(\vec{r}) = \int T(\vec{x}) T(\vec{x} + \vec{r}) d^3x$. By restricting the antenna response $A(\ell, m, \nu)$ to a finite width of $0 \leq \ell, m \leq \theta$, the frequency sub-band to B , and a change of coordinates from $(\ell - \ell', m - m', \tau - \tau') \rightarrow (\ell_r, m_r, \tau_r)$ yields

$$\begin{aligned} |\tilde{V}(u, v, \tau)|^2 \approx & \left(\frac{2k_B}{\lambda^2} \right)^2 \theta^2 B \int \hat{\zeta}_{21}(u, v, \tau) \times e^{-2\pi i(u\ell_r + vm_r + \nu\tau_r)} d\ell_r dm_r d\nu_r \\ & = \left(\frac{2k_B}{\lambda^2} \right)^2 \frac{\theta^2 B}{X^2 Y} \int \hat{\zeta}_{21}(u, v, \tau) \times e^{-2\pi i(u\ell_r + vm_r + \nu\tau_r)} d^3r, \end{aligned} \quad (2.12)$$

which allows the conversion of sky and frequency coordinates to cosmological transverse and line-of-sight distances using Equations 2.7 and 2.8. The result of these coordinate

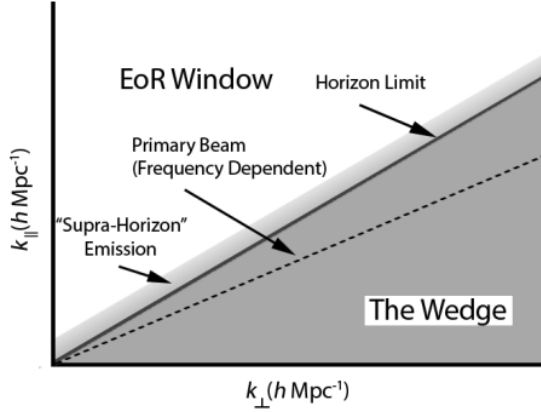


FIGURE 2.2: An illustration of the EoR window in 2D cylindrical k -space. The wedge (modes below the horizon limit) are dominated by bright Galactic and extragalactic foreground emissions. The chromatic response of the primary beam adds additional complexities at the supra-horizon as it pushes foreground power into the EoR window (Thyagarajan et al., 2016a). Taken from Pober et al., 2014.

conversions is the relationship

$$|\tilde{V}_{21}(u, v, \tau)|^2 \approx \left(\frac{2k_B}{\lambda^2}\right)^2 \frac{\theta^2 B}{X^2 Y} \hat{P}_{21}(\vec{k}), \quad (2.13)$$

which provides an estimate of the 21cm power spectrum from the Fourier transform of an interferometric visibility. The 21cm EoR power spectrum is analyzed in Chapter 3 as a method to quantify sensitivity improvements in terms of galactic and extragalactic foreground mitigation.

2.5 21cm Cosmology Problems Addressed in this Thesis

2.5.1 Foregrounds and the EoR Window

A more extensive introduction to foregrounds along with their avoidance strategies within the picture of the EoR window and horizon limit is addressed in Chapter 3, but for this section I outline the general concept of the EoR window. Working from the result in §2.4, the power spectrum can be cylindrically binned into a 2D k -space. This 2D power spectrum k -space is represented by the transverse modes of $k_\perp$ and the line-of-sight mode of $k_\parallel$. A natural consequence of observing with an interferometer and the nature of foregrounds when

moving to the $(k_{\perp}, k_{\parallel})$ -space is the ideal separation between foregrounds and the 21cm EoR signal. This result is what is commonly referred to as the “EoR Window” as shown in Figure 2.2. Due to the spectral smoothness of foregrounds which limits emissions from the Milky Way and extragalactic sources to low $k_{\parallel}$ modes along with the delay of these sources on the sky depending on the length of the baseline observing them; there is a ‘wedge’ of emissions that increasingly pushes foregrounds to higher $k_{\parallel}$ modes as you get to larger $k_{\perp}$ modes. This relationship between $k_{\parallel}$ and $k_{\perp}$ is found to be

$$k_{\parallel,hor} = \left(\frac{1}{\nu} \frac{X}{Y} \right) k_{\perp} \quad (2.14)$$

(Pober et al., 2014), where the maximum $k_{\parallel}$ at some $k_{\perp}$ is referred to as the horizon limit $k_{\parallel,hor}$. It would be ideal to measure the 21cm EoR signal at these low $k_{\parallel}, k_{\perp}$ modes because the EoR signal is brightest at the largest scales here, but most foreground avoidance strategies use a ‘scorched Earth’ approach to foregrounds, removing both EoR and foreground signals. The concept of the EoR window provides an information dense approach to understanding the interplay between the interferometer, the EoR signal, and foreground contamination.

2.5.2 RFI Contamination

21cm Radio Cosmology differs from CMB Cosmology in some very distinct ways, most notably being that the observing band at high frequencies ($10^1 - 10^3$ GHz) for CMB measurements is much more quiet than at lower frequencies ($10^0 - 10^2$ MHz). And with that limitation comes some unique challenges such as anthropogenic Radio Frequency Interference (RFI). Common RFI contaminants that must be avoided for 21cm cosmological studies originates from any naturally occurring non-astrophysical source or human made device that transmits near or within a bandwidth of 100 to 200 MHz. This includes sources such as lightning, airplanes, orbital satellite communications, radio and television broadcasts, and hand-held radios. RFI directly contaminates visibility measurements but is usually confined to a few frequency channels. It becomes a nuisance however in the delay spectrum and therefore in EoR power spectrum measurements and presents itself in two distinct ways.

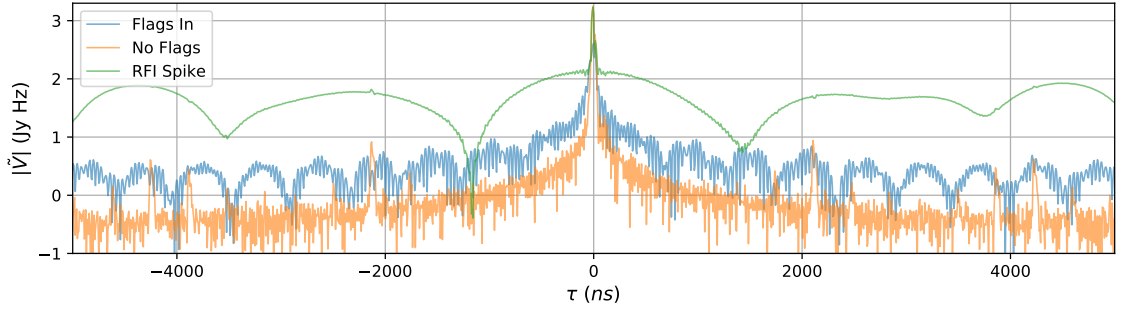


FIGURE 2.3: Delay spectra demonstrating the contamination from RFI present in a single baseline visibility simulated using FHD. The ideal delay spectrum (orange), the same delay transformed visibility with an RFI spike present (green), and the indirect example of delay transforming with RFI flags present (blue).

In the most direct case the power spectrum becomes overpowered by the introduction of bright RFI in observations. In the other case RFI indirectly causes leakage into higher delay modes when time-frequency channels are flagged (zeroed). This creates a complicated sampling function $W(t, f)$ which introduces itself as a ‘spectral floor’ in delay space. An example of each one of these cases is demonstrated in Figure 2.3 using a simulated model visibility from HERA. The RFI spike (green curve) shows how detrimental leaving RFI in the visibilities is to measuring the EoR signal because of the significant amount of leakage at all delay modes. It’s also important to point out that the RFI spike in this example has an SNR of 2, which is well below the average SNR of RFI in HERA observations. In Chapter 4 I go more in depth on the topic of RFI in HERA data by developing a novel approach for identifying RFI using machine learning.

2.5.3 Calibration

Calibration of radio interferometry data serves to remove effects from the instrument and to properly relate the overall measurement to a physical sky. This process is heavily dependent on proper processing of each visibility measurement to account for malfunctioning antennae or RFI contamination. For an imaging based calibration, both of these issues can seriously impede the calibration process when calibrating against a model of the sky. In comparison to sky calibration, is redundant calibration, which makes no reference to a sky model and

requires a subsequent absolute calibration for scaling to a physical sky. This technique was developed and deployed on arrays such as PAPER (Parsons et al., 2012b; Ali et al., 2015), which required a high degree of redundancy and was also explored on non-redundant arrays such as the Murchison Widefield Array (MWA) (Li et al., 2018). Redundancy has the benefit of improved sensitivity because the array is simultaneously measuring many of the same Fourier modes on the sky. However, one of the primary assumptions of redundant calibration is that every antenna in the array is operating identically and that visibilities with the same baseline distance and orientation are in fact redundant. While this is a good approximation, it is generally not true. The HERA calibration pipeline uses redundant calibration followed by an absolute calibration with a minimal sky model and no knowledge of the HERA primary beam. In Chapter 5 I demonstrate a wide-field sky calibration on a preliminary 18 night HERA observation dataset using a more complete sky and HERA primary beam model.

Chapter 3

PAPER

3.1 Background

Galactic and extragalactic foregrounds dominate the cosmological signal in 21cm EoR experiments, and must be removed if a detection of the 21cm signal is to be achieved. Several foreground mitigation techniques have been applied independently to arrays such as PAPER, MWA, and LOFAR (see e.g. Liu and Tegmark 2011; Vedantham, Udaya Shankar, and Subrahmanyan 2012a; Parsons et al. 2012a; Dillon, Liu, and Tegmark 2013; Wang et al. 2013; Parsons et al. 2014; Liu, Parsons, and Trott 2014a; Liu, Parsons, and Trott 2014b; Dillon et al. 2015; Pober et al. 2016; Trott et al. 2016). Some novel examples include a non-parametric GMCA wavelet decomposition (Chapman et al., 2013) used by LOFAR (Patil et al., 2017) and an SVD technique that is shown to isolate smooth spectrum foregrounds in GMRT visibilities while making no a priori assumptions about their form (Paciga et al., 2013).

The point of reference for this chapter begins with the foreground filtering technique used in Ali et al. (2015) on PAPER-64 observations. This filtering technique uses a wide-band iterative deconvolution filtering approach for removing bright foregrounds, which I describe in more detail in Section 3.3. This filtering however comes at the cost of losing

all EoR information at the lowest Fourier modes where foregrounds reside.¹ Foreground subtraction is a method employed by both the MWA and LOFAR (Jacobs et al., 2016; Beardsley et al., 2016; Patil et al., 2017), where one simulates visibilities corresponding to an accurate foreground model and subtracts them from the data. Unlike filtering, this method should ideally retain EoR information at all Fourier modes.

The motivation behind integrating the foreground subtraction and filtering techniques into a single power spectrum pipeline is to further reduce band-limited spectral leakage² of foreground power into high cosmological line-of-sight Fourier modes, $k_{\parallel}$, beyond what is capable with a filter alone (Parsons et al., 2012a; Vedantham, Udaya Shankar, and Subrahmanyam, 2012b; Morales et al., 2012; Pober et al., 2013). Analogously with a time series of finite data, a windowing function can be applied to reduce spectral leakage. This provides a reasonable way to mitigate any power seeping into neighboring bins in frequency space when taking the Fourier Transform, but it comes at the expense of reduced sensitivity and correlated measurements. Subtracting a foreground model can reduce the dynamic range required in the frequency Fourier transform to $k_{\parallel}$, mitigating the need for more aggressive windowing functions (e.g. a Blackman-Harris convolved with itself, as in Thyagarajan et al. 2016b).

3.2 The Donald C. Backer Precision Array for Probing the Epoch of Reionization

The PAPER array described in this chapter was composed of a 64 element drift-scan array³ situated in the Karoo Desert, South Africa. PAPER was designed with the pioneering concept of redundancy for both calibration and sensitivity purposes (Parsons et al., 2012b).

¹This filtering technique is widely associated with the terminology of “foreground avoidance.” The goal is to keep high $k_{\parallel}$ modes (the so-called “EoR window”) of the 21cm power spectrum clear of foreground contamination, at the cost of indiscriminately removing power from low $k_{\parallel}$ modes. Hence, these low $k_{\parallel}$, foreground contaminated modes are “avoided.”

²Commonly in Digital Signal Processing, spectral leakage is associated with power contaminating neighboring bins in frequency when Fourier transforming from a time series to frequency spectrum. In the work outlined here I refer to spectral leakage as contamination in nearby delay (τ) bins when transforming from a frequency space.

³PAPER had several build-out stages where it took data as an 8 (Parsons et al., 2010), 32 (Parsons et al., 2014), and 128 element array, however none of the results presented here are from those stages.



FIGURE 3.1: A single dual-polarization dipole antenna from the Donald C. Backer Precision Array for Probing the Epoch of Reionization. The wire mesh reflectors decrease the response to sources from the horizon. The stage of the PAPER array used for the analysis in this chapter was composed of 64 of these dipole antenna.

This meant that its array layout was rectilinear so as multiple baseline measurements measured identical Fourier modes on the sky simultaneously. The construction of each antenna element consisted of dual-polarization sleeved dipoles. Wire mesh reflectors were attached at 45° angles to reduce sensitivity to the horizon which slightly decreases the field-of-view to being just below horizon to horizon ($\sim 160^\circ$) (Parsons et al., 2010). An example of a single PAPER antenna element can be seen in Figure 3.1. The observing frequencies covered the wide band of 100 - 200 MHz which correspond to the redshift range of $\sim 6 < z < 13$.

3.3 Wide-band Delay Filtering for Foreground Avoidance

To understand the wide-band delay filtering method employed in PAPER I begin with an explanation of the delay spectrum approach. I use the alternative form of the standard radio interferometer visibility equation

$$\begin{aligned} V_b(\nu) &= \int dl \, dm \, A(\nu, l, m) I(\nu, l, m) e^{-2\pi i(ul+vm)} \\ &= \int dl \, dm \, A(\nu, l, m) I(\nu, l, m) e^{-2\pi i\nu\tau_g}, \end{aligned} \quad (3.1)$$

where $V_b(\nu)$ is the visibility of baseline b at observing frequency ν , $A(\nu, l, m)$ is the antenna response, $I(\nu, l, m)$ is the specific intensity distribution on the sky, (l, m) are the direction cosines on the sky, and (u, v) are the projected lengths of the baseline in units of wavelength. This form of the interferometer visibility equation is especially useful for demonstrating the delay spectrum approach because I can associate sources on the sky with a geometric delay, τ_g . This geometric delay

$$\tau_g = \frac{\vec{b} \cdot \hat{s}}{c} = \frac{1}{c}(b_x l + b_y m) \quad (3.2)$$

corresponds to the light travel time distance between two antennas for an emission from the direction $\hat{s}$, and directly relates to the baseline lengths $\vec{b} = (b_x, b_y)$ and the visibility domain coordinates with $\vec{u} = \nu\vec{b}/c$. By then applying a window function $W(\nu)$, and taking the Fourier Transform over frequency of Equation 3.1,

$$\begin{aligned} \tilde{V}_b(\tau) &= \int_{-\infty}^{\infty} d\nu W(\nu) V_b(\nu) e^{-2\pi i\tau\nu} \\ &= \int_{-\infty}^{\infty} d\nu W(\nu) \left[\int dl \, dm \, A(\nu, l, m) I(\nu, l, m) e^{-2\pi i\nu\tau_g} \right] e^{2\pi i\nu\tau} \\ &= \int dl \, dm \left[\tilde{W}(\tau) * \tilde{A}(\tau, l, m) * \tilde{I}(\tau, l, m) * \delta(\tau - \tau_g) \right]. \end{aligned} \quad (3.3)$$

I have what is called the Delay Transform, as found in Parsons and Backer (2009). This means that point sources on the sky can be mapped from the visibility to a geometric delay in delay space. Flat spectrum sources are constrained to a maximum geometric delay, τ_h ,

the horizon limit.⁴ This maximum delay is achieved when the direction from the baseline to source on the sky, $\hat{s}$, is parallel to the baseline $\vec{b}$, which means the source is located at the horizon. When measuring the relatively weak 21cm EoR signal, spectral leakage when taking the Delay Transform is significant. The foreground dominated delays at $|\tau| \leq |\tau_H|$ can be approximately 10^5 larger in amplitude than the EoR modes being contaminated at $|\tau| > |\tau_H|$.

I also want to be able to associate the geometric delays with cosmological scales which I can do with

$$\begin{aligned} k &= \sqrt{k_{\parallel}^2 + k_{\perp}^2} \\ &\approx \frac{2\pi H(z)}{\lambda(1+z)} |\tau| \end{aligned} \tag{3.4}$$

that directly relates the cosmological line of sight $k_{\parallel}$ and its perpendicular $k_{\perp}$, where λ is the band center wavelength, z is the redshift of the observation, c is the speed of light, and $H(z)$ is the Hubble parameter. For the case of moving between delays and k space for PAPER-64 I can approximate $k_{\perp} \approx 0$ because of the short baselines involved; this is not true for the MWA analysis which spans baselines comparable to and much longer than PAPER-64.

A toy model demonstrating an example of spectral leakage can be seen in Figure 3.2, where a single point source spectrum near the horizon is shown in visibility space (top) at full intensity (blue) and reduced by 60% (green). The delay spectrum (bottom) shows the full and reduced intensity point source. Both have sidelobes that contaminate higher delays, but the reduced source has a proportionally reduced spectral leakage in delay space. Thus, even a partial foreground subtraction can free up additional sensitivity to k -modes where one is more likely to make a statistical EoR detection.

The approximation of foreground sources being confined within a maximum delay, τ_h , gives a natural motivation to filter out low delays. This filtering removes the foregrounds that occupy low delay modes, and in turn will reduce spectral leakage of foreground power

⁴The intrinsic spectrum of the source expands its footprint in delay space beyond that of a delta function, but smooth spectrum sources (like foregrounds) will still have compact footprints and limited extent beyond the baseline's delay horizon, τ_h , which is a direct consequence seen from Equation 3.3.

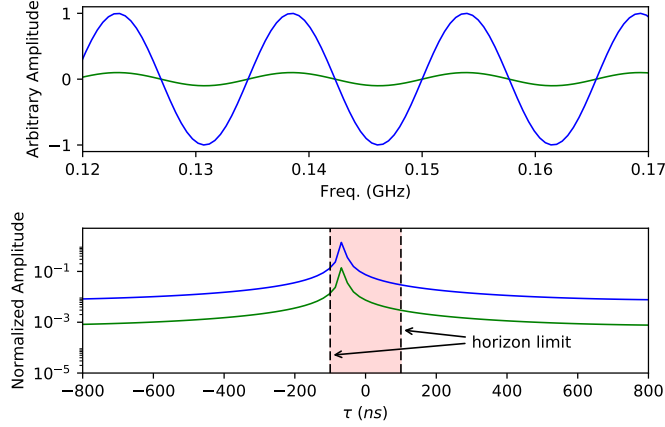


FIGURE 3.2: A toy model example of a single source near the horizon as it appears in a baseline visibility amplitude (top) and its delay transform (bottom), demonstrating how it appears near the horizon limit. The green line represents a 60% partial removal of the blue point source. Even the partial subtraction of a source can produce a non-negligible reduction in spectral leakage to delay bins outside of the horizon limit.

into modes beyond τ_h . However, due to discontinuities in the visibility sampling function (e.g. missing data from radio frequency interference (RFI) and a finite band i.e. convolution with a top-hat window function), foreground power is scattered to high delays and thus a simple delay-space filter is insufficient for mitigating spectral leakage (Parsons and Backer, 2009; Parsons et al., 2012a; Parsons et al., 2014). The delay filtering in the PAPER-64 power spectrum pipeline uses the application of a per baseline and per time wide-band deconvolution filter that iteratively fits components in delay space to remove foregrounds. I describe how the wide-band delay filter is implemented here in a succinct manner so as to motivate the foreground subtraction approach for mitigating any spectral leakage from this process. A more in-depth explanation of the filtering approach is outlined in Parsons et al. (2012a).

The deconvolution process works in delay space. Since the sampling function in frequency is known, one can calculate the delay-space convolving kernel created by the sampling. The algorithm then simply finds the brightest delay mode and subtracts a fraction of its power, accounting for the convolving kernel that correlates it with all other delay

modes.⁵ The range of modes that are fit in delay space can be restricted to be within the horizon limit (or, typically, a small buffer of 15 ns beyond the horizon limit) so that no signal is removed from the higher delays. This process is iterated over until a predetermined foreground residual threshold is met. This threshold is determined by the ratio of the initial amplitude to the final amplitude, which for the purpose of demonstrating this technique I use 10^{-9} . I choose 10^{-9} because it provides the minimum amount of worst-case signal loss (<5% signal loss) at the modes nearest the horizon (Ali et al., 2015) and deconvolves the maximum amount of the sampling function while still converging. Decreasing the order of this threshold or using a more sparse sampling/windowing function can lead to the algorithm’s inability to converge giving results with reduced filtering. I provide additional statistics in Appendix A for understanding how both a Blackman-Harris or Top-hat window function with various thresholds affect the signal loss when applying the iterative wide-band deconvolution filter.

3.4 Foreground Subtraction Methodology

Foreground subtraction is performed by forward modeling a source catalog to match observations. This model can come from an external catalog, or can be created from the data itself, typically through a source-finding algorithm in image space. Because of its limited imaging capabilities and poor PSF, it is inadvisable to use data from PAPER to create a foreground model. Rather, I use the GaLactic and Extragalactic All-sky Murchison Wide-field Array (GLEAM; Hurley-Walker et al. 2017a), described in Section 3.4.1. To generate and subtract model visibilities, I use the Fast Holographic Deconvolution (FHD; Sullivan et al. 2012) software package. The specifics of the visibilities generated for this work are described in Section 3.4.2.

⁵In previous publications, this filtering has been referred to as CLEAN or a variant thereof, leading to some confusion. This iterative deconvolution process *is* based on the CLEAN algorithm as outlined in Högbom, 1974, but there is no imaging or other steps associated with typical radio astronomy CLEAN algorithms. The PAPER wide-band delay filtering is similar only in the iterative process of deconvolving the maximal point in 1D delay space.

3.4.1 Model Catalog

The standard MWA GLEAM catalog contains approximately 300,000 point sources across 30,000 deg². For this work, I use a modified GLEAM catalog which includes Pictor A⁶ and Fornax A, both of which are treated as point sources given PAPER’s resolution.

It should also be noted that the catalog contains an observing gap from 6h 30m to 7h 30m LST to avoid the Galactic plane, which lies at the tail end of the available PAPER-64 observational data. I choose to focus on the LST range (1-6 h) where known sources were subtracted and thus ignore the observational data where gaps in the GLEAM catalog exist. PAPER also has a non-trivial response to bright sources above the maximum GLEAM declination of +30° (e.g. Cygnus A). These sources have not been modeled or removed from the data.

3.4.2 Model Visibility Parameters

FHD has three primary functions: Simulation, Deconvolution and ‘Firstpass’. The Deconvolution feature allows for full deconvolution of sources using a modified CLEAN algorithm (Högbom, 1974; Sullivan et al., 2012). Using FHD’s full deconvolution is both time consuming and computationally expensive, and thus is not reasonable to fully deconvolve foregrounds from a 128 day dataset. I choose to use the *Firstpass* feature that accepts a source catalog, a primary beam model (Pober et al., 2012), and the antenna layout which is then forward modeled to match a given observation. This reduces the time for modeling (approximately 10,000 point sources) and subtraction of a single 10 minute PAPER-64 observation to ~ 20 minutes using considerable computing resources.⁷ I force all FHD foreground subtractions to model the GLEAM catalog as point sources, as PAPER-64 lacks the capability to properly resolve any of the extended sources in its FoV. An example of a single observation showing poorly resolved extended sources and the successful subtraction

⁶I use a value of Pictor A’s flux density which is 10% higher than that from Jacobs et al., 2013. This factor is an approximate correction for the flux density scaling issues between the VLSS survey (Cohen et al., 2007) and the VLSSR survey (Lane et al., 2014) used as part of the MCMC pipeline of Jacobs et al., 2013.

⁷For a single observation modeled with a uv -plane resolution of 0.35λ I used 3 cores from an Intel Xeon E5-2650 v4 with 80 GB of RAM with a runtime of ~ 20 minutes. Comparatively for a resolution of 0.1λ using the same 3 cores and 100 GB results in a runtime of ~ 45 minutes

of Fornax A is shown in Figure 3.3, where Fornax A (3^h22^m , $-37^\circ12'$) is near the center of the FoV.

FHD’s Firstpass mode also includes a calibration of the data using the model visibilities and the method described in Salvini and Wijnholds, 2014. However, the PAPER data analyzed here is already calibrated using the methods described in Ali et al., 2015. Hence, I allow FHD to fit a 3rd order polynomial to the time averaged per-frequency solutions for both phase and amplitude to correct for any gain scale or phase center discrepancies between the PAPER calibration and the GLEAM catalog. I calibrate and model 10,000 point sources of fixed flux density (> 1 Jy) and position from the GLEAM catalog. Point sources are modeled in FHD as having a flat spectrum.

3.4.3 Excess High Delay Power in FHD Models

When modeling sources in FHD, each GLEAM point source is modeled as flat across frequency. In the absence of numerical errors, I would therefore expect that the only power that is pushed to higher delay modes in the delay transform is due to spectral leakage from a finite band and from the primary beam’s spectral structure. In principle, FHD can simulate precise and realistic spectral responses when computational resources are abundant, however in this analysis approximations must be made. Numerical artifacts from these approximations in the visibility simulation process can introduce spurious spectral structure in the model visibilities. One particular source of artificial spectral structure in the algorithm used by FHD comes from its discrete sampling of the uv plane at a finite resolution. In this work, to balance the needs of precision and computational efficiency with a 128 day data-set, I model the uv plane with a pixel size of 0.35λ on a side. This does not completely alleviate the numerical artifacts in the source spectrum; instead, in delay space, I find an aliasing floor beyond the horizon that is ~ 3 order of magnitude smaller than the foreground amplitude, but still several orders of magnitude greater than a 21cm EoR signal. When these model visibilities are subtracted from the PAPER visibilities, this high delay power is transferred into the residual visibilities. With the sensitivity of this dataset, this spurious

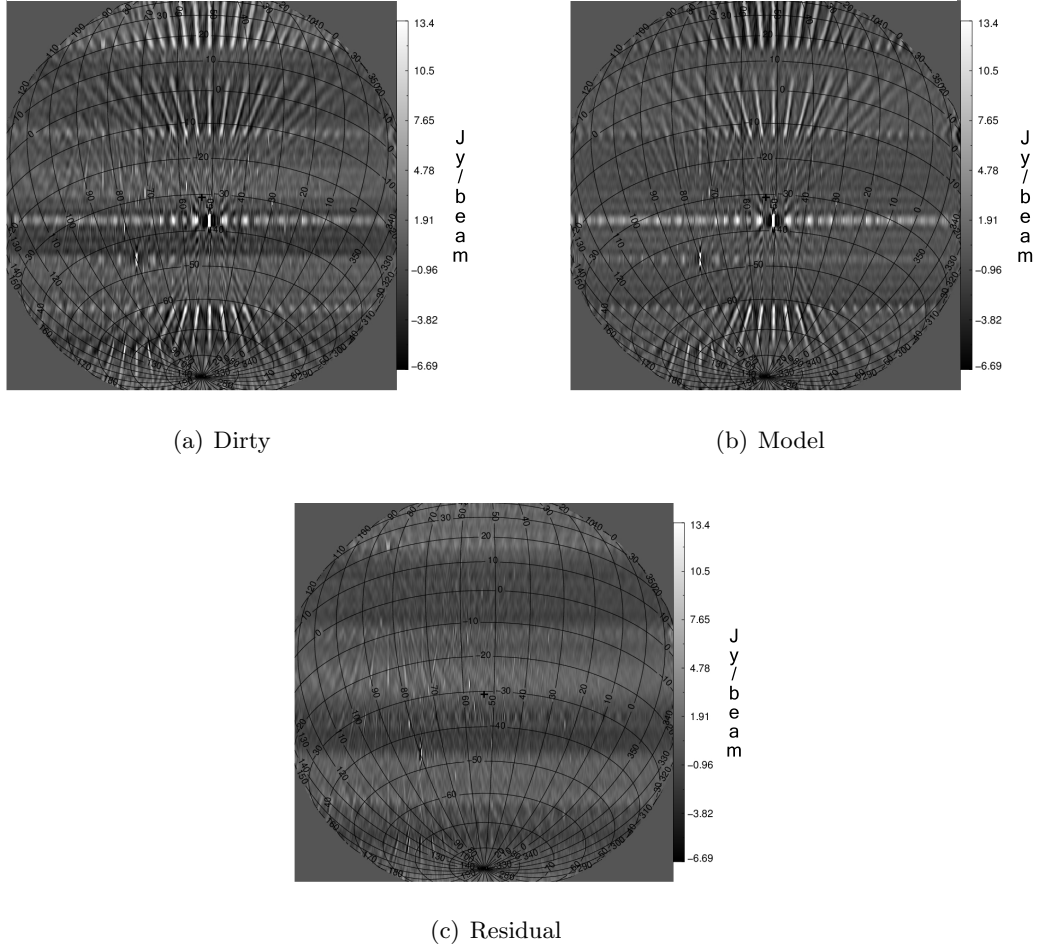


FIGURE 3.3: FHD generated Dirty (a), Model (b), and Residual (c) sky images of a single 10 minute PAPER-64 observation. The two brightest sources in the FoV are Fornax A (3^h22^m , $-37^\circ12'$) and Pictor A (5^h19^m , $-45^\circ46'$). In the Residual sky image, it can be seen that Fornax A is removed almost entirely, while Pictor A is subtracted but to a lesser degree. The partial subtraction of Pictor A is most likely due to a poor calibration and flux model in FHD, where the brightest sources (Fornax A) dominate the calibration process. Cygnus A can be seen towards the upper-left of the Dirty and Residual sky images, but is not present in the model because it is outside of the MWA GLEAM survey field.

spectral structure in the foreground model becomes the limiting factor in the analysis and sets a floor to any potential 21cm signal limit.

To make progress, I low pass filter the model visibilities using the iterative deconvolution filter with limits of $|\tau| \geq 90$ ns (for 30 m baselines). This removes a minimal amount of the model within the horizon limit while giving the dynamic range to ensure that I am not adding power that would contaminate the EoR signal at high k-modes and near the horizon. I present a more in-depth analysis into this source of spurious spectral structure and the filtering I perform on model visibilities in Appendix B.

This filter indiscriminately removes high delay power from the model visibilities, including potential effects from the intrinsic spectral structure in the foregrounds or the instrument response. In this work, the foregrounds are modeled as flat spectrum, and so possess no intrinsic high delay power; however, I do use the full frequency-dependent PAPER beam model in creating the model visibilities. Ultimately, I would like to include the full frequency-dependent model of the instrument and a frequency-dependent sky model in this kind of analysis. At present, however, the spurious structure limits the analysis and filtering is a necessary step to reach sensitivities relevant to the analysis. I further discuss the implications of this analysis choice in Section 3.8.

3.5 Data and Preprocessing

3.5.1 PAPER

The drift-scan PAPER observations used for this analysis originate from the November 2012 - March 2013 observing season of 128 days using the 64 element dual-polarization PAPER array. The dataset is compressed in both time and frequency by averaging from 10.7 s time integrations to 42.9 s and from frequency channel widths of 97.7 kHz to 492.6 kHz using the delay/delay-rate filtering scheme described in Parsons et al., 2014. I identify and flag RFI at the 6σ level prior to calibration. Observations are then redundantly calibrated, also known as OMNICAL calibrated (Zheng et al., 2014), and absolute calibrated. Additional details of this dataset can be found in Ali et al., 2015. Following these initial pre-processing

steps I begin the foreground filtering and filtering+subtraction pipeline analysis as laid out in Figure 3.4. It should be noted that these visibilities undergo an additional calibration in FHD (step B), but after subtraction I return each visibility to their original OMNICAL calibration. In doing this, I should be mitigating any flux density scaling discrepancies, coordinate, or pointing errors.

3.5.2 MWA

The MWA observations used are 2.6 hours from EoR0 ((RA,Dec: $0^h, -27^\circ$) which is a region of the sky with minimal Galactic emissions (Beardsley et al., 2016)) where each visibility has a time integration of 2 s as opposed to the 42.9 s time averaged integrations of PAPER. The MWA Phase I EoR observations (Jacobs et al., 2016; Trott et al., 2016; Pober et al., 2016) have an instantaneous bandwidth of 167-197 MHz across 384 channels. The narrow-band EoR redshift observations are taken over 95 channels (7.68 MHz) which correspond to a $\Delta z \approx 0.3$ for most measurements. The pipeline for comparing filtering vs. filtering+subtraction for the MWA data is only similar to PAPER-64 up to step D and I omit the RFI flagging after the wide-band delay filter (step C).

3.6 Implementation on PAPER Data

The analysis focus for this work begins after the calibration steps described in Section 3.5. I emphasize only the critical steps required for the application of the combined foreground subtraction and filtering approach; a full description of the PAPER-64 analysis pipeline can be found in Ali et al., 2015. An abridged outline of the power spectrum pipeline can be seen in Figure 3.4, which shows the placement of the foreground subtraction step I add to the analysis. I choose to forgo the fringe rate filtering and the inverse covariance weighting at the end of the traditional PAPER-64 power spectrum pipeline. While these steps can improve the overall limit one can place on the 21cm signal, they obscure many of the main effects seen in the analysis and complicate interpretation of the results. However it is possible to include inverse covariance weighting, fringe rate filtering, and signal loss in the filtering+subtraction power spectrum pipeline. An increased scrutiny must be applied to

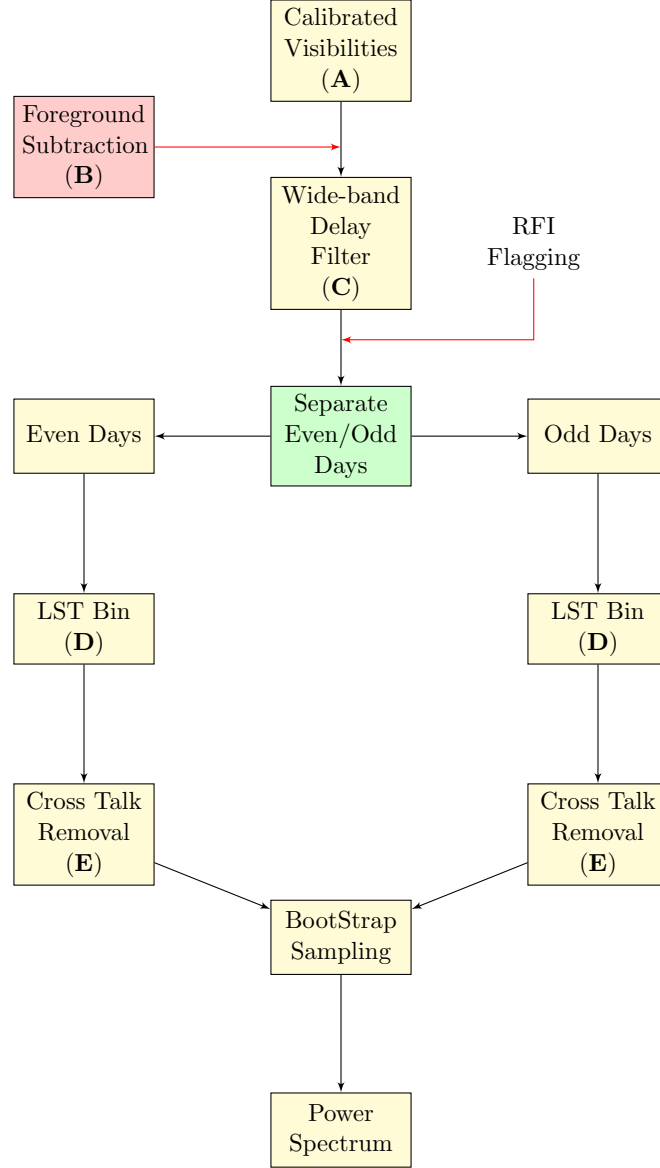


FIGURE 3.4: Abridged PAPER power spectrum pipeline used in processing visibilities for EoR measurements. After calibration (step A; described in Section 3.5), there is the option to subtract model foregrounds (step B). After the wide-band filtering (step C; described in Section 3.3), additional weak RFI is removed by flagging data with $> 3\sigma$ deviation. Visibilities are split into even/odd days and averaged in local sidereal time (LST; step D). I cross multiply the even/odd LST binned visibilities to obtain an unbiased power spectrum estimator. Cross-talk removal (step E) is performed by using the full 8.5 hrs of observations to ensure at least one full fringe period has elapsed. The PAPER power spectrum analysis from Ali et al., 2015 additionally uses a signal loss correction loop and inverse covariance weighting; these steps are not included here as they do not contribute to the understanding of filtering+subtraction. The letters labeling each step correspond to the steps in Figure 3.5.

any resulting power spectra from these additional analysis techniques such as was covered in Cheng et al., 2018; where mode coupling during inverse covariance weighting caused additional signal loss and signal loss was more than was estimated.

To determine the effect of foreground subtraction on the analysis, I split the pipeline into the traditional PAPER-64 power spectrum pipeline, which I refer to as **filtering**, and one with FHD foreground subtraction included, referred to as **filtering+subtraction**. Figure 3.5 shows the effect each of these foreground removal steps has on the visibility amplitudes of 8.5 hrs of data from a 30 m baseline. The first three steps (A,B,C) involve no averaging over days of observation; however, in panel D LST binning is used to average over 64 days of observation (while the total data set is ~ 128 days, even and odd days are averaged separately). Note that omitting foreground subtraction (step B) results in plots C through E which are visually indistinguishable from the version shown. The white regions in Figure 3.5 are missing data due to flagging intermittent RFI which are recovered in step D due to LST binning. The RFI flagging that is consistent among all steps centered at 137 MHz with no recovery from LST binning is the ORBCOMM satellite system which is persistent across all observations and confined to ~ 4 frequency channels.

Depending upon the RFI strategy employed, frequency/time samples can be falsely identified as containing RFI, which is the case when flagging between steps C to D. The wide-band delay filtering decreases the narrow-band variance in the band center due to the Blackman-Harris window function; however, the wide-band variance remains comparatively large. This means that a reduction in the false-positive RFI identification rate can be achieved from either reducing the wide-band power (foreground subtraction) or using a different window function in the filtering step prior to RFI flagging. I address additional benefits of using a different window function in the filtering step and how it affects the power spectrum in Section 3.6.2.

In the final pre-processing step prior to the power spectrum estimate I apply cross-talk removal to all visibilities. I average per frequency over the entire 8.5 hrs of visibilities which is then subtracted from each observation. By doing this I am able to account for the offset in the visibility complex plane due to cross-talk that becomes apparent after one full fringe

period in the 30 m baselines.

For determining the statistical significance of the filtering+subtraction technique I empirically estimate errors on the PAPER-64 power spectra in Section 3.6.1 by bootstrapping. This bootstrapping is done in accordance with modifications suggested in Cheng et al. (2018). I use 400 bootstrap samples taken from three redundant 30 m baseline types and over LST. To properly account for the error of differenced power spectra in Sections 3.6.1, 3.6.2, and 3.6.3, I bootstrap over the distribution of differenced samples rather than use the combined error from each independent power spectrum. The differenced power spectra are created in parallel to make certain that the same random sample from LST and baseline are chosen.

3.6.1 Effect of Foreground Subtraction

I use the PAPER-64 power spectrum pipeline from Ali et al. (2015) as the point of comparison for determining how applying foreground subtraction prior to filtering can increase the sensitivity. The significance of the improved sensitivities are estimated by bootstrap sampling over the distribution of power spectra and differenced power spectra. While the foreground filter is performed across the entire 100 MHz PAPER band (hence the name “wide-band filter”), cosmological power spectra are restricted to a narrower range of frequencies to limit redshift evolution. In this analysis, I use 20 MHz of data for each cosmological power spectrum, but use a Blackman-Harris window function in this “narrow band” delay transform, giving an effective bandwidth of 10 MHz.

Power spectra from the band center at 150 MHz ($z = 8.4$) can be seen in Figure 3.6(a), where I compare each successive technique for foreground removal: no foreground mitigation (green), foreground subtraction (red), foreground filtering (blue), and filtering+subtraction (purple). The shaded gray region corresponds to modes inside the horizon, and the dashed black line corresponds to a delay of 15 ns beyond the horizon; the wide-band deconvolution is only allowed to fit modes with delays less than this value. I see that subtraction drops the foreground power by an order of magnitude within the horizon limit, with marginal improvements at higher delays. (Because of the Blackman-Harris window function in the

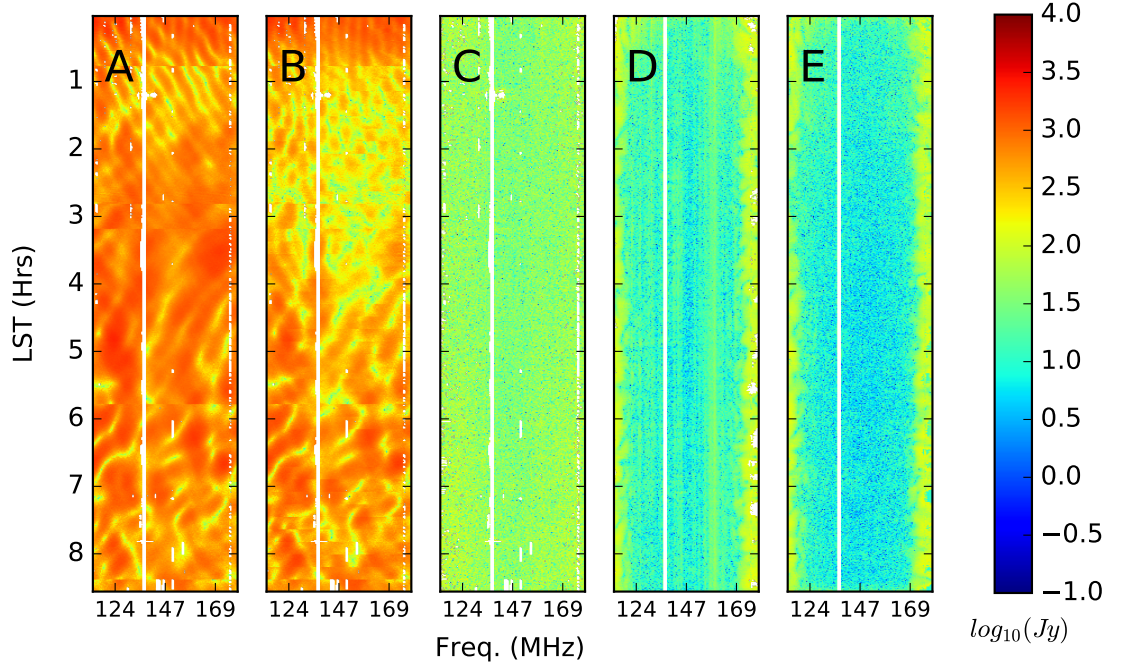


FIGURE 3.5: Sequential foreground removal steps of a fiducial 30 m PAPER-64 baseline. The steps are **(A)** OMNICAL calibrated baseline visibility, **(B)** foreground subtraction, **(C)** foreground filtering, **(D)** LST binning, and **(E)** cross-talk removal. Foreground subtraction can be seen to reduce power primarily due to Fornax A around 1-4 hrs LST in **(B)**, and additional foreground power is removed from smaller transiting sources up until the GLEAM catalog cutoff at 6h 30m LST. The end result of applying the PAPER-64 power spectrum pipeline techniques results in ~ 4 orders of magnitude in amplitude of foreground suppression. A diagrammatic overview of how the pipeline progresses by step is shown in Figure 3.4. The discontinuities seen in LST in both A & B steps originate from gain variations due to applying the OMNICAL calibration on single 10 minute observations.

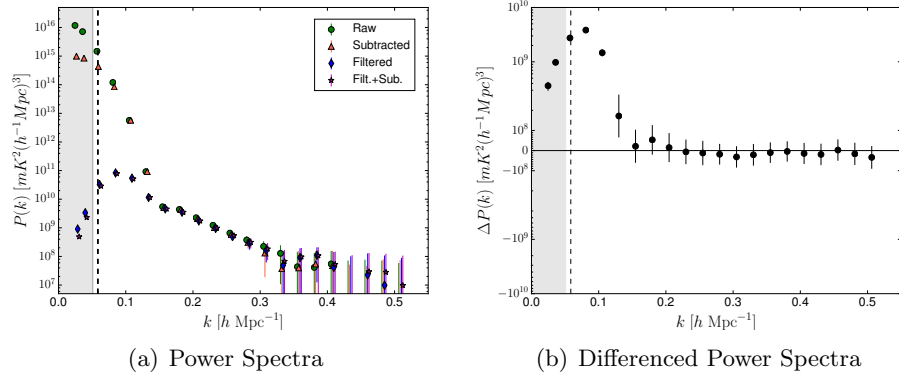


FIGURE 3.6: (a). Resulting power spectra from the 128 day LST binned PAPER-64 dataset at band center ($z=8.4$) focused on 1-6 hrs LST, where the shaded region represents the horizon limit for a 30 m baseline and the broken vertical line is a 15 ns filtering extension. Power spectra correspond to each step of foreground removal: the raw PAPER-64 power spectrum (green), foreground subtracted (orange), foreground filtered (blue), and the filtered+subtracted technique (purple). (b). The differenced filtered and filtered+subtracted power spectra showing where I see improvements in sensitivity; more positive means greater sensitivity improvement due to reduction in spectral leakage. The first few modes nearest the horizon ($k \approx 0.06$) show a statistically significant (2σ) improvement through the implementation of filtering and subtraction.

narrow band delay transform, adjacent k modes in the power spectrum are highly correlated, and so the improvements in the first mode beyond the horizon should be interpreted with caution.) The foreground filtering removes a very large fraction of power within the horizon limit, but also leads to significant improvements in the first several modes beyond the horizon. This behavior is indicative of reduced spectral leakage from modes within the horizon.

The power spectrum difference between filtering and filtering+subtraction can be seen in Figure 3.6(b), which shows a statistically significant (2σ) improvement of modes within and near the horizon. At the most basic level, this result is expected. The model-based foreground subtraction has made the job of the wide-band delay filter easier: there is less power it needs to remove. In particular, when foregrounds are subtracted from the raw visibilities, I reduce the amount of power that rings off of sharp discontinuities in the frequency sampling function when taking the delay transform. In the PAPER data set, the ORBCOMM satellite band centered at 137.5 MHz spans several channels and is always

flagged. This creates a very sharp discontinuity that leads to strong ringing of foreground power throughout visibility space. Because the frequency sampling function is known, the iterative wide-band filter can, in principle, remove this scattered power itself. But, the deconvolution algorithm is not perfect, and it appears that reducing the total amount of foreground power available to scatter in a delay transform leads to better performance of the wide-band deconvolution filter. Put another way, the wide-band foreground filter has an effectively fixed dynamic range that it can achieve between the brightest foreground contaminated modes and the limits in the EoR window. Even though the model visibilities contained no intrinsic power in the EoR window, reducing the power in the brightest foreground contaminated modes improves the end result because I am limited by the dynamic range of the wide-band filter.

3.6.2 Performance Near Band Edges

One of the strengths of the PAPER experiment is its wide instantaneous bandwidth which covers from 100 to 200 MHz ($z \sim 6 - 13$). The data analyzed in Section 3.6.1 were centered around 150 MHz, where past analyses have shown the wide-band foreground filter (using a Blackman-Harris window function) to be most effective (Jacobs et al., 2015). In this section, I demonstrate the effectiveness of filtering+subtraction nearer to the edges of PAPER’s frequency band with power spectra at 124 and 169 MHz (corresponding to $z = 10.4$ and 7.4, respectively).

The green points in Figure 3.7 show the difference in power spectra made using filtering and filtering+subtraction at all three redshifts: $z = 10.4$ (Figure 3.7(a)), 8.4 (Figure 3.7(b)), and 7.4 (Figure 3.7(c)). (Note that the green points in Figure 3.7(b) are the same data shown in Figure 3.6(b).) These data show that filtering+subtraction removes significantly more power from the first few modes outside the horizon than filtering alone. Compared with the $z = 8.4$ results in Figure 3.6(a), filtering+subtraction removes one to two more orders of magnitude of foreground power than it does at the band center. This increased effectiveness of the technique at the band edges comes more from foreground subtraction itself than any interplay between the subtraction and the filtering, which drove the improvements at band

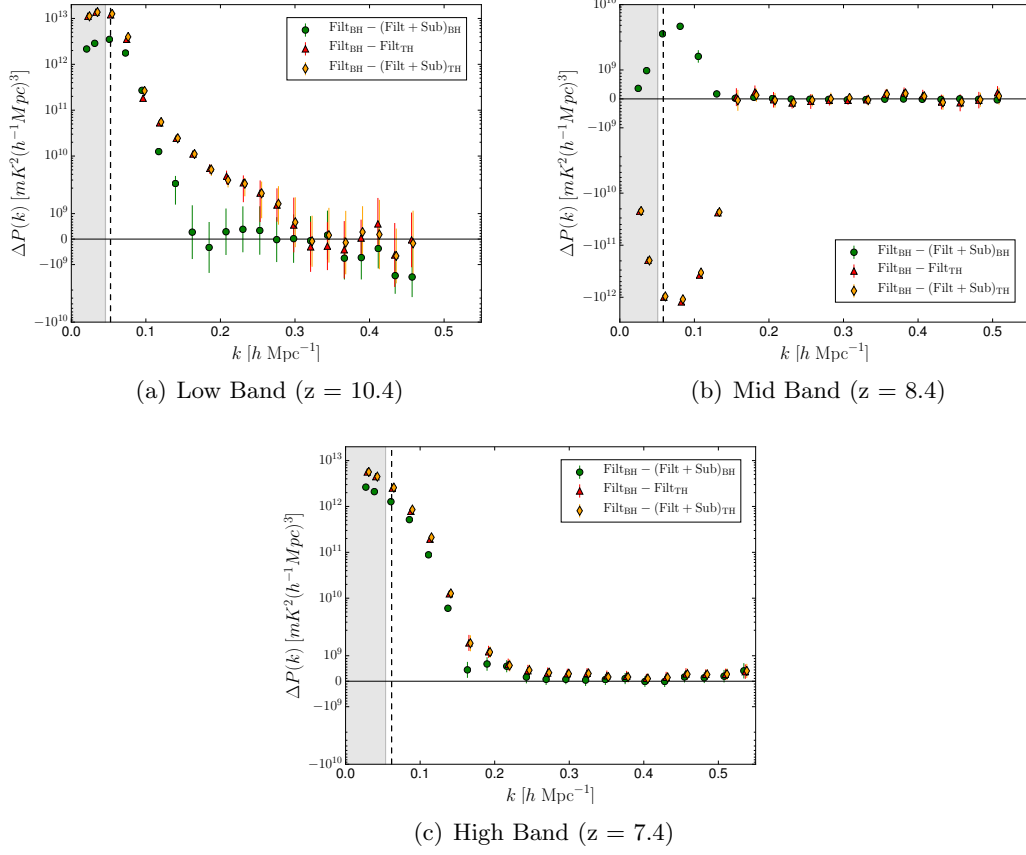


FIGURE 3.7: The power spectrum difference between filtering and filtering+subtraction at different band locations comparing the effectiveness of window function choice on power removed during the wide-band filtering step, where again more positive shows an improvement by both filtering and subtracting foregrounds. I compare the standard Blackman-Harris (BH) to the Top-hat (TH) window function. I should expect that the Top-hat window function gives the maximum amount of foreground removal at band edges because of the equal weighting across the entire band. A symmetric improvement about band center is shown in (a) & (c) where the switch from BH to TH window function increases the reduction in spectral leakage in the low (high) band by $\sim 145\%$ (113%) and with subtraction $\sim 169\%$ (125%) at $k \approx 0.11 \text{ hMpc}^{-1}$. The band center plot (b) demonstrates that the BH window function is the ideal choice regardless of filtering or filtering+subtraction when compared to the TH window.

center. This effect is primarily because the Blackman-Harris window in the wide-band filtering step specifically down-weights the band edges. Therefore, I expect the wide-band filter to perform comparatively poorly at the band edges, as was seen in Jacobs et al., 2015. Subtraction should, however, significantly help at the band edges since it is able to remove power here that the wide-band filter cannot. Alternatively, I might expect that a different choice of window function in the filtering step should give dramatic improvements at high/low redshifts when compared to the standard pipeline.⁸

The PAPER power spectrum pipeline uses the Blackman-Harris window function when applying the iterative deconvolution filter, which gives the result of maximum foreground avoidance at band center. This choice of window function, however, produces limited improvements at the band edges due to down-weighting prior to delay transforming. Typically, if one were to forgo the use of a window function (i.e. using a Top-hat window) in the filtering of data with discontinuities this would result in significant spectral leakage, but this is not the case here. I can overcome spectral leakage in this instance because I am deconvolving both the window and sampling function from the filtered visibility. There are however limitations in doing this because deconvolution is a computationally demanding task. Furthermore, the product of the window and sampling functions can be complicated, and varies from visibility to visibility, so in some cases the iterative nature of this filtering may fail to converge. I therefore might expect the Top-hat window function to be an improvement over the Blackman-Harris near the band edges because it equally weights every frequency channel (as long as convergence in the iteration is reached). The points in Figure 3.7 are a difference of power spectra demonstrating the effect of window function choice in the wide-band filter (more positive is better). For these points, the foreground filtering+subtraction analysis was applied in two cases (green; orange) to show the effect without a Top-hat window function used in the wide-band transform and after using the Top-hat window. I also show the change from Blackman-Harris to Top-hat (red) in a filtering only case to provide a point of reference for the inclusion of the foreground subtraction

⁸Alternatively, Liu, Parsons, and Trott (2014b) describe a method for quantifying the leakage between k-bins from choice of window function therefore reducing emphasis on which window function is used; I do not explore this effect here.

in the filtering+subtraction. I see that the Top-hat filter leads to significant improvements at the band edges — almost an order of magnitude larger near the horizon than the improvements I see from filtering+subtraction with a Blackman-Harris window. At the band center ($z = 8.4$), however, the Top-hat filter results in a significantly worse result near the horizon: below k of $\sim 1.5 \, h\text{Mpc}^{-1}$, the Top-hat filter left significantly more residual foreground emission, making the difference negative. I expect that the Blackman-Harris should outperform the Top-hat window function in this case, since the Top-hat has more difficult sidelobes to deconvolve.

Lastly, the orange points in Figure 3.7 show the difference between filtering with a Blackman-Harris window and filtering+subtraction with a Top-hat window. In all three bands, filtering+subtraction with a Top-hat leads to even more of an improvement than filtering with a Top-hat alone (red). At the band edges, the majority of the improvement over filtering with a Blackman-Harris appears to come from the switch to the Top-hat window, but the best results come when first subtracting a foreground model and then applying the wide-band deconvolution filter using a Top-hat window. At the band center, the Blackman-Harris proves the best choice of window function in the wide-band filtering step.

3.6.3 Improved measurements within the horizon

Delay filtering of foregrounds has been shown to be a powerful technique for 21cm data analysis, but it does have significant limitations. Modes within and directly outside the horizon (due to 15 ns extension see §3.3) have effectively 100% EoR signal loss when filtering and therefore become useless for cosmological measurements (see Appendix A). This is of course unfortunate, as the EoR signal is brightest at these large scales. To begin to recover some of these modes, I can potentially use foreground subtraction prior to filtering and then reduce the filter’s extent in k back towards or even inside of the horizon. This means that I can maximize the effect of subtraction mitigating the foreground contamination by purposefully reducing the performance of filtering at the cost of sensitivity at

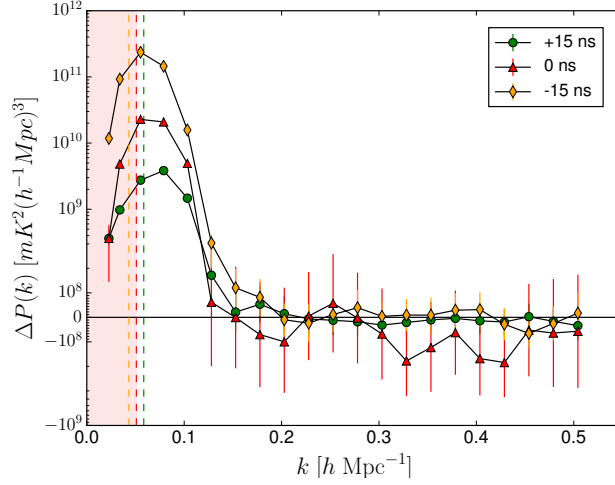


FIGURE 3.8: Power spectra differences (filtering minus filtering+subtraction) for different sizes of filter: 15 ns beyond the horizon limit (green), at the horizon (red), and 15 ns inside the horizon (orange). Smaller filters leave more of k space available for EoR measurements (mode within the filter have effectively 100% signal loss), and subtraction is seen to result in bigger improvements when smaller filter sizes are used. With improved sky models, subtraction can therefore play an important role in maximizing the recovery of the 21cm signal.

higher k -modes. Figure 3.8 shows power spectrum differences between filtering and filtering+subtraction for three choices of filter width: at the horizon (0 ns, red), 15 ns beyond the horizon, as used in the analysis above (green), and 15 ns inside of the horizon (orange). I see that the improvements coming from the subtraction step grow larger the less aggressive the filter, i.e., filtering+subtraction results in the biggest improvement when the filter is set to 15 ns inside the horizon (-15 ns). Overall, the total residual foreground power in k modes near the horizon is minimized with the most aggressive filter (+15 ns), but as I continue to improve foreground models and the efficacy of subtraction, it will be worthwhile to revisit this analysis and see whether k modes at or near the horizon can be turned into useful limits even in a foreground filtering approach.

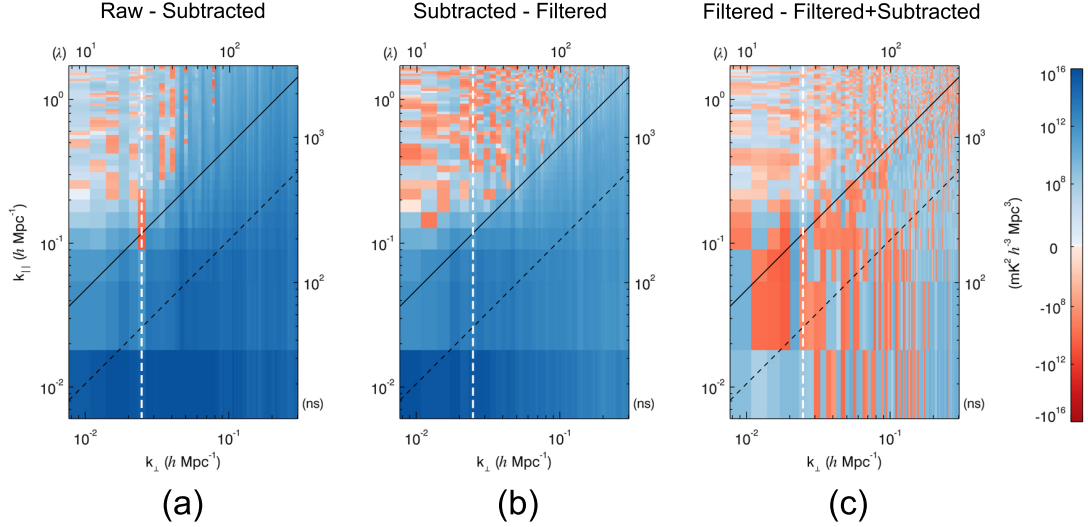


FIGURE 3.9: MWA 2D wedge power spectra difference plots showing the comparison between each foreground removal step where blue (red) indicates an improvement (worsening) between steps. The solid diagonal line represents the horizon limit while the broken line represents the Full Width Half Max of the MWA primary beam (Sutinjo et al., 2015). The vertical white broken line marks a 30 m baseline, equivalent in length to that used in the PAPER-64 analysis. The difference between the raw power spectrum with no foreground removal and one with FHD foreground subtraction is shown in (a) which gives an expected improvement inside of and beyond the horizon limit. The comparison between FHD foreground subtraction and wide-band filtering (b), shows an additional clear improvement over the standard MWA subtraction pipeline at the majority of $k_{\perp}$ modes nearest the horizon. Finally, comparing the use of filtering with both filtering+subtraction in panel (c), I do not see any clear improvement, but comparison with the PAPER-64 results suggest that I have reached a noise limit where any improvements are well below the sensitivity available with only three hours of data.

3.7 Hybrid Foreground Removal Technique

Applied to MWA Phase I

To demonstrate the robustness and general applicability of the hybrid foreground removal technique across multiple instruments, I apply it to the three hours of MWA Phase I observations as described in Section 3.5.2. In many ways, these observations represent the extreme opposite of PAPER-64. The MWA Phase I has 128 tiles⁹ laid out in a minimum redundancy configuration, giving it a much fuller (u, v) coverage than PAPER-64. It also has baselines up to nearly 3 km in length, giving it a resolution of order a few arcminutes, a factor of ~ 10 improvement over PAPER-64. FHD was also designed specifically for use with MWA interferometry data and the GLEAM catalog used for subtraction is a direct MWA product. I should thus expect better relative foreground subtraction when compared to PAPER-64.

The results of applying the hybrid technique on approximately 3 hours of MWA Phase I observations at band center (182.415 MHz) are shown in Figure 3.9 in ‘xx’ instrument polarization. For the most accessible comparison to PAPER-64 hybrid foreground removal I continue to use the difference between power spectra but in the form of the cylindrically averaged 2D differenced power spectra. This provides the ability to approximately compare similar MWA and PAPER 30 m baseline difference power spectra and to view the effect of foreground filtering+subtraction across multiple MWA baseline types. In these difference plots, blue indicates a region where the first step of the analysis has more power, and red indicates a region where the second step has more power. In Figure 3.9(a) (left-hand panel) I show the difference between raw data (calibrated, but no foreground subtraction) and the standard foreground subtraction used in the MWA power spectrum pipeline (as performed in e.g. Pober et al. 2016, Jacobs et al. 2016, and Barry et al. 2016). I see, unsurprisingly, that foreground subtraction removes a significant amount of power inside the wedge, but there are also improvements well beyond the horizon limit (shown as a solid diagonal line). These improvements are on the same order of magnitude ($10^{16} \text{ mK}^2(\text{h}^{-1}\text{Mpc})^3$ within the

⁹MWA tiles consist of 16 individual dipole antenna elements that are steered with an analog beamformer, as opposed to the single dipole element used for drift scanning in PAPER-64.

horizon) as seen in Figure 3.6(a) between the raw (green) and subtracted (red). Figure 3.9(b) compares foreground subtraction with the wide-band deconvolution foreground filtering. Filtering removes more power within the horizon than subtraction, but I also see an improvement immediately outside of the horizon limit across all $k_{\perp}$. Note that no “buffer” was used outside the horizon in the foreground filtering, so any power removed outside the horizon is almost certainly due to reduced spectral leakage. At low $k_{\perp}$ and high $k_{\parallel}$ I approach a noise floor; since only three hours of data were used, it is considerably higher than the noise floor in the PAPER-64 analysis. In Figure 3.9(c), comparing the filtered and filtered+subtracted power spectra, I see a similar noise level, but throughout all of k space. If I assume the results for the MWA should closely mirror those from PAPER-64, then improvements at band center due to filtering+subtraction should be on the order of $\sim 10^9 \text{ mK}^2(\text{h}^{-1}\text{Mpc})^3$ which is below the noise level I see here.

I also see similar results to the PAPER-64 analysis when looking at the MWA band edges (high/low redshift). Based on the PAPER-64 analysis, I would expect improvements of $\sim 10^{12} - 10^{13} \text{ mK}^2(\text{h}^{-1}\text{Mpc})^3$ from using filtering+subtraction over just filtering. This is consistent with the results shown in Figure 3.10, which has a uniform decrease in power within the horizon and a slight, though noisy, improvement in the window. While I have shown a Top-hat to be a much improved wide-band window for analysis near the band edges, MWA visibilities have a 2 s time resolution, compared to 10 s with PAPER. This lower SNR coupled with a Top-hat window can make it more difficult for the iterative deconvolution filter to converge. Using a wide-band Blackman-Harris window function, even when analyzing data near the band edges, is therefore important to help convergence in this instance.

3.8 Conclusions

The next generation of radio interferometers attempting to make a statistical detection of the 21cm EoR power spectrum will require the removal of as much foreground power as possible. To date, even the most promising foreground removal techniques leave behind residual contamination which is well above the most optimistically bright EoR signals. In

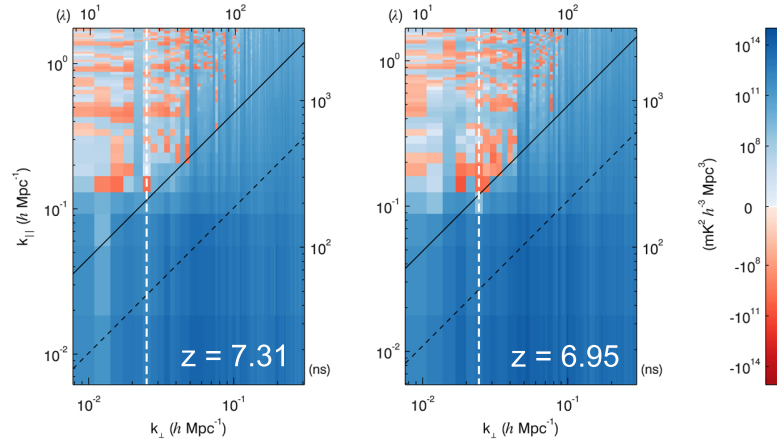


FIGURE 3.10: MWA 2D power spectra difference plots between filtering and filtering+subtraction at low band (7.31) and high band (6.95). The results are generally similar to the equivalent PAPER-64 difference power spectra at the band edges. Improvements near the horizon are due to subtraction compensating for the down-weighting of band edges when windowing.

this work, I have presented a hybrid technique which combines both model based foreground subtraction and delay based foreground filtering. In doing so, I have demonstrated the following:

- Even in parts of the EoR window, residuals from foreground filtering exist which are brighter than an EoR signal.
- Foreground subtraction can be successfully applied to arrays with sparse (u, v) coverage (e.g. PAPER-64) using source catalogs derived from other arrays with high (u, v) coverage (e.g. the MWA). The limiting factor in the ability to perform this subtraction is the numerical fidelity with which I can forward model visibilities for the large (1088 hour) PAPER data set.
- Combining filtering and subtraction results in statistically significant improvements across the entire band, with the greatest improvements at band edges.
- The choice of window function is an important ingredient in wide-band iterative deconvolution foreground filtering, especially when analyzing data at frequencies away from the band center.

- Foreground subtraction can potentially enable less aggressive foreground filtering (i.e. removal of fewer modes near/inside the horizon), allowing for EoR measurements within the horizon.
- Despite a more complicated frequency response (including the coarse band harmonics) and beam shape, foreground filtering can be successfully applied to MWA visibilities and results in a significant improvement over subtraction alone.
- Improvements with filtering+subtraction are more readily seen as SNR improves, therefore I should expect for any improvements to sensitivity to become more visible with longer integration times.

The EoR signal is brightest at low k -modes, so foreground removal nearest (or inside) the horizon limit can have the greatest impact on overall sensitivity (Pober et al., 2016). I have seen that subtraction with a reduction in the net power prior to filtering can reduce leakage into the modes right outside the horizon. It would therefore be reasonable to subtract as much foreground power — both diffuse and point sources — prior to filtering for the most minimally contaminated EoR power spectrum estimates across all redshifts. I demonstrate this hybrid foreground subtraction and filtering technique as a proof of concept, and with additional computational resources to increase the uv -plane resolution to below the 0.35λ used here, I expect that this technique could be drastically improved upon. This would alleviate the need to low-pass filter the model visibilities and should result in near-horizon improvements by both modeling the spectral response of the instrument and reducing the overall dynamic range that the iterative deconvolution filter needs to overcome.

Furthermore, with an improved sky and instrument model I should expect subtraction to provide a more consistent improvement across all visibilities and not just ones containing the brightest foreground sources (Fornax A and Pictor A). More specifically, as weak and smaller sources are included in the analysis I approach a ‘subtraction floor’ due to position errors between PAPER observations and the GLEAM catalog. This leads to improper or partial subtractions that result in no net removal of foreground power. Therefore improvements resulting from these changes in the case of PAPER-64, would stem directly from improved

catalog positions and flux measurements, and the introduction of a diffuse model; but not necessarily from the additional modeling and subtraction of more weaker (< 1 Jy) point sources.

Implementing the iterative deconvolution wide-band filter can be a computationally expensive process, especially when compounded with foreground subtraction which itself is laborious. The filtering step can also be accomplished without the use of the iterative deconvolution filter by doing a more traditional filtering in the delay domain. A sparse sampling function with missing data will however give significant ringing in the visibility. Ringing occurs when a signal is band-limited, which is approximately the case for a raw visibility where the foregrounds are $> 10^5$ when compared to anything beyond the horizon. This ringing is substantially reduced when subtraction is applied prior, potentially bringing the residual foregrounds to a point where they are no longer within the approximate band-limited regime. If subtraction can remove enough power to enable the use of an FIR foreground filter, the computational load and pipeline processing time can be significantly reduced.

Future work will rely on applying this hybrid foreground removal approach to the newest season of observations from the Hydrogen Epoch of Reionization Array (HERA). In this thesis FHD is applied to HERA data in Chapter 5 but not for the purpose of foreground subtraction in a power spectrum. In terms of hardware improvements, HERA-37 alone should increase the sensitivity by a factor of ~ 5 over PAPER-128 (DeBoer et al., 2017). HERA’s beam in terms of complexity positions it between PAPER and MWA, while having better horizon suppression than either of them (Thyagarajan et al., 2015). This places additional importance on mitigating residual foreground contamination and puts it in a new regime for testing the hybrid foreground removal limitations as sensitivities increase. The inclusion of this hybrid approach to the most sensitive 21cm EoR interferometer should lead to the tightest, least contaminated power spectrum constraints over the redshift range of $7 < z < 11$.

Acknowledgments

A version of this chapter was previously published as Kerrigan et al. 2018 ApJ, and is reproduced with the permission of all coauthors and the copyright holder. Copyright 2018 American Astronomical Society.

Chapter 4

Machine Learning for Improving Sensitivity through RFI Mitigation

4.1 Interference in Radio Observations

The next generation of low-frequency radio interferometers are now beginning to become operational. These arrays are looking to detect and measure one of the weakest cosmological signals the Universe has to offer, the brightness-temperature contrast of the 21cm signal during the Epoch of Reionization (EoR). By measuring this highly redshifted signal the progression of the EoR can be characterized. The understanding gained from this characterization has the potential to help unravel how the first stars and galaxies formed and reionized their surrounding neutral hydrogen. While instruments like the Hydrogen Epoch of Reionization Array (HERA) (DeBoer et al., 2017) have the intrinsic sensitivity required to detect the EoR signal through a power spectrum, they are afflicted with anthropogenic noise which is referred to as Radio Frequency Interference (RFI). Interference from RFI in 21cm EoR observations is an especially significant obstacle because it can have a brightness anywhere from on the order of the EoR signal to orders of magnitude beyond even Galactic and extra-galactic foregrounds. RFI unfortunately introduces a reduction in sensitivity in two separate but distinct ways, one being the direct contamination by having similar spectral characteristics and overpowering of the 21cm signal, and the other being

the introduction of a complex sampling function due to missing data. This produces correlations between modes when computing the Fourier transform along the frequency axis (Offringa, Mertens, and Koopmans, 2019). It is therefore important to strike a balance between identifying RFI while not falsely identifying non-RFI as RFI, which leads to further complicating the sampling function over frequency. Many approaches have recently been developed to identify and extract RFI from radio telescope data. RFI algorithms of particular interest include AOflagger (Offringa, van de Gronde, and Roerdink, 2012), which uses a Scale-invariant Rank operator to identify morphologies that are scale-invariant in time or frequency which is a characteristic of many RFI signals. This RFI detection strategy has been used successfully on instruments such as the MWA (Offringa et al., 2015) and the Low-Frequency Array (LOFAR) (Offringa et al., 2013). Alternative approaches to RFI identification include the application of neural networks. More specifically, a Deep Fully Convolutional Neural Network (DFCN) based on the U-Net architecture (Ronneberger, Fischer, and Brox, 2015) has been used on single dish radio telescope data (Akeret et al., 2017), and a Recurrent Neural Network (RNN) has been applied to signal amplitudes from radio interferometer data (Burd et al., 2018).

In this chapter I expand upon the RFI identification approach using a DFCN developed in `Tensorflow` (Abadi et al., 2016) with the use of both the amplitude and phase information from an interferometric visibility. This technique is prompted by examples such as what is shown in Figure 4.1, which demonstrates how the phase of time-ordered visibilities (waterfall visibilities) can provide supplemental information in identifying RFI beyond that of an amplitude-only approach. Note that in this chapter, all time-ordered visibility plots of real data are in the yellow-purple palette (e.g. Figure 4.1) whereas all simulated data is in the blue-white palette (e.g. Figure 4.7). To understand the improvements afforded by the joint amplitude-phase network I compare it to both an amplitude only network and the Watershed RFI algorithm which is the current RFI-flagging algorithm of choice for the HERA data processing pipeline.

This chapter is outlined as follows. Section 4.2 introduces the architecture of the network, discusses how it compares to previous work, and describes the training dataset. I

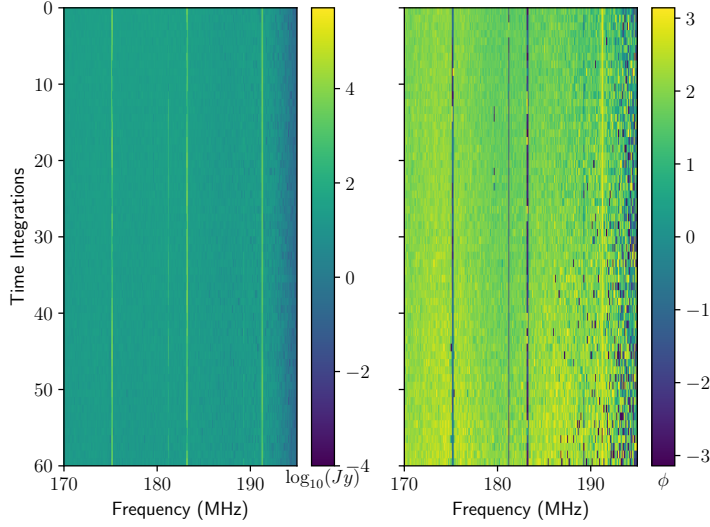


FIGURE 4.1: An example of a HERA 14m baseline waterfall visibility between 170-195 MHz in amplitude (left) and phase (right). The phase waterfall visibility demonstrates how it can provide complementary information about the presence of RFI such as in the 181.3 MHz channel which has constant narrow-band RFI and the more spontaneous ‘blips’ in the 179.5 MHz channel at time integrations of 13, 22, and 23. The significant contrast between the phase of the sky fringe, and how it’s restricted to a narrow-band is an obvious indication of being RFI.

then demonstrate the effectiveness by evaluating both DFCNs on simulated and real HERA observations in Section 4.3. Finally in Section 4.4 I conclude with discussion of further applications and future work.

4.2 Method

4.2.1 DFCN Architecture

The standard 2d convolutional neural network (CNN) (LeCun and Bengio, 1998) is structurally similar to that of a typical Artificial Neural Network (ANN) (Lecun et al., 1998), but it differs to an ANN’s dense layers of ‘neurons’ by its successive convolutions of an input image, which preserves spatial dependence. Each convolutional layer contains a set of learn-able filters which represent a response for particular shapes at different scales (e.g. the edges of an object in an image). The convolved output for every layer is then typically

downsampled using a process known as max pooling that strides a window across the image keeping the highest pixel value within the window. Max pooling provides both a computational improvement due to a decreased image size, and an added level of abstraction relative to the initial image. After the convolution and max pooling layers the image typically is then passed through a non-linear activation function (e.g. sigmoid function) which produces a spatial activation map describing the convolutional layer’s response to every pixel contained within the image. The eventual output of these successive convolution, max pooling, and activation layers is then used to predict (or regress) based on the classification of the image. The error between the predicted class and the true class is then computed through a loss function such as the cross-entropy loss (or mean squared error for regression) and the error is back-propagated through the network updating the learn-able parameters.

The style of network I describe in this chapter deviates from a traditional CNN by existing as a multi-input/output Fully Convolutional Neural Network, where the number of inputs is equal to outputs (i.e. pixel-wise prediction). For a deeper understanding of this kind of network architecture, see Krizhevsky, Sutskever, and Hinton, 2017. I begin with a Deep Fully Convolutional Network architecture similar to the U-Net RFI (Ronneberger, Fischer, and Brox, 2015; Akeret et al., 2017) implementation. However instead of using a uniform number of feature layers per convolutional layer I increase the feature layers as you get deeper into the network. This approach should offer an increase in performance as the feature map for each successive convolution shrinks the layer size decreasing the convolution space. Some caution should be taken here as a reduced number of feature layers for the input convolutional layers could cause prediction performance to suffer as the neural network may be incapable of capturing important fine structure features in your images. This initial layer feature size is another hyperparameter that needs to be tuned where the fewest number of layers required to model the data is optimal. Each stacked layer in the max pooling stages has the dimensions of $(\frac{H}{2^L} \times \frac{W}{2^L} \times 2^L F)$ where F is the number of feature layers, H and W are the layer height and width in pixels, and L is the layer of interest. Every output for each stacked layer contains a 1x1 convolutional layer with a ReLU activation function which is analogous to the ‘Network in Network’ introduced in Lin, Chen,

and Yan, 2013. By introducing 1x1 convolutions with non-linearities to each output I yield a per-pixel discrimination independent of neighboring pixels which increases the network’s ability to model higher-level abstractions. This is important because abstractions of the data are typically invariant to initial changes (Bengio, Courville, and Vincent, 2012) thus having the potential for a more robust predictive power.

To adapt the network to use the visibility phase component, I mirror the amplitude only network as shown in Figure 4.2. I then combine successive amplitude & phase convolution layers at each transpose convolution layer with the technique known as ‘skip connections’ introduced in Long, Shelhamer, and Darrell, 2014 and He et al., 2016. This is implemented by taking the output of a downsampled convolutional layer and concatenating it with an upsampled transpose convolutional layer of equal time, frequency, and feature dimensions. By using these skips in the convolutional pathway, the network is provided with a “template” from which to make small deviations. More simply speaking, the skip connections re-introduce higher resolution information that may have become lost due to the successive convolution and pooling operations that further remove the network’s ability to relate higher order abstraction to the raw input images. This fixes an issue within deep networks where fits to nonlinearities become dominant in a layer, leading to training and overfitting issues. Empirically, I find that using skip connections in conjunction with phase information allows for training a deeper network that converges in fewer iterations than the simple amplitude-only network. This effect is most likely due to the fact that while the amplitude and phase visibilities are broadly dissimilar the one aspect where they can agree is where RFI exists. Layers deeper into the network rely on more abstract features and by stacking these skip connections from amplitude and phase pathways, activations due to RFI present in pixels strengthens inferences at later layers.

For each of the skip layer concatenations between the amplitude and phase pathways, I subtract the mean and normalize over both time and frequency, which assists in standardization as amplitude and phase features can be quite disparate. The amplitude only DFCN I use has $\sim 6 \times 10^5$ trainable parameters, while the addition of the phase downsampling layers for the amplitude & phase DFCN pushes the number of trainable parameters

layer type	kernel size	stride	filters	depth
convolution	3x3	1	16	2
convolution	1x1	1	16	1
maxpool	2x2	2		1
batch norm.				
convolution	3x3	1	32	2
convolution	1x1	1	32	1
maxpool	2x2	2		1
batch norm.				
convolution	3x3	1	64	2
convolution	1x1	1	64	1
maxpool	2x2	2		1
batch norm.				
convolution	3x3	1	128	2
convolution	1x1	1	128	1
maxpool	2x2	2		1
batch norm.				
convolution	3x3	2	256	2
convolution	1x1	1	256	1
maxpool	2x2	2		1
batch norm.				
transpose conv.	3x3	2	128	1
transpose conv.	3x3	2	64	1
transpose conv.	3x3	2	32	1
transpose conv.	5x5	4	2	1

TABLE 4.1: Architecture overview of the DFCNs demonstrated in this analysis. The colored rows correspond to the concatenations on the outputs between those respective layers, where prior to the concatenation each layer undergoes a batch normalization. The depth of a layer here means that there are multiples of the layer stacked all having the same properties. The amplitude-phase DFCN has two input pathways mirrored up until the first transpose convolution layer.

to $\sim 9 \times 10^5$. The specific per layer attributes employed in the networks can be seen in Table 4.1, where it should be noted that per layer dimension sizes are not specified because this style of network is agnostic to the input height and width. Additionally all convolutional layers are zero padded so that input height and width dimensions are equal to the output which helps with reducing the minimum allowable dynamic input size.

To optimize the network hyperparameters, a coarse grid search was performed over activation function, dropout rate, learning rate, and batch size; the optimal results from this search are found in Table 4.2. The depth of the convolutional layers are chosen to maximize learning and minimize prediction times, while trying to retain abstractions of the input visibilities that can properly describe the RFI. These dimensions are thus determined by initially training at an arbitrarily high number of feature layers and scaling back to the minimum number of layers I need to retain for convergence of the training loss.

4.2.2 Data Preparation

The analysis in this chapter is performed entirely on HERA data (both simulated and real) and therefore should be noted that any data preparation techniques outlined here may be unique to HERA. This does not imply that they are unsuitable for other radio interferometers but additional precautions may need to be taken into consideration. To prepare the amplitude-phase input space to be as robust to as many visibility scenarios as possible, I must adopt several standardizations. The amplitude of the visibility can vary drastically by local sidereal time (LST), day, and baseline type while having significant differences in dynamic range. In contrast, the phase of a visibility is intrinsically more standardized: it is constrained between $-\pi \leq \phi \leq \pi$ and should have a mean that is approximately $\mu_\phi = 0$, so I should only expect substantial deviations across baseline type, which are due to changing fringe rates. Therefore to lessen the dynamic range issues in amplitude, I standardize the waterfall visibilities $V(t, \nu)$, according to $\hat{V}(t, \nu) = (\ln|V| - \mu_{\ln|V|}) / \sigma_{\ln|V|}$, by subtracting the mean, $\mu_{\ln|V|}$, and dividing by the standard deviation, $\sigma_{\ln|V|}$, across time and frequency of the logarithmic visibilities.

TABLE 4.2: Parameters and network architecture features that were determined by grid-search cross validation which included the activation function but not the loss function. The dropout rate is uniform across all nodes as highlighted in Figure 4.2.

Parameter	Values
Batch size	256
Optimizer	ADAM ¹
Learning rate	0.003
Activation function	Leaky Rectified Linear Unit ²
Dropout rate	0.7
Loss Function	Cross Entropy

¹Kingma and Ba (2014)

²Maas, Hannun, and Ng (2013)

To further increase the robustness and generalizability of the network for different time and frequency sub-bands, I slice the HERA visibilities into 16 spectral windows of dimensions 64 frequency channels by 60 time integrations ($6.3 \text{ MHz} \times 600 \text{ sec}$). I then pad both time and frequency dimensions by reflecting about the boundaries, extending the dataset in both directions. This allows one to make more reliable predictions for the edge pixels, which otherwise would be ignored due to the convolution layer kernel size of 3×3 ($98.44 \text{ kHz} \times 30 \text{ s}$). Furthermore, I want to use square input channels to maintain a 1:1 aspect of time to frequency pixels. These considerations inform the decision to use a size of 68×68 for the input waterfall visibility.

Combined with the training batch size, $N = 256$, for the amplitude-phase DFCN, the input training space size is $N \times H \times W \times C = (256 \times 68 \times 68 \times 2)$, where C is the number of input channels (e.g. $C_0, C_1 = \hat{V}(t, \nu), \phi(t, \nu)$).

4.2.3 Training Dataset

The training dataset was composed of simulated HERA visibilities using the simulator, `hera_sim`.¹ This simulator creates visibilities according to a ‘pseudo-sky’, which means that modeled point sources have no relationship, in either time or frequency, to any real

¹https://github.com/HERA-Team/hera_sim

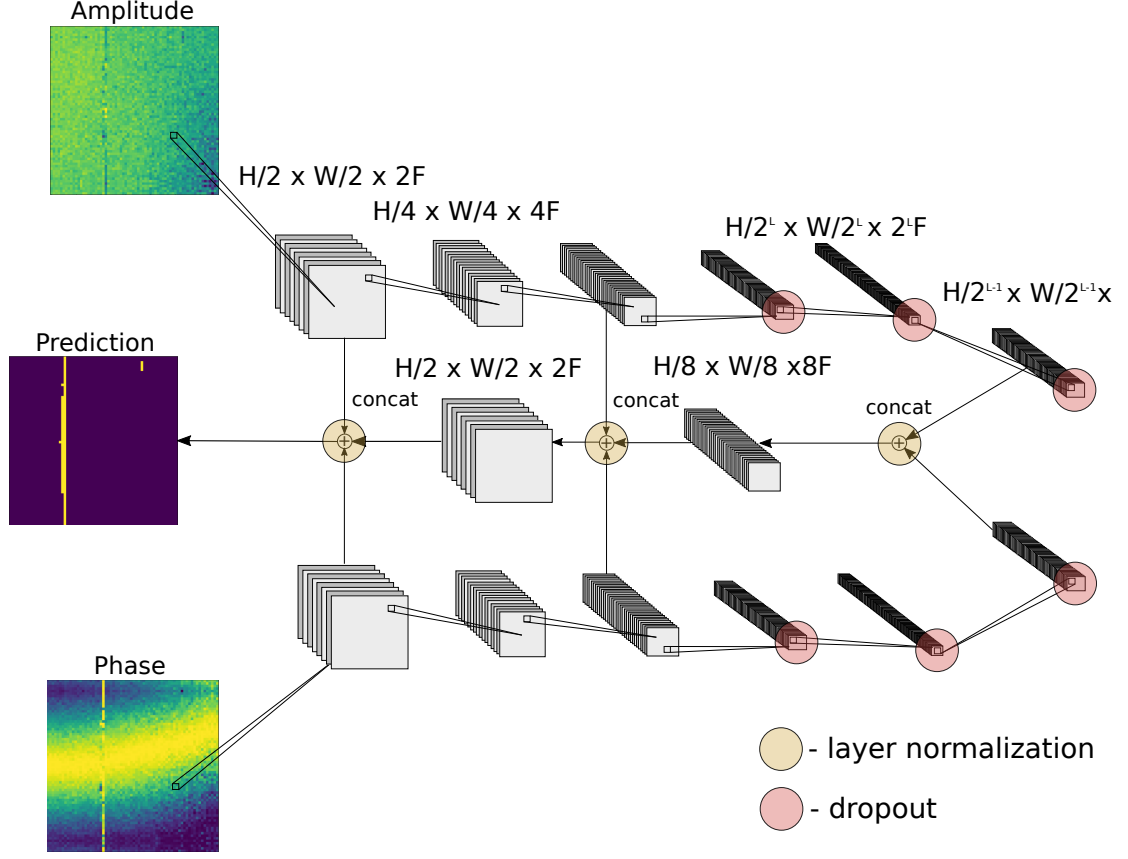


FIGURE 4.2: The general architecture for the amplitude-phase DFCN demonstrating the sliced in frequency, padded in both time and frequency, and finally normalized amplitude & phase input layers. H and W correspond to the input visibility dimensions in time and frequency, while F is the number of filter layers with L corresponding to the total number of layers between input and the fully convolutional layer. For reasons explained in Section 4.2.2, I use layer normalization at each skip connection and concatenation due to the difference in distributions of the amplitude and phase downsampling pathways. Every convolutional layer in the downsampling pathway is a $3 \times$ stacked set of convolutional layers with 3×3 kernels leading into an output convolutional layer with a 1×1 kernel, similar to the ‘Network in Network’ architecture of Lin, Chen, and Yan, 2013. All convolutional layers including the final output layer use the Leaky ReLU as their activation function.

extragalactic source on the sky (e.g. Fornax A). Extragalactic point sources are modeled using the discrete form of the visibility equation (Parsons et al., 2012a)

$$\tilde{V}(t, \nu) = \sum_n [\tilde{A}(\tau, \hat{s}) * \tilde{S}_n(\tau) * \delta(\tau_n - \tau)], \quad (4.1)$$

which depends upon the source delay position on the sky, τ , the source spectrum, $S_n(\nu)$, and the delay-dependent interferometer gains, $\tilde{A}(\tau, \hat{s})$, where a tilde represents the Fourier transform converting between frequency ν and delay τ and $*$ represents a convolution. Point source flux densities are drawn from a power-law distribution of the form $\Pr(S > S_{0.3 \text{ Jy}}) = \left(\frac{S}{S_{0.3 \text{ Jy}}}\right)^{-1.5}$ with a lower bound of 0.3 Jy. The spectral indices for these sources are then assigned uniformly at random between $-1 \leq \alpha_r \leq -\frac{1}{2}$ as per Hurley-Walker et al. (2017b), where $S_\nu \propto \left(\frac{\nu}{\nu_{\text{center}}}\right)^{\alpha_r}$. The source delays (sky positions) are also chosen according to a uniform random distribution. Each simulated waterfall visibility contains between $10^3 \leq N_{\text{srcs}} \leq 10^4$ sources with the aforementioned characteristics. I simulate diffuse galactic emissions with the de Oliveira-Costa et al. (2008) Global Sky Model (GSM) and an analytic form of the HERA primary beam (Parsons, 2015). GSM diffuse emissions are not precisely modeled but created to give a sky-like realization by sampling across LST and frequency, and applying a filter in time that has a fringe-rate corresponding to the baseline type being simulated. The visibility baseline types are uniformly sampled across LST, where baseline length, $|\vec{b}|$, is chosen according to a half-normal distribution with $\mu_{|\vec{b}|} = 7.5 \lambda$ and $\sigma_{|\vec{b}|} = 150 \lambda$. This is done to closely resemble the distribution of baseline lengths seen in HERA which is weighted towards short baselines. The learned model can be further tuned as longer baseline types are introduced.

I model RFI with four distinct classes: narrowband persistent (e.g. ORBCOMM), narrowband burst (e.g. ground/air communications), broadband burst (e.g. lightning), and random single time-frequency ‘blips’. Narrowband persistent constitutes the majority of RFI and are most often the brightest sources in HERA observations; these are empirically modeled. Narrowband bursts have no preference in duration or frequency but typically persist > 30 s and are simulated with a Gaussian profile in time to mimic the roll on/off

seen in HERA observations. Broadband bursts are rare events that exist across the entire HERA band at specific time integrations. These events are introduced in only 3% of the training data. I randomly inject ‘blips’ that are RFI with a duration of $\Delta t \leq 10$ s and frequency width, $\Delta \nu \leq 100$ kHz, which when taking into account HERA’s time and frequency resolution places this class of RFI into single visibility pixels.

To create the most comprehensive HERA visibility simulations to mimic real observations I include simplistic models of several important effects seen in the HERA signal chain. These effects include:

Cross-talk - An effect due to over-the-air coupling between nearby HERA receivers and dipole-arm coupling. This spurious correlation is mocked by convolving the simulated visibility with white noise.

HERA bandpass - Empirically derived from HERA bandpass measurements and fit to a 7th order polynomial (Parsons and Beardsley, 2017).

Gain fluctuations - Fluctuations are applied to the analytic HERA bandpass by introducing individual phase delays with a uniform spread between $-20 \leq \delta \tau \leq 20$ ns.

I simulate a training dataset of 1000 HERA observations of 10 minutes (60 time integrations) over the frequency range of 100-200 MHz (1024 frequency channels). The mean RFI occupancy rate for these simulated observations was $\sim 10\%$. This value differs from the $\sim 3\%$ which is the typically observed RFI environment in the Karoo Desert, South Africa seen in past HERA observations (Kohn, 2016) which used a simple statistical thresholding RFI algorithm. The comparison between the simulated and more recent real HERA dataset RFI occupancy rates across the band can be seen in Figure 4.3. I further expand this training dataset by performing data augmentation techniques on the reduced spectral windows. These techniques include mirroring over time and frequency, Gaussian random noise injection (correlated between amplitude and phase) with an amplitude that is at most 10% of

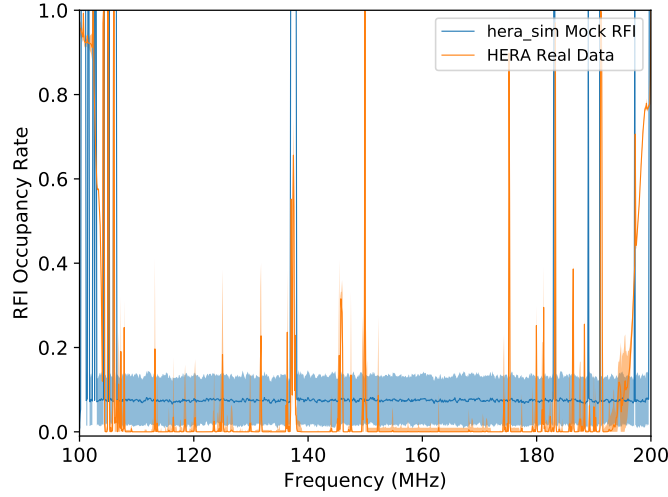


FIGURE 4.3: The hera_sim mock RFI (blue) occupancy rates across the band with its variance (blue region), as compared to RFI flagged in the real HERA data evaluation dataset (orange) with its own variance (orange region). The simulated RFI is overemphasized ($> 10\%$) in the training dataset. This is done in an attempt to balance the training due to RFI being a significantly sparse class which without would lead to more significance placed on Non-RFI when computing the loss.

$|V|$ the visibility amplitude and by translating a spectral window across the band creating unique window samples at varying frequency intervals. Using a translation in frequency has the intent of reducing over-fitting to steady state narrow-band RFI (e.g. ORBCOMM) because of repetitive sub-band samples entering the training dataset.

After increasing the simulated dataset volume through augmentation it is sliced into 16 spectral windows and padded according to Section 4.2.2 which results in 44800 unique spectral window visibilities each of size $68 \text{ time samples} \times 68 \text{ frequency channels}$. I separate this simulated dataset by an 80-20 split, where 80% of the simulated dataset is used for training and 20% is used for validation.

4.3 Evaluation

For the evaluation of the networks I used several datasets unseen in training. The real observed dataset used for evaluation consisted of HERA observing data from the 2017 -

2018 season, more specifically between the Julian Dates of 2458098 - 2458116, which I will just refer to as the real HERA dataset. The real HERA data were composed of raw uncalibrated visibilities that have been visually inspected and manually flagged by hand with high- and low-frequency band edges removed. Hand flagging was accomplished by looking in both amplitude and phase for sharp discontinuities and structure that exhibited an increase in power when compared to a fringing sky signal. The band edge removal is a precaution due to the large dynamic range roll-off, which makes discriminating between RFI and sky observations nearly impossible for humans and algorithms alike. This reduces the actual data evaluation passband to 896 frequency channels which covers $106 \leq \nu \leq 194$ MHz.

A simple approach to evaluation would rely on using the accuracy of predictions meaning I only look at the number of correctly predicted RFI pixels relative to all RFI, although this metric hides important details to the performance of the networks. This is an important consideration in this instance because the HERA data observations contain on average 3% of data corrupted by RFI, which means a blanket classification of “No-RFI” would yield an accuracy rate of 97%. To account for this class imbalance I evaluate the effectiveness of the networks by using several metrics commonly employed for classification. I use the standard metrics of Recall and Precision, which are defined as

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{RFI}_{\text{Correct}}}{\text{RFI}_{\text{Correct}} + \text{RFI}_{\text{Incorrect}}} \quad (4.2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{\text{RFI}_{\text{Correct}}}{\text{RFI}_{\text{Correct}} + \text{NoRFI}_{\text{Incorrect}}} \quad (4.3)$$

where I consider True Positives (TP) to be correctly predicted RFI pixels, False Negatives (FN) are RFI pixels identified as No-RFI, and False Positives (FP) are No-RFI pixels incorrectly identified as RFI. I can therefore understand Recall as the fraction of all RFI events identified by the flagging algorithm, and Precision the fraction of identified RFI that is actually RFI.

I also need a metric that is sensitive to a dataset with a sparse class, as in the case where

Algorithm	RFI Classes				
	No RFI Accuracy (%)	Narrowband Recall - Precision - F_2	Narrowband Burst “ ”	Broadband Burst “ ”	‘Blips’ “ ”
Amp DFCN	94	0.98 - 0.82 - 0.94	0.77 - 0.65 - 0.74	0.16 - 0.67 - 0.19	0.35 - 0.01 - 0.07
Amp-Phs DFCN	98	0.99 - 0.83 - 0.95	0.77 - 0.67 - 0.74	0.18 - 0.68 - 0.21	0.35 - 0.02 - 0.08
Watershed RFI	98	0.49 - 0.95 - 0.54	0.32 - 0.97 - 0.37	0.99 - 0.74 - 0.98	0.71 - 0.73 - 0.71

TABLE 4.3: RFI recovery metrics based on individual type of simulated RFI. I look at the Recall, Precision, and F_2 score for each of the three algorithms as simulated with hera.sim. All metrics here are determined under the condition that an optimal threshold value was found by maximising the F_2 score. The Recall and Precision rates are the average over 1000 simulated waterfall visibilities with the same simulation parameters for foregrounds and signal chain outlined in Section 4.2.3. Values in bold indicate the best achieved rate within each RFI type across algorithms.

RFI represents $< 3\%$ of the observations, and one that can condense the overall performance into a single metric. I therefore can use the binary classifier performance metric, the F-score which has the general form of

$$F_\beta = (1 + \beta^2) \frac{\text{Recall} \cdot \text{Precision}}{\beta^2 \cdot \text{Precision} + \text{Recall}}, \quad (4.4)$$

where I set $\beta = 2$ to preferentially weight Recall.² The F_2 score is therefore providing a metric that has an aggressive stance towards RFI while still being somewhat sensitive to false positive flagging. Due to the nature of measuring the 21cm EoR signal with HERA where I have collected sufficient data to not be noise limited, I can sacrifice good quality observations for the sake of reducing as much RFI contamination as possible; thus allowing a higher rate of FPs.

I compare three distinct algorithms: the amplitude-only DFCN, amplitude-phase DFCN, and the Watershed RFI algorithm. For a fair comparison I evaluate the DFCNs after both have converged independently of the number of training epochs, ensuring that they have learned the training dataset to their maximum capability. The networks are then checked for over-fitting by comparing that the evaluation loss is similar to that of each networks training loss when applied to the unseen 30% of simulated visibilities set aside for evaluation.

Before I approach the evaluation on the HERA data I can further optimize how the networks handle the shift in domain from simulation to observed. This is done by looking

²An F_β score with $\beta < 1$ describes a preference for weighting Precision over Recall.

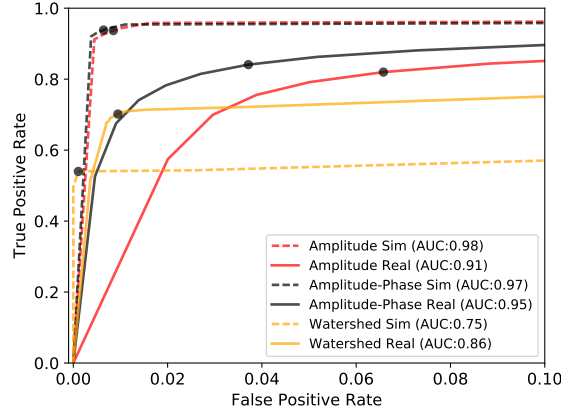


FIGURE 4.4: ROC curve showing a broad model comparison between all three RFI flagging algorithms, Amplitude DFCN (Red), Amplitude-Phase DFCN (Black), and Watershed (Orange). The ROC curves were derived from each algorithm predicting on real HERA data visibilities (solid) and simulated HERA visibilities (broken). Black circles represent the optimal F_2 score. The Area Under the Curve (AUC) metric condenses the overall performance of the algorithms and finds that the Amplitude-Phase network exhibits the best response on the real data with an AUC of 0.95. The TPR and FPRs for the real data (solid) are based on manually flagging RFI to the best of my ability to discern RFI from signals on the sky and therefore should not be taken as a ground truth.

at the Receiver Operating Characteristics (ROC) curves in Figure 4.4 of both the networks and the Watershed algorithm. The ROC curve gives a general idea of the performance of the networks by looking at how the True and False Positive rates respond to different thresholding values of the networks' softmax output layer. An optimal threshold is found per network/algorithm and per dataset allowing the freedom to tune the predictions for a particular instance. I determine this optimal thresholding value by using the maximum F_2 score across all thresholds which is shown to find a reasonable balance in TPR and FPR by locating the 'knee' of the ROC curve.

The results of each algorithm along with their performance metrics are reported in Table 4.3 as applied to simulated datasets. These metrics are performed on data that are unique from the previous training/evaluation datasets and include a control dataset which has no RFI present and four others that contain a single distinct class of RFI. An example of what each RFI class is modeled after is shown in Figure 4.5. In doing this, I can gauge how sensitive each algorithm is to a certain class of RFI. The DFCN networks

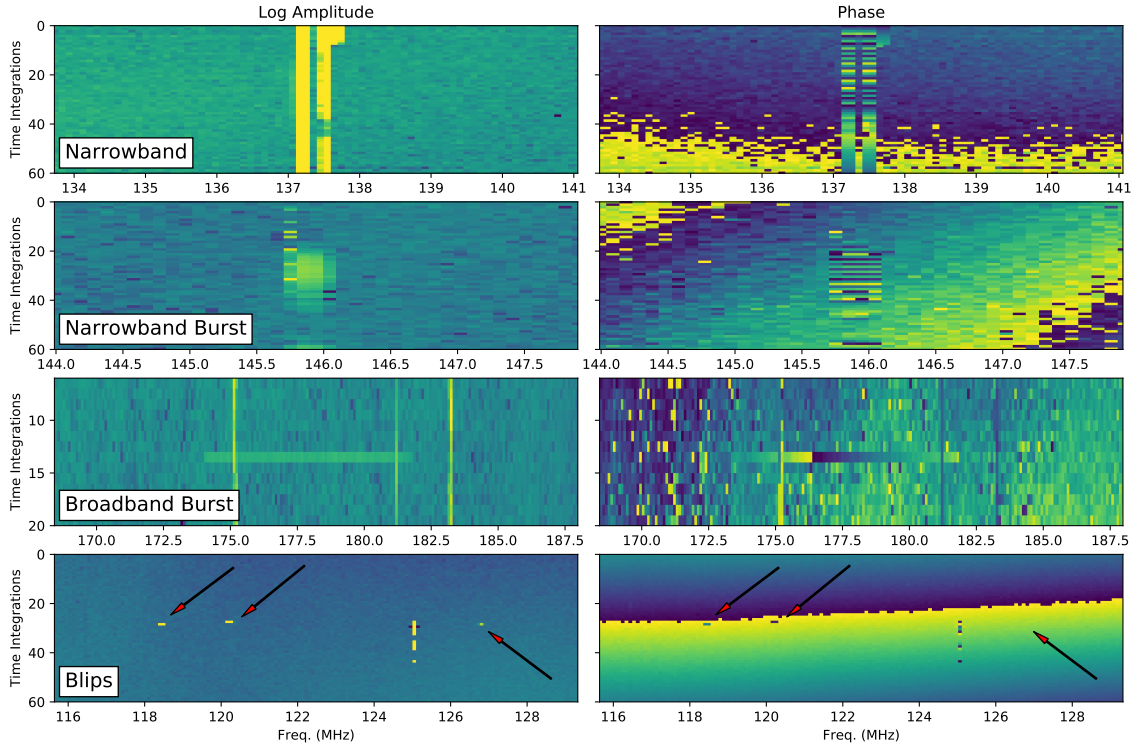


FIGURE 4.5: Examples of the four RFI classes from HERA data as they appear in amplitude and phase that I model in the simulations. Note the different time and frequency scales on each plot. The narrowband example (row 1) centered at a frequency of ~ 137.2 MHz is the ORBCOMM satellite system which is occasionally intermittent. Narrowband burst (row 2) is typically limited across only a few frequency channels (≤ 500 kHz) and has no consistent operating pattern over time. Broadband burst events (row 3) are short time duration (≤ 40 s) and can exist across the entire band (e.g. lightning) or in a sub-band as seen here flanked by the South African Broadcasting Corporation’s channel 4 video (175.15 MHz) and audio (181.15 MHz) broadcasts (Kohn, 2016). The ‘blips’ (row 4) demonstrate the one off nature of this sparse class as compared to the intermittent transmitter at frequency 125 MHz.

both perform well on the narrowband time persistent and burst RFI which is unsurprising as these simulations closely resemble the evaluation dataset and only differ in occurrence of events. However, both networks are inadequate for identifying Broadband Bursts and ‘blips’. This is understandable from a training perspective as both of these classes of RFI are going to be the last to be modeled in the networks as they account for only a minor fraction of all simulated RFI and lead to little overall optimization of the loss. This could potentially be remedied by placing more emphasis on these two classes of RFI in the training dataset or a much more in depth hyperparameter optimization of per-layer kernel sizes.

I compare all three algorithms as applied to a simulated `hera_sim` and real HERA dataset in Table 4.4. Looking at the prediction rates, both DFCN networks display an immense improvement over the Watershed RFI algorithm, boasting rates of $32\times$ and $22\times$ better than the Watershed, for the amplitude and amplitude-phase networks respectively. The faster amplitude only prediction rate compared to the amplitude-phase is unsurprising, as the number of parameters involved in an amplitude-phase prediction is roughly $1.5\times$ more and scales approximately proportional with the prediction rate. An example of these results, which serves to give an appropriate idea of the average performance as applied to a real HERA waterfall visibility is shown in Figure 4.6; how each compares on simulated data can be seen in Figure 4.7. Both DFCN networks have a tendency to over-predict RFI where there may not be any, however in the case of the narrowband RFI seen in frequency channels 175 MHz and 189 MHz, it may not be unreasonable to be more aggressive as leakage into adjacent channels can occur. This can be difficult to quantify of course as the ground truth of the real HERA data is unknown and RFI leakage can be easily masked by the sky.

4.4 Conclusions

Machine learning applications in the fields of astronomy and cosmology are rapidly developing, and in many cases are beginning to outperform the classical algorithms by way of increased speed and more accurate predictions. In this chapter I described an RFI identification approach to using a Deep Fully Convolutional Neural Network, which combined

Algorithm	hera_sim	HERA real	Prediction Rate
	Recall - Precision - F_2	“ ”	waterfall/h/GPU
Amp DFCN	0.90 - 0.61 - 0.82	0.76 - 0.42 - 0.65	2.4×10^5
Amp-Phs DFCN	0.90 - 0.82 - 0.88	0.81 - 0.58 - 0.75	1.6×10^5
Watershed RFI	0.53 - 0.95 - 0.58	0.64 - 0.88 - 0.68	7.4×10^3

TABLE 4.4: RFI recovery metrics for hera_sim simulated data containing signal chain effects with all classes of RFI and raw (uncalibrated) HERA observations from the 2017 - 2018 observing season. All results in the real HERA data column are based off of manually identified RFI and therefore the ground truth is uncertain especially in the low SNR limit for RFI. The real HERA data included observations from LSTs of $0 \leq t \leq 5$ h and across baseline lengths of $7 \leq |b| \leq 100 \lambda$. Values in bold correspond to the best achieved result for that metric. All three algorithms are ran on identical GPU resources.

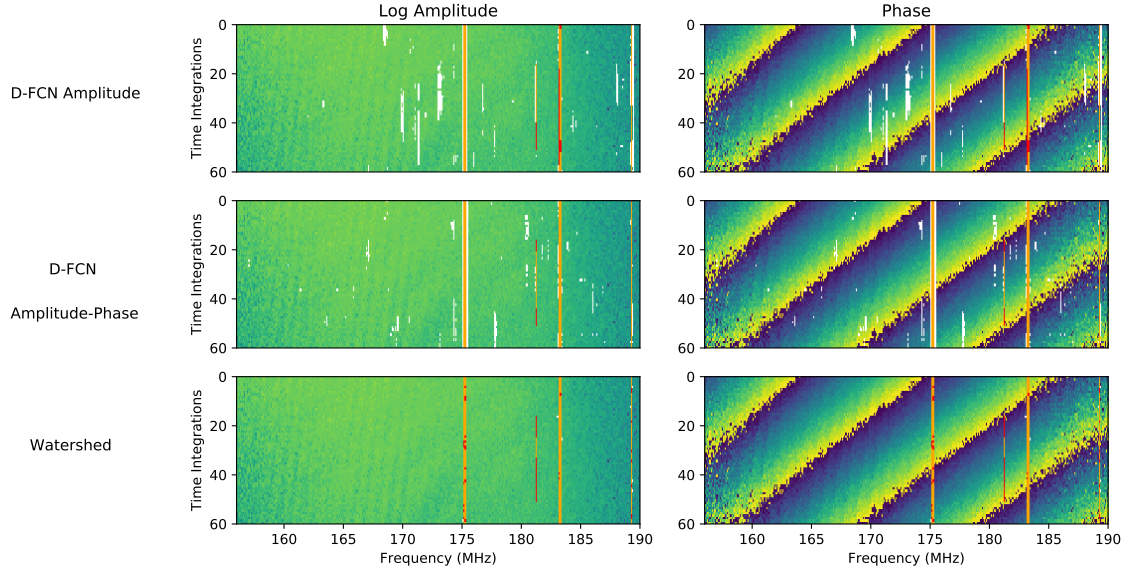


FIGURE 4.6: A comparison between the three flagging algorithms described in this chapter as applied to a sub-band (157 - 193 MHz) from the real HERA dataset, which has been flagged manually and is used as the approximate ground truth. Orange indicates true positives, white is false positives, and red represents false negatives. In this example the amplitude-phase fed DFCN ultimately has the best true positive outcome but, as seen in Table 4.4, both the DFCN algorithms take a more aggressive stance towards RFI resulting in higher rates of false positives when compared to the Watershed algorithm.

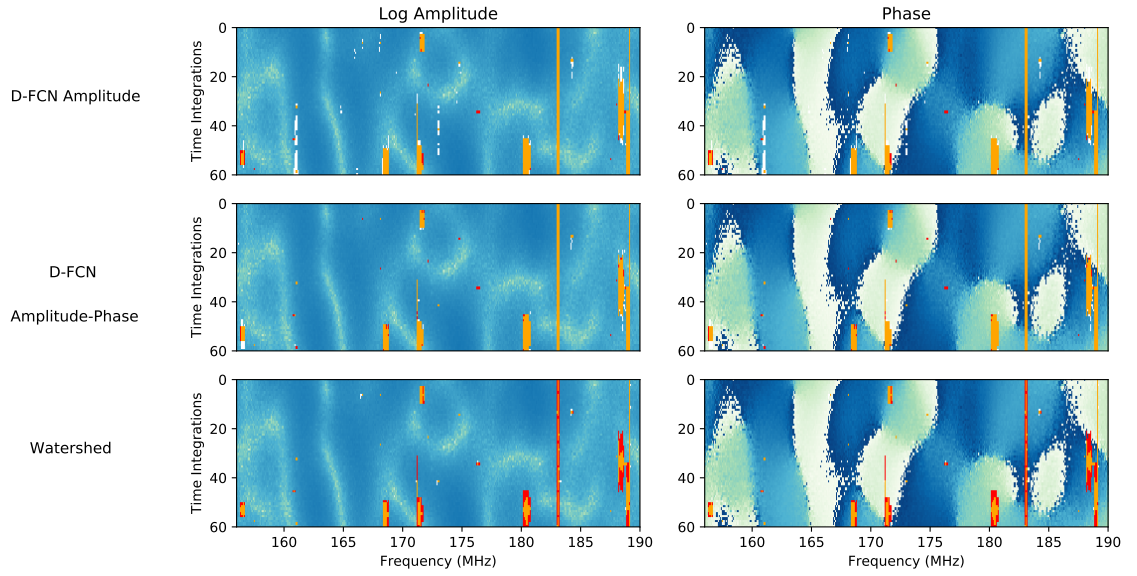


FIGURE 4.7: A similar comparison as in Figure 4.6 demonstrating how each RFI flagging algorithm performs on simulated HERA data from `hera_sim`. Orange indicates True Positive, white is False Positive, and red is False Negative. The simulated waterfall visibility is of a 25λ baseline dominated by strong diffuse emissions from the de Oliveira-Costa et al. (2008) GSM. The Watershed algorithm’s inability to discriminate between RFI and sky, as indicated by its higher false negative rate, in this instance hints that there is required fine-tuning of its kernel size and initial threshold hyperparameters, due to the spectral structure in the simulations.

the amplitude and phase of an interferometer’s measured visibility to predict which time-frequency pixels contained RFI. I compared this result to the Watershed RFI algorithm in the HERA data processing pipeline, and demonstrated that the DFCN approach was vastly more time efficient in its prediction with comparable to improved RFI identification rates. I also show that by including the phase component of the visibility I can mitigate the effects of domain shift between an entirely simulated HERA visibility training dataset and the observed validation dataset. This means that by improving the simulated model for HERA visibilities, coupled with an amplitude-phase DFCN I should be able to achieve an extremely effective first-round RFI flagger that reduces a common pipeline bottleneck. I do however recognize that the DFCN approach can have issues with identifying RFI bursts that occupy single time-frequency samples, what I called ‘blips’, and broadband bursts. An issue most likely due to an imbalanced representation in the training dataset, and the loss optimization not being rewarded enough to drive the DFCNs to learn a subclass that appears at a rate of $< 0.1\%$. This can be potentially overcome by fine-tuning the model by using transfer learning (Yosinski et al., 2014), and would involve a training dataset which consists almost entirely of these two subclasses, where the trained DFCN model shown here would serve as the starting point.

In near future build-outs of HERA there will need to be an extreme importance placed on reducing bottlenecks in the HERA data processing pipeline. The current Watershed RFI flagging algorithm does not scale particularly well, which puts this class of fully convolutional neural network as an ideal alternative. The eventual number of HERA dishes will total 350, which for a single 10 minute observation gives 61,075 unique waterfall visibilities. In the amplitude-phase DFCN design outlined in this chapter the RFI flagging throughput is 1.6×10^5 waterfalls/h/gpu³ which compares to the Watershed RFI flagger at 7.4×10^3 on the same resources.

Future work related to the amplitude-phase DFCN could include a modification to a similarly styled comprehensive data quality classifier which should in-turn lead to improved results for sky based (Barry et al., 2016) and redundant calibration (Zheng et al., 2014),

³Performed on a single NVIDIA GeForce GTX TITAN

both of which requires exceptionally conditioned data. A strict binary classifier could be achieved by developing a training dataset that doesn't use a mock sky, but an accurately modeled sky with a proper HERA beam model. Of course it would also be possible and might be better suited by developing an observation derived training dataset in this instance, as failure modes are generally easier to identify in visibilities as opposed to contamination by RFI.

It should also be possible to extend this work to arrays with better temporal resolution such as the MWA (Tingay et al., 2015) in the search for transients like Fast Radio Bursts (FRBs, Zhang et al. 2018). The additional phase information could potentially reduce the low-end limit of fluence for identification due to a more significant contrast between RFI and sky fringes.

The github repository for the RFI DFCN described in this chapter can be found at https://github.com/UPennEoR/ml_rfi.

Acknowledgments

A version of this chapter is currently under review as Kerrigan et al. 2019 MNRAS, and is reproduced with the permission of all coauthors and the copyright holder. Copyright 2019 Monthly Notices of the Royal Astronomical Society.

Chapter 5

HERA

In this chapter I introduce the Hydrogen Epoch of Reionization Array (HERA) in its under construction state as of writing this thesis. The purpose of this chapter is to establish a precision wide-field sky calibration procedure for HERA which also has the potential for identifying systematics such as signal chain reflections, cross-talk, and as of yet unidentified issues. I systematically describe each necessary component for successfully performing sky calibration on HERA observations. Following this, I look at two unique observing case studies of bright extragalactic sources on the sky as seen by HERA. Finally, I inspect the night-to-night and antenna-to-antenna variation to determine the stability of the array over a multi-night observing period.

5.1 Hydrogen Epoch of Reionization Array

HERA is a drift-scan array situated at the same Karoo Desert South Africa site as the former PAPER array. HERA is a designated precursor instrument for the future Square Kilometer Array (SKA). The HERA array at the time of this analysis was composed of 52 parabolic dishes in a compact configuration which I aptly refer to as HERA-52 for the remainder of this chapter. The antenna layout of the HERA-52 configuration is shown in Figure 5.1. Each dish has a diameter of 14.5m and uses re-purposed sleeved dipoles from the PAPER array for the antenna feeds. The observing band covers the frequency range of

100 - 200 MHz (redshift of $6.1 < z < 13.2$) with a frequency resolution of 97.8 kHz and has a field-of-view of 9° at 150 MHz (DeBoer et al., 2017).

5.2 Sky Calibration

Calibration is the process of correcting for instrumental effects present in the array and associating it with a proper physical scaling relative to the sky. Calibration has proven to be one of the more difficult aspects of pursuing the 21cm reionization signal. This has been such a concern that many 21cm cosmology radio interferometers have been purpose built to leverage redundancy in their layout to improve calibration. The strength of redundant calibration is that no *a priori* assumptions are made concerning the sky (Liu et al., 2010; Zheng et al., 2014). Unfortunately, this leaves the interferometer physically unrelated to the sky meaning the array is only internally consistent. Additionally, redundant calibration relies on every baseline visibility being perfectly redundant but this is typically not the case due to individual antenna characteristics. To relate the redundantly calibrated interferometer visibilities to the sky, a source catalog of extragalactic point sources must be used to fit for both an absolute amplitude and phase. The default approach for HERA is to do this with the Common Astronomy Software Applications (CASA) package and one or a few strong point sources in the sky to calibrate. This may not be optimal as many locations in the sky do not have very strong calibrators and the few locations in the southern sky that do are usually extended sources which add an additional layer of complexity to the calibration of a low spatial resolution interferometer. In addition to these limitations CASA doesn't support the use of a primary beam model. This pipeline for HERA can be found at https://github.com/HERA-Team/hera_cal which includes redundant calibration (Zheng et al., 2014; Liu, Parsons, and Trott, 2014c), CASA absolute calibration (Kern and Carilli, 2018), and averaging of nights in LST (*aka* LST binning) (Ali et al., 2015). I therefore look at an alternate approach for calibrating HERA observations by using the radio interferometry simulation and calibration software known as Fast Holographic Deconvolution (FHD).

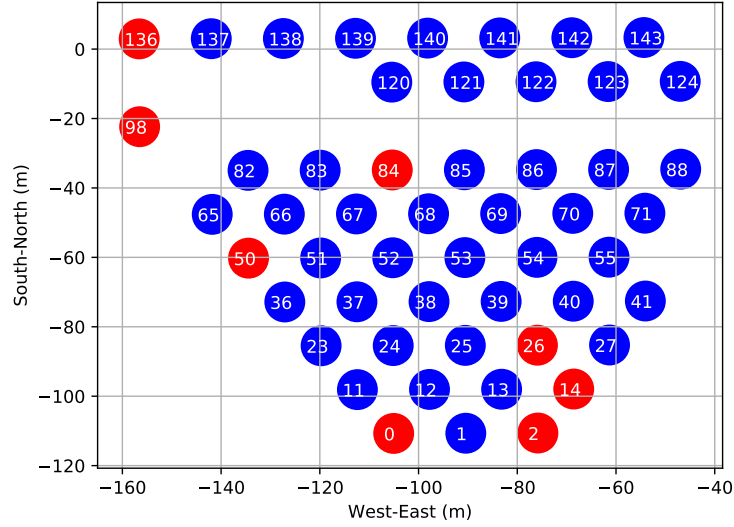


FIGURE 5.1: The antenna layout at the time the HERA-52 data was observed. Blue antennas are deemed to be suitable while red antennas are bad. Some noticeable differences between this work and Kern and Carilli, 2018 is that I determine antennas **11** and **121** to be stable when compared to the average of antenna calibrations and that antenna **136** is unstable (this antenna was not available at the time of the analysis presented in Kern and Carilli, 2018).

5.3 HERA Data

The methods outlined in this chapter work with raw preliminary HERA-52 data with RFI flagging performed using the legacy XRFI watershed algorithm which is the only post-processing from the Initial Data Release (IDR) 2 dataset. This means that no redundant calibration, removal of malfunctioning antennas, or handling of other array systematics has been allowed prior to handing off to FHD. I intend to use results of the FHD sky calibration process to measure and identify details such as bad data from malfunctioning antennas and array systematics. The dataset consists of observations from JD 2458098-2458116 with each daily observing period covering $1h32m \leq \text{RA} \leq 10h9m$. Each 10 minute observation is split into 12 smaller observations of 50 seconds to minimize error when calibrating. Therefore calibration solutions are fit per every 50 s observation per frequency channel and per antenna with no further attempt at reducing the degrees of freedom.

5.4 Fast Holographic Deconvolution

Fast Holographic Deconvolution¹ or FHD as it will be referred to in the remainder of this chapter is a software package for simulating and calibrating radio interferometry data. It is written in IDL and driven by bash scripts for performing batch processing. Running FHD for any specific radio interferometer requires several key components a **primary beam model**, **source catalog**, **parameter script**, and **observation data**.

HERA Primary Beam Model

Ideally, a primary beam model should simulate both the spatial and spectral response precisely because inaccuracies in either could introduce aliasing, contamination, or leakage into the power spectrum. One such example that is a concern for low frequency EoR studies is from primary beam chromaticity² (Thyagarajan et al., 2016a) which can introduce foreground power well into the EoR window therefore obscuring a detection. The HERA dish/antenna was purposefully designed with this element in mind and should have a reduced response to the beam chromaticity. The **beam model** supplied to FHD can be either an analytic representation that is computed on the fly (e.g. MWA Beam Model Version #2), a simple analytic model such as an Airy disk / Gaussian, or a pre-stored simulation. For this HERA-52 analysis I use a pre-stored beam model simulation³. The simulated beam model is from a CST EM⁴ simulation that numerically solves Maxwell's equations for simulated 3D RF engineering applications. The simulated HERA primary beam model is mapped to the sky in HEALpix (Górski et al., 2005) coordinates with an NSIDE of 128 and provides the resolution of a 1 MHz/channel spectral response of a single antenna over the observing band. This spectral response is more finely sampled using a cubic interpolation down to a frequency resolution of 97.8 kHz/channel to match the HERA passband resolution, and is inspected to ensure there is no aliasing.

¹The FHD development repository can be found at <https://github.com/EoRImaging/FHD>. This analysis was performed using the FHD branch HERA with a githash of 417241627c0bdd7c1dee3fffc0820c31d2fb354f.

²Not to be confused with the chromaticity introduced from the frequency dependence of the measured uv -plane Fourier modes, synthesized beam.

³HERA RAW CST primary beam models can be found at <https://github.com/Nicolas-Fagnoni/Simulations>.

⁴CST is the Computer Simulation Technology 3D simulator for electromagnetic fields.

Source Catalog

The **source catalog** used is typically the Galactic and Extragalactic All-Sky MWA (GLEAM) catalog which contains over 300,000 unique extragalactic point sources as found by the MWA. FHD does not require any particular source catalog and user defined or created ones can be used. These catalogs are simple arrays with information such as the (RA,Dec) positioning on the sky, flux of the source, α which is the source spectral index, and the frequency the flux measurement was observed at f_0 . Some acceptable versions of the GLEAM catalog for FHD include '[mwa_calibration_source_list_gleam_kgs_fhd_fornax.sav](#)', '[GLEAM_plus_rlb2017.sav](#)', and '[GLEAM_v2_plus_rlb2019.sav](#)'⁵. These catalogs are all found in the **FHD/catalog_data** directory and can be accessed independently from IDL by using scipy's `readsav()` function. Calibration of HERA data in this chapter exclusively used the '[GLEAM_plus_rlb2017.sav](#)' catalog.

Parameter Script

A **parameter script** is simply the script that sets important imaging, simulation, calibration, or deconvolution settings in the underlying FHD processes. A sample of some critical parameters used by FHD for HERA-52 calibration is shown in Table 5.1. These parameters are instrument dependent and therefore not all parameters used for one interferometer will translate well to another.

Observational Data

Observational data is important if using FHD for calibration and deconvolution/foreground subtraction. This data can be in the form of raw or pre-calibrated (redundantly calibrated) visibilities but must be in the **uvfits**⁶ file format. Overly flagged data due to RFI can present issues during calibration because of lack of convergence and the same goes for missed RFI which pollutes the image causing low SNR.

⁵Credit to Ruby Byrne for compiling the GLEAM+ catalogs which consist of the standard GLEAM components in addition to several extended and point sources.

⁶pyuvdata can be used for converting between filetypes such as Miriad, uvfits, and measurement sets. It can be found at <https://github.com/RadioAstronomySoftwareGroup/pyuvdata>

5.4.1 Capabilities

FHD can perform calibration, simulation, and imaging. The process of calibration involves the creation of model visibilities implying that for every calibration you are also performing a simulation. If foreground subtraction is performed FHD can calibrate the data against a source catalog that it will simulate model visibilities for providing an output of calibrated observed, model, and residual visibilities. Residual in this case means the subtraction of the model visibilities from the calibrated observed visibilities and is typically referred to as foreground subtraction.

5.4.2 Under the hood

FHD relative to other radio interferometer simulators such as PRISim⁷ (Thyagarajan et al., 2015) and pyuvsim⁸ (Lanman et al., *In Review*) is fast but potentially trades off precision for that speed. This speed increase is directly attributed to the holographic mapping formulation

$$\mathbf{H}(\mathbf{u}, \mathbf{u}) = \tilde{\mathbf{B}}^T(\mathbf{u}, \mathbf{V})\mathbf{B}(\mathbf{V}, \mathbf{u}), \quad (5.1)$$

which reduces computational time by directly mapping the model visibilities $\mathbf{B}(\mathbf{V}, \mathbf{u})$ and holographic antenna beams $\tilde{\mathbf{B}}^T(\mathbf{u}, \mathbf{V})$ from their locations in the model uv plane $\mathbf{u}$ to their position in the holographic map of visibilities $\mathbf{V}$ i.e. $\mathbf{V} \leftrightarrow \mathbf{u}$ (Sullivan et al., 2012). This is in contrast to the need for other simulators to integrate across the uv -plane to create model visibilities which is very time consuming. The holographic mapping function $\mathbf{H}(\mathbf{u}, \mathbf{u})$ is computed for each observation and polarization.

For calibration FHD performs a least squares regression on the χ^2 equation

$$\chi^2 = \sum_{ij} \frac{|v_{ij} - g_i g_j^* m_{ij}|^2}{\sigma_{ij}^2}, \quad (5.2)$$

where m_{ij} is the FHD generated model visibility, g_i the gain from the i th antenna, v_{ij} the raw visibility which contains noise, σ_{ij} the noise variance of the i, j baseline visibility and *

⁷<https://github.com/nithyanandan/PRISim>

⁸<https://github.com/RadioAstronomySoftwareGroup/pyuvsim>

represents the complex conjugate. This is solved per frequency channel and per polarization. By default FHD solves for an average calibration across all times which can be problematic if the sky transits substantially over an observation, which for a standard 10 minute HERA observation gives a difference of $\delta\theta = 2.5^\circ$ between the initial and final time integrations. Applying these time-averaged calibrations over all integrations would introduce increasing amplitude and phase errors approaching both ends of the observation. This is accounted for in this work by slicing HERA-52 observations into smaller 50 s ($\delta\theta = 12.5'$) observations, down from their typical 10 minute totals. Doing this should reduce any time dependent calibration errors down by a factor of ~ 5 provided the center RA of the observation is used for calibration.

In this analysis I look at only the XX⁹ polarization, because while the YY polarization beam model is nearly identical to the XX beam model up to a 90° rotation this was not properly performed in the primary beam model on hand. In the default calibration pipeline for the Murchison Widefield Array (MWA), smoothing is performed independently on amplitude and phase with a low order polynomial using the FHD keywords `cal_amp_degree_fit` and `cal_phase_degree_fit`, with the amplitude order being set to 2 and phase order set to 1. I do not perform any smoothing of the calibration solutions in this chapter. However, if I were to smooth the calibration solutions for HERA, a significant increase in the amplitude degree would be required due to the more complex nature of the HERA passband. In that instance, a polynomial of $\mathcal{O}(p) \geq 5$ has been found to be sufficient.

5.4.3 Caveats

FHD has been proven to have issues in the past, most notably pointed out in Chapter 3 and Appendix A (Published in Kerrigan et al., 2018) where Precision Array for Probing the Epoch of Reionization (PAPER) observations were modeled in FHD. The resulting model visibilities had issues at high-delay modes (into the EoR window) where aliasing due to *uv*-gridding causes these errors to be transferred to the residual foreground subtracted observations. This high delay aliasing can be reduced with increased sampling rates of

⁹FHD uses the XX, YY polarization convention in contrast to the convention adopted by the HERA collaboration which uses E for XX and N for YY.

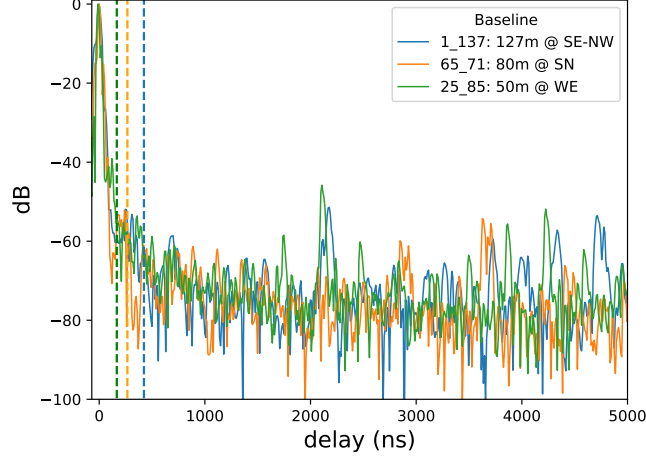


FIGURE 5.2: FHD simulated model delay spectra demonstrating the aliasing errors from the uv gridding process for several baseline lengths at different orientations. The vertical dashed lines represent the horizon limit for each baseline type. As per Kerrigan et al., 2018, to drive down this high delay aliasing ‘noise’ the uv -plane gridding resolution must be significantly increased but this has the considerable trade-off of more demanding computational needs. The uv -plane gridding resolution used here is 1.27λ .

the uv -plane. However it becomes extremely costly computationally. Unfortunately these high-delay modeling errors still persist for HERA (see Figure 5.2).

Therefore as of now there are only two relevant solutions that exist without dealing with this directly within the FHD code. One is to filter out high delays in the model visibilities as was performed in Chapter 3, removing this aliasing which in turn also removes a good deal of the beam main- and side-lobe leakage that we’d like to mitigate in the first place. The other is calibrating an entire observing night of HERA observations using an extremely fine uv -plane resolution to reduce the aliasing and then transfer the model and calibration solutions to every other observing night. The latter is of course the more appropriate solution as it still allows for capturing and correcting of the beam leakage into the EoR Window as long as the high delay calibration errors can be reduced to a sub-EoR level. Obtaining a more accurate beam model is more important before this effect becomes relevant. At this stage I do not perform either of these solutions as there is a more pressing issue of understanding the instrument calibration and reliability.

An additional caveat includes the inability to currently use diffuse maps for calibration

or subtraction; this is due to both not having precise diffuse catalogs and a not entirely fleshed out diffuse modeling code infrastructure. A first attempt at integrating a diffuse map for calibration could potentially be made by using the de Oliveira-Costa et al., 2008 GSM diffuse sky map or the newer more high resolution GSM from Zheng et al., 2017. These diffuse sky maps do pose some conflicts if being used in conjunction with the previously mentioned GLEAM and modified GLEAM point source catalogs, because the diffuse maps also include extended sources such as Fornax A. If calibrating off a complete diffuse + point source sky there will be significant flux over-counting and therefore either these extended sources need to be pruned from the point source catalog or regions of the sky need to be avoided. As of this thesis however these features are still intended to be developed as options for FHD and will be included in future versions.

5.5 Sky Calibration

5.5.1 Runtime

The runtime for FHD calibration on a single sliced (5 time integrations by 1024 frequency channels) HERA observation using the parameters from Table 5.1 on the resources of 2 cpus with 20 GB of memory is ~ 30 minutes. This runtime is highly dependent on the FHD parameters. The key parameters are the FoV and dimension which implicitly set the resolution and dimension of the uv plane which dramatically increases memory requirements for more dense uv plane samplings.

5.5.2 Calibration Case Studies

To gain an understanding of how well I am calibrating the observations I use two distinct cases. In the extreme case I look to an observation containing the extended extragalactic source of Fornax A (RA 3h 22m). For the more mild case I use the case study from Kern and Carilli, 2018 (RA 2h 00m) which contains two bright point sources at the center of the beam. Additional concerns that I cannot and do not attempt to control for in these two instances is diffuse emission and polarization. While these issues are not the primary

Parameter	Values (units)	Definition
FoV	45 ($^{\circ}$)	Field of view, directly sets the extent in uv -space
dimension	1024 (px)	Dimension of the image in pixels
instrument	hera	Important for initializing several array dependent properties (e.g. beam model)
freq_start	110 (MHz)	Low end bandpass cutoff
freq_end	190 (MHz)	High end bandpass cutoff
nfreq_avg	1	Rate of frequency channel averaging
n_pol	2	Number of polarizations to calibrate (4 pol would include XX,YY,XY,YX)
calibrate_visibilities	1	Allow FHD to calibrate visibility data
flag_calibration	1	Allow FHD to flag poor calibrations
calibration_polyfit	0	Apply a polynomial fit to the amplitude and phase solutions
bandpass_calibrate	0	Apply a global bandpass fit to antennas
firstpass	1	Calibrate and subtract data using pre-made source catalogs
min_baseline	0 (λ)	Minimum baseline to simulate and image
min_cal_baseline	25 (λ)	Minimum baseline to use during calibration

TABLE 5.1: FHD parameters for calibrating HERA-52 data. These parameters have been determined by maximizing the amount of power removed from the visibilities during foreground subtraction, implying that a proper calibration should result in correspondence to the model visibilities simulated by FHD.

concern for this work they do require a more in-depth exploration such as was covered in Kohn et al., 2018 and Dillon et al., 2018. The FHD calibration software does have the ability to perform calibration using data from all 4 polarizations (XX, YY, XY, YX) provided the correct E-field beam is used.

Fornax A: Bright Extended Off Zenith Source

In this section I try to understand how a bright extended source that is off zenith affects the calibration results. I approach this by looking at how I model Fornax A as either a many component extended source, or a single point source and additionally how it affects the calibration as it moves in and out of the sidelobes. In this analysis it is important to note that the HERA beam model is relatively untrustworthy below 10% of its maximum. This is a problem in all cases but is more important here because Fornax A never transits through the beam center and will spend the majority of its time within the poorly modeled portion of the main lobe and will remain a continued nuisance well into the beam’s sidelobes. Due to the longest baseline $b_{\text{max}} = 92\lambda$ in this HERA-52 dataset giving an angular resolution of $\delta\theta = 37.4'$ at 150 MHz, HERA-52 is right at the edge of being able to resolve Fornax A which has an angular size (radio lobe-to-lobe) of $\sim 60'$. This means that an extended source model for Fornax A (shown in Figure 5.4) should be used. Indeed when attempting to run this exact same calibration scheme but with Fornax A modeled as a point source (flux summed into a single component) the results are poor with a residual image that has substantial over-subtraction of Fornax A and its sidelobes. The Dirty, Model, and Residual sky images of Fornax A are shown in Figure 5.3. As Fornax A is the brightest extragalactic source in HERA’s observing field, artifacts from malfunctioning antennae, a poor primary beam model, phase errors, dipole-arm leakage, or source position errors should worsen the calibration and final sky image. Looking at the residual image there are two mis-subtractions located at $(-28^\circ, 50^\circ)$ and $(-21^\circ, 41^\circ)$ which are left over from the sidelobes of Fornax A. Incorrect modeling of Fornax A in this instance could suggest it’s a primary beam model issue because if the source model was incorrect I should see mis-subtraction doubles (positive and negative source residuals) in more locations.

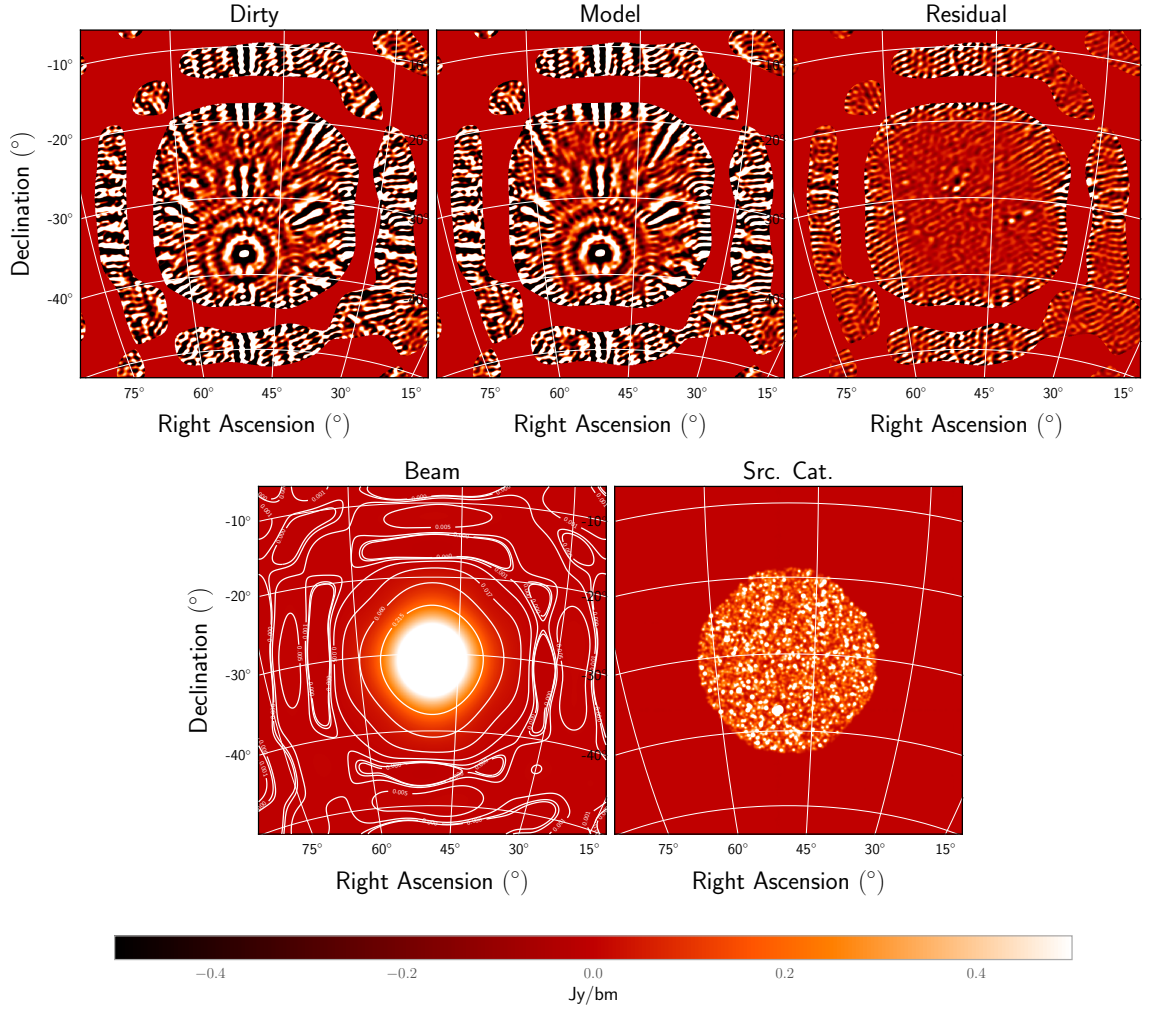


FIGURE 5.3: Observing JD 2458098 of Fornax A as Multi-Frequency Synthesis (MFS) imaged with HERA-52 and calibrated using FHD with the GLEAM source catalog. The calibration and subtraction parameters used for these images were the same as used from §5.1. In this case Fornax A is modeled as several individual components ($N_{\text{comps}} = 1925$) as opposed to a single approximate point source; an example of this extended source model can be seen in Figure 5.4. An indication that the beam model is not correct can be seen in the mis-subtraction of Fornax A’s sidelobes.

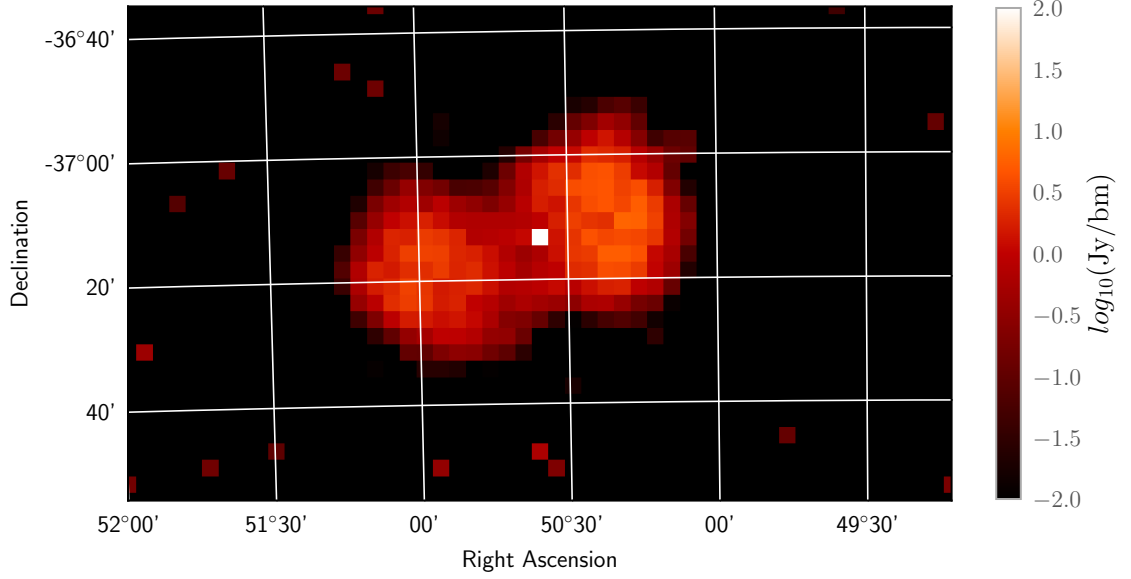


FIGURE 5.4: The extended source GLEAM model used for Fornax A at 150 MHz with a peak flux of ~ 313 Jy. A total of 1925 components are used for this model.

GLEAM J0200-3053: Point Source

For a comparison to the absolute calibration work performed by Kern and Carilli, 2018 I look at the exact same case of point source J0200-3053 where zenith of the observation is RA 2h 00m 12.7s. This source is ideal as a calibrator in comparison to Fornax A because it is approximately a point source and transits directly through the peak of the primary beam, where beam model errors should be minimal. Looking to Figure 5.5 the dirty, model, and residual sky images are shown and at first glance the dirty and model appear to be in agreement. Compared to the Fornax A case both of the sources near zenith appear accounted for in the main lobe of the primary beam. The flux density remaining at the first sidelobes of the primary beam however are somewhat expected as the HERA primary beam model at those levels is highly questionable.

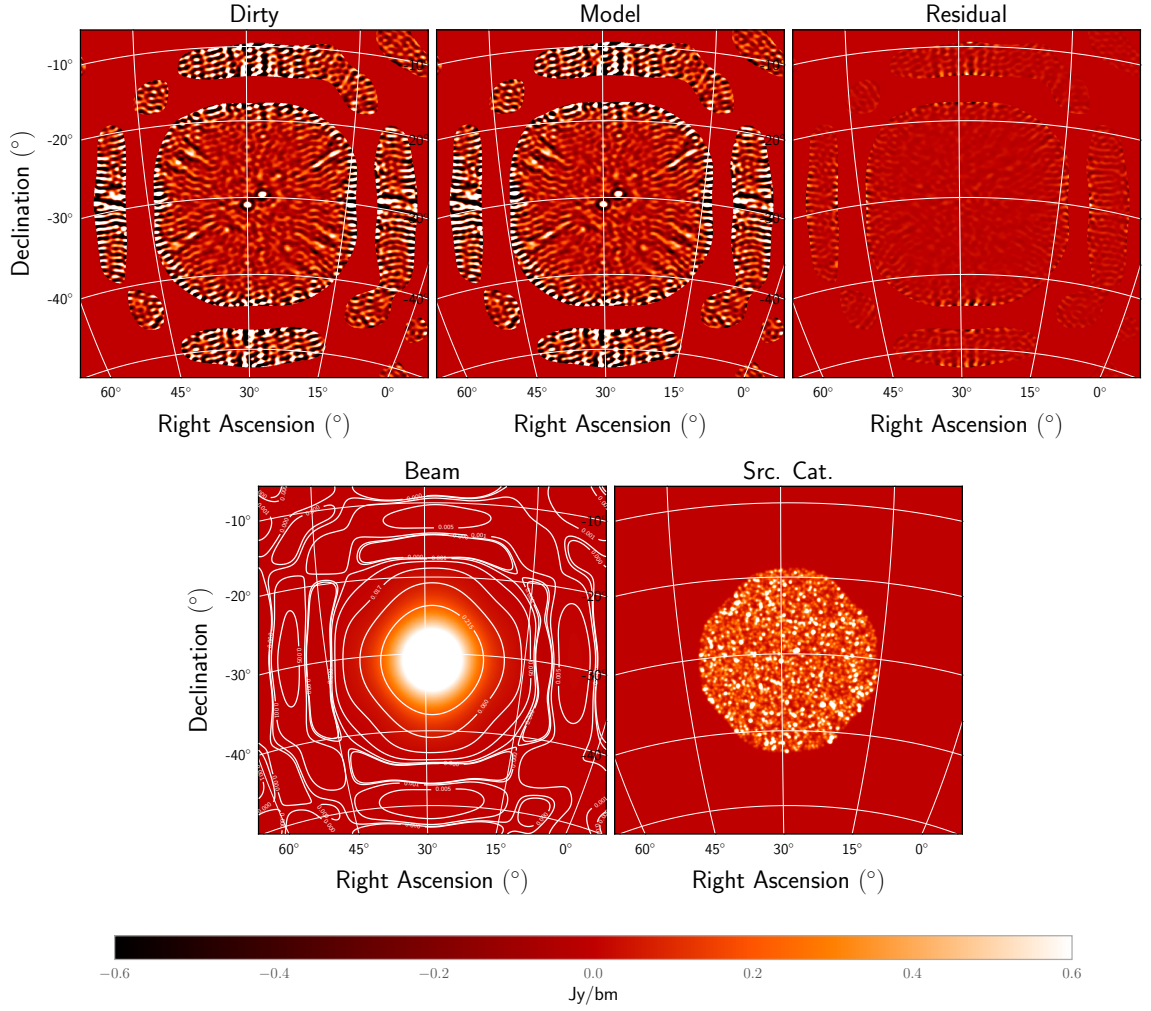


FIGURE 5.5: The same field as covered in Kern and Carilli, 2018 which used GLEAM J0200-3053 as a primary calibrator. In this example I attempt to calibrate using an entire field of sources ($N_{\text{sources}} = 8238$) all located within 10% of the beam max. I can see that with sources that are much better approximations to point sources and are located beam center I have significant improvements in the Residual image for both the main- and side- lobes.

5.6 HERA-52 Calibration Stability

5.6.1 Per Day Calibration Solution Waterfalls

The HERA-52 dataset consists of 18 nights of observation and to get an idea of how stable the array is night to night I can look to the calibration solutions over all nights and LSTs. A waterfall sampling of antenna 53's gain amplitude calibration solutions across all frequency channels (1024) and from $1.5 < t < 11.5$ h LST are shown in Figure 5.6. Two features that stick out are the large gain amplitudes (dark purple) centered at ~ 8 h LST and the contrast in the gain drop from Fornax A at 3.2 h LST. The feature at 8 h LST is from the sky model deteriorating as the GLEAM catalog loses point sources to the Galactic plane which causes the calibration to try and account for power in the sky that is in the raw visibilities but is not in the model visibilities. The sharp drop in the gains during the Fornax A transit is less obvious however but points heavily to poorly modeled beam sidelobes. The sudden decrease in gain amplitude is more easily seen in Figure 5.7, where slices through LST over the frequency channels of 130 MHz (red) and 160 MHz (black) demonstrate the time variability of calibration solutions. The double trough of the gain drop between $2.5 < t < 3.5$ h appears to peak at the beam midpoint in RA as Fornax A goes between off-center RA and DEC.

Generally speaking, antenna 53 over the course of the 18 observing nights appears to be quite stable. The night-to-night gain calibrations are within $\sim 2\%$ of the initial observing JD with little to negligible drift if any at all. To extrapolate this finding to the broader array I can compare the calibration solutions of antenna 53 to the remaining antennae in the array.

5.6.2 Per Antenna Correlation

To demonstrate the degree to which calibration amplitude solutions vary between antenna to antenna I form a correlation matrix in Figure 5.8 of the gains $|g(\nu)_i|$ from antenna i with

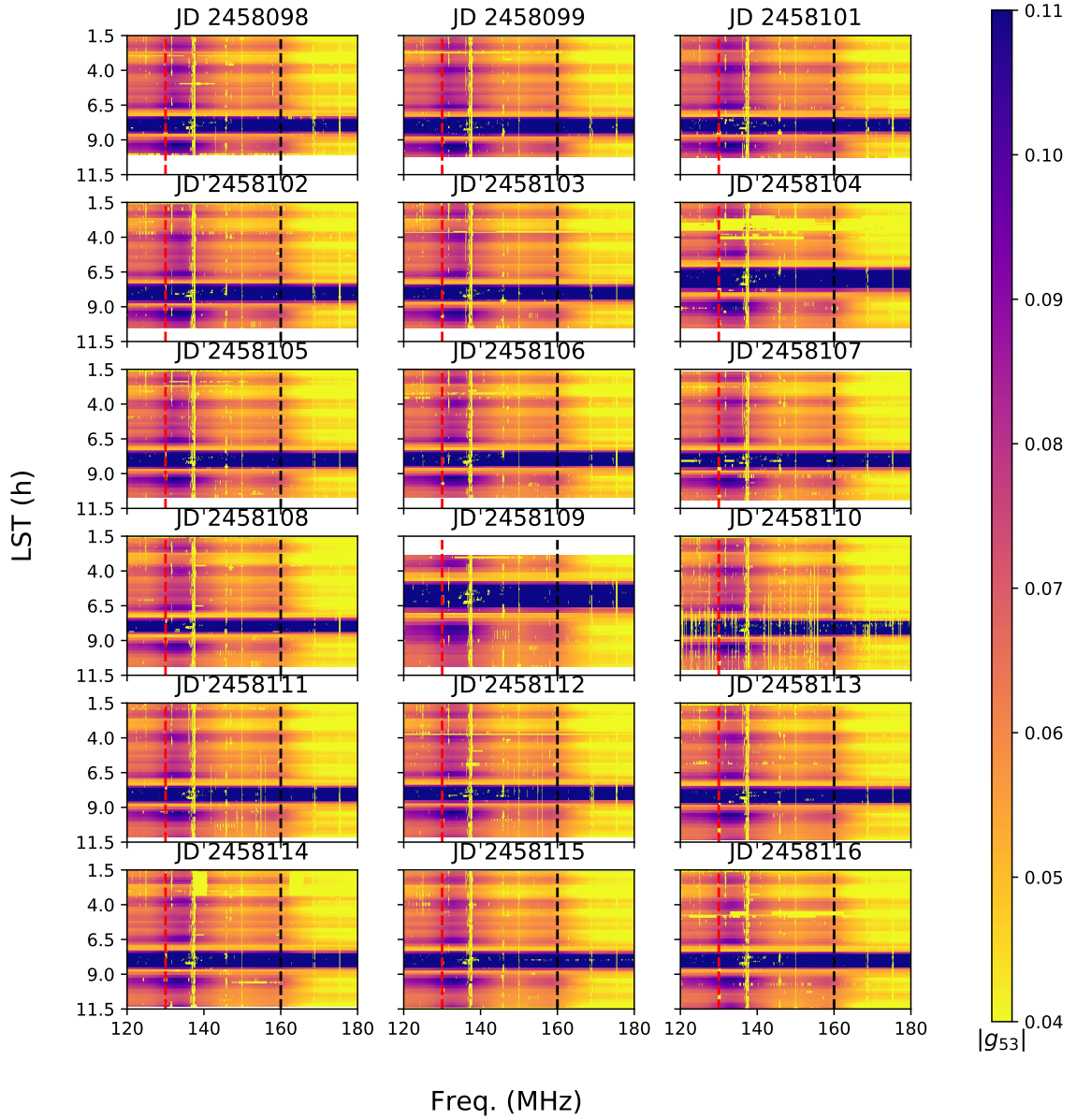


FIGURE 5.6: The per observing night gain calibration amplitudes for antenna 53. The regions of missing gains are from RFI removal performed prior to calibrating in FHD. The bright(dark purple) broad feature in the gain amplitudes is due to FHD attempting to account for power on the sky from our own galaxy which GLEAM does not provide sources for resulting in meaningless amplitude solutions. The vertical dashed red and black lines correspond to slices taken in frequency and shown in Figure 5.7

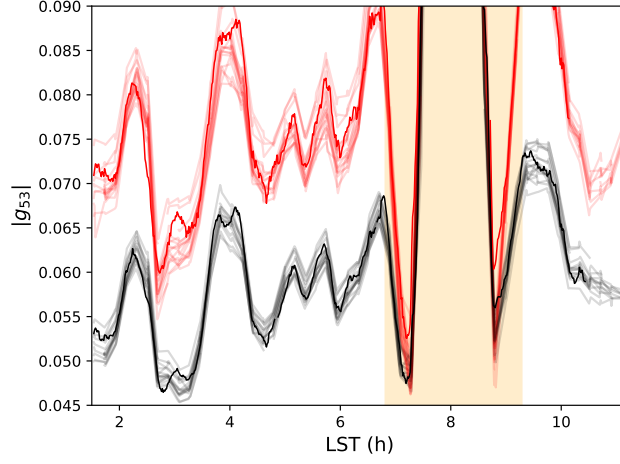


FIGURE 5.7: Gain amplitude slices for antenna 53 over LST at center frequencies of 130 MHz (red) and 160 MHz (black) demonstrating the time dependent structure from calibration. Gains were sampled roughly every 10 minutes for the background ensemble of calibrations that have reduced opacity. The opaque red and black lines are more finely sampled at 10 second intervals from JD 2458101 to give a better idea of the structure over LST. The dip seen centered at ~ 3.4 hrs is when Fornax A is at zenith. The orange shaded region represents the incomplete sky catalog due to the Milky Way.

the gains $|g(\nu)_j|$ from antenna j . Therefore each (i, j) pixel is the correlation coefficient of

$$R_{i,j} = \frac{C_{i,j}}{\sqrt{C_{i,i} * C_{j,j}}}, \quad (5.3)$$

where $C_{i,j} = \text{cov}(|g(\nu)_i|, |g(\nu)_j|)$. I must take additional care when building these correlations as the discontinuities introduced by RFI flagging produce stronger correlations between antennae than I should expect and I therefore mask this missing data. Looking at Figure 5.8 I can see that there are two classes of broadband gain amplitude solutions between antenna pairs. The two distinct modes can be seen in the passband gain solutions in Figure 5.9 and appear to be linked to the difference in front-end modules used. For reference, front-end modules here contain a series of Low Noise Amplifiers (LNAs) which are then attached to a hybrid coupler (balun) and should be impedance matched (DeBoer et al., 2017) with the output signal path to reduce ripples in the passband. Antenna's having a front-end with a nomenclature of 'FEA' are **51, 65, 66, 70, 82, 83, 98, 120, 121, 122, 123, 124, 141, 142, 143** and correspond to the blue curves from Figure 5.9. Using

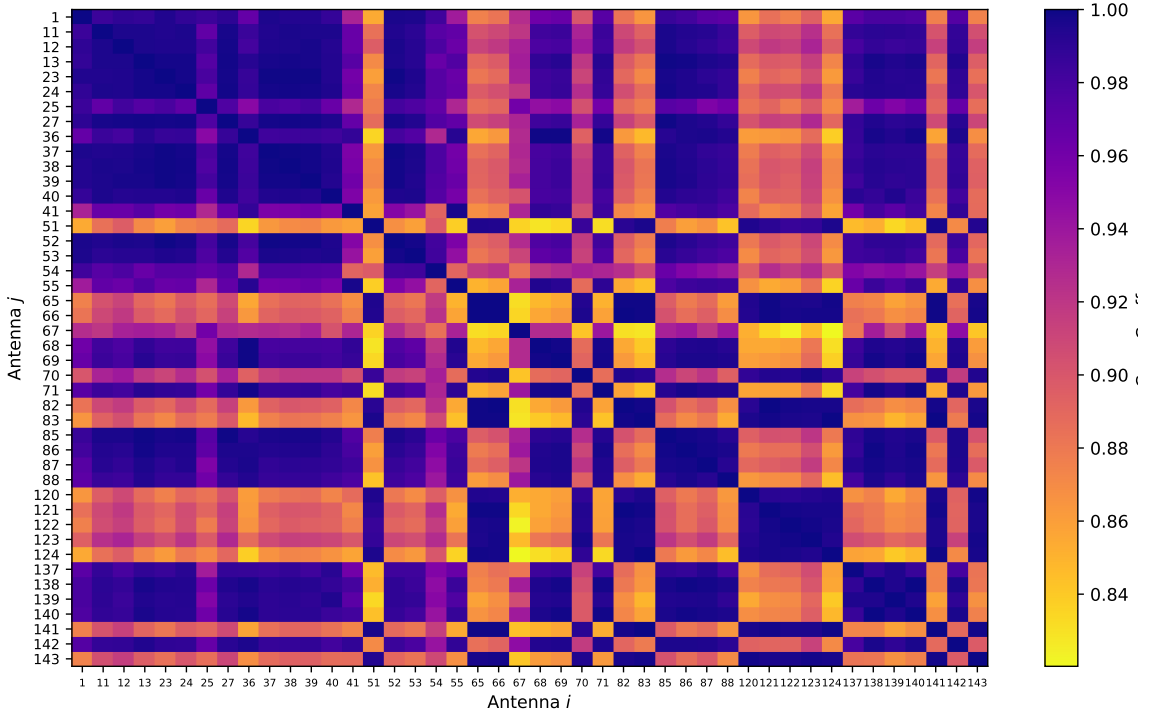


FIGURE 5.8: The correlation matrix between all antennas that were deemed functional from JD 2458101 at 2.36h LST. The two antenna modes can be easily distinguished here as their per frequency gain amplitude solutions do not correlate particularly well with one another.

both Figure 5.8 and Figure 5.9 the degree to which the broader array and the intra-mode correspondence can be approximated. The average antenna to antenna intra-mode variation for the passband across mode 1 is $< 9\%$ and for mode 2 is $< 12\%$. According to the correlation matrix antenna **67** is behaving uniquely in so much that it does not fit either of the two modes identified. Also interesting is that while antenna **142** has a front-end ‘FEA’ module and should be within the secondary mode, it corresponds more closely to the first mode implying that its passband ripples may not be as attenuated. If both antenna **67** and **142** are removed the intra-mode variation for mode 1 reduces to $< 7\%$.

5.7 Calibration Results

I find that overall the preliminary HERA-52 dataset and array is nominally stable over the 18 nights of observation. However, several precautions should be taken for the purpose

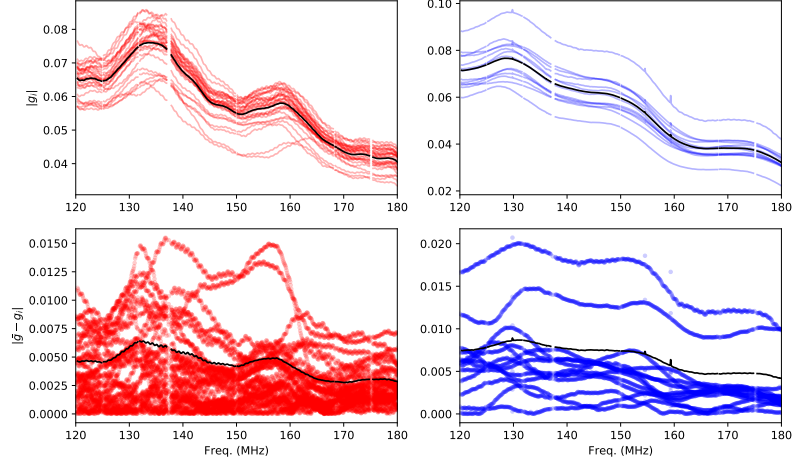


FIGURE 5.9: Plots demonstrating the two distinct modes (left and right columns) seen in the HERA-52 dataset. The bottom two figures show the difference between each antenna’s gains g_i and the mode group average $\bar{g}_i$. The black curves represent the average of antennas per mode per frequency (top) and the standard deviation per antenna mode per frequency (bottom).

of performing any type of power spectrum analysis such as removal of antenna **67**, and **142** as they both seem to be exhibiting behavior outside the two operating modes. For cuts along observing night, JD 2458104 and 2458109 should be excluded as they appear to have some time stamp issues and don’t align time-wise with the rest of the observing nights. Based on the calibration of these 18 nights the average variation across the array and bandpass is $\sim 10\%$, with the night-to-night variation being $\sim 2\%$. Looking to the FHD sky calibrated waterfall visibilities in Figure 5.10, there is close agreement with the FHD model to the observed data. Baseline ‘13.25’ displays fairly close to the same amount of power post-subtraction due to the calibration baseline length threshold set in the initial parameters and also because diffuse emissions are stronger in short baselines. Residual power seen in the mid and long baselines see a substantial amount of foreground power removal. The longer baseline waterfall visibilities where power was left behind in the wake of Fornax A’s removal could very well be due to the HERA primary beam model’s spectral response being insufficiently modeled or possibly the requirement for an even higher fidelity Fornax A model. In either case an improved HERA primary beam model will be needed as its limitations are still apparent from Figure 5.7 where calibration gain amplitudes attempt to account for bright sources away from beam center and down into the sidelobes.

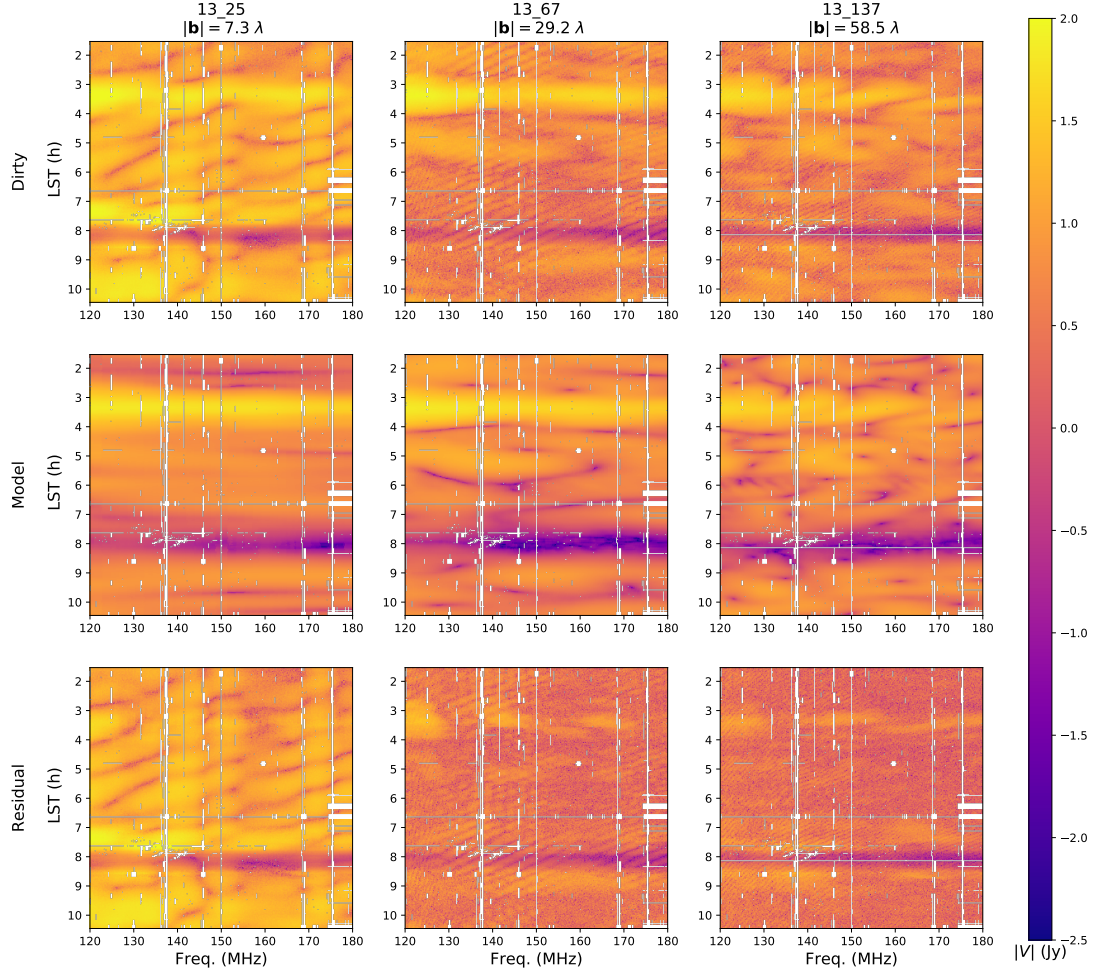


FIGURE 5.10: Waterfall visibilities of the FHD calibrated HERA data (top row), Model (middle row), and Residual (bottom row) for baselines (13,25) (shortest), (13,67), and (13,137) (longest). As should be expected, baseline (13,25) has the most power due to galactic diffuse emission and also minimal foreground mitigation in the residual waterfall visibilities. Baselines (13,67) and (13,137) were included in the calibration process while (13,25) was not due to the minimum baseline threshold of 25λ as I don't want to attempt a calibration of the shortest baselines without a diffuse sky model. Missing data (white) is due to RFI flagging.

5.8 Conclusion

The calibration pipeline and analysis presented here offers an additional avenue for validating the default HERA redundant and absolute calibration pipeline. Recent efforts have been made to explore the limitations of both sky calibration (Barry et al., 2016) and redundant calibration (Byrne et al., 2019), which places more emphasis on being aware of the bias inherent in either calibration procedure. A common theme for limiting the precision of these calibrations stems from an incomplete sky model. While the GLEAM catalog is an amazing resource for extragalactic point sources as of right now, the inclusion of a diffuse sky model native to the EoR observing band could lead to a significant leap in calibration results. A proper diffuse sky catalog is however left to a future endeavor as there are more immediate needs for the development of HERA. For example a simulated complex E-field primary beam would give FHD the ability to calculate the full proper Jones matrices and calibrate across polarizations to account for polarization leakage (Carozzi and Woan, 2009; Geil, Gaensler, and Wyithe, 2011; Kohn et al., 2018). This is a project that is being worked on as of writing this thesis, with the results of this chapter being a perfect comparison for any proceeding HERA calibration performed with the complex E-field beam.

Chapter 6

21cm EoR Parameter Estimation

The 21cm EoR signal is expected to have a non-Gaussian redshift dependent signature (Majumdar et al., 2018). Neural networks have shown to be highly capable of regressing and classifying signals with non-Gaussian features. In this chapter I outline a convolutional neural network for handling cosmological redshift slice cubes for exploiting this ability. More importantly I introduce a method for quantifying the intrinsic uncertainty from a neural network model coupled with the noise estimate of the observed data and how they propagate to the output regressed parameters. The work presented in this chapter is preliminary, however it does establish a foundational approach to 21cm parameter estimation using neural networks.

6.1 21cm Simulations Dataset

The cosmological dataset used for this parameter estimation analysis originates from La Plante et al., 2014. These cosmological simulations are designed using a mixed hydrodynamic and semi-analytic approach. To create an extensive dataset covering many realizations of reionization, additional hydrodynamic simulations which include radiative transfer are run using smaller volumes. These limited size volumes then provide a statistical measure for determining the time dependence of reionization and can be applied in a semi-analytic sense to much larger volumes with differing distributions of matter. I will now provide a

short introduction into these simulations to better understand their assumptions but for a more in depth understanding see Battaglia et al., 2013. In the formalism for these semi-numerical simulations the matter overdensity field is defined as

$$\delta_m(x) = \frac{\rho_m(x) - \bar{\rho}}{\bar{\rho}}, \quad (6.1)$$

which for large scales ($> 1 h^{-1}\text{Mpc}$) closely traces the reionization field. The reionization field can be defined as

$$\delta_z(x) = \frac{[z_{re}(x) + 1] - [\bar{z} + 1]}{\bar{z} + 1}, \quad (6.2)$$

where $\bar{z}$ is the mean redshift of reionization and z_{re} is the reionization field at some position x . This approximation can be made based on the ‘inside-out’ reionization scenario where the denser regions containing stars and galaxies reionize first and then emit UV radiation beyond which reionizes the lower density regions (La Plante et al., 2014). The correspondence between these fields can be quantified by looking at the difference in amplitude using the bias parameter

$$b_{zm}^2(k) = \frac{\langle \delta_z^* \delta_z \rangle_k}{\langle \delta_m^* \delta_m \rangle_k}, \quad (6.3)$$

which is parameterized with

$$b_{zm}(k) = \frac{b_0}{(1 + k/k_0)^\alpha}, \quad (6.4)$$

and is defined in k -space where k is a spherical wavenumber and b_0 is determined in the limit where $k \rightarrow 0$ (La Plante and Ntampaka, 2018; Barkana and Loeb, 2004). The values of k_0 and α are free parameters meant to be set based on the underlying statistics of a more comprehensive simulation and together with a chosen average redshift of reionization $\bar{z}$, a reionization scenario is specified. For creating scenarios where reionization was quick or drawn out k_0 , and α act as tunable parameters. The semi-analytic parameters k_0 and α for the simulations used in this work were fit from the underlying statistics of a hydrodynamic N-body simulation with radiative transfer. While these semi-analytic simulations are based on some approximations, they stand apart from the assumptions used in the semi-analytic 21cmFAST simulator (Mesinger, Furlanetto, and Cen, 2011) but offer a complementary

approach. Furthermore, they suffice to provide a proof-of-concept demonstration of my approach, which can then be applied to even more complex simulations at a later date. For a more complete understanding of this semi-analytic reionization technique see La Plante et al., 2014; Battaglia et al., 2013.

In general, the simulation dataset consists of several thousand unique cosmological realizations of reionization over the redshifts of $6.0 \leq z \leq 14.7$. Each one of these realizations is made of 30 redshift sliced 21cm signal ‘layers’ with an area of $4 (h^{-1}\text{Gpc})^2$. The line-of-sight evolution of reionization over a single slice is minimal to negligible (La Plante et al., 2014) and therefore the slices can be considered comoving. An example of a single realization of reionization over redshift is shown in Figure 6.1. I define a realization here as any reionization scenario that has unique initial conditions (i.e. the 3 free simulation parameters are unique). The parameters that I want to estimate through regression are the midpoint of reionization $z_{50\%}$, the mean redshift of reionization $\bar{z}$ set in the initial simulation, and the duration of reionization Δz . The midpoint $z_{50\%}$ is the redshift in the reionization simulation at which 50% of the cosmological volume becomes ionized. The mean of reionization $\bar{z}$ is the average redshift of reionization which has an important distinction from $z_{50\%}$ because there exist asymmetric reionization scenarios. The duration of reionization is defined as $\Delta z = z_{25\%} - z_{75\%}$ or the difference in redshift between when the cosmological volume was 25% and 75% reionized. All of these parameters can be extracted from the simulations given the proper preparation of the dataset. This preparation is described in Figure 6.2, which demonstrates how I can build the cosmological slice per redshift cubes. Every slice in the cubes contain important spatial information about the state of reionization, such as the distribution of the 21cm signal. To closely match the results as seen by an interferometer, the ‘foreground wedge’ and angular resolution of HERA-350 is added to the cosmological cubes (see Chapter 2). This is performed by removing Fourier modes where $k_{\perp} > k_{\perp, \text{max}}$ to match the angular resolution of HERA-350 and $k_{\parallel} \leq m k_{\perp}$ to mimic extreme foreground mitigation, where $k_{\perp, \text{max}} = 0.5 h^{-1}\text{Mpc}$ and $m \approx 3$ (La Plante and Ntampaka, 2018).

The stacked redshift slice cube thus contains the optimal amount of information concerning the reionization history and motivates a perfect use case for a convolutional neural

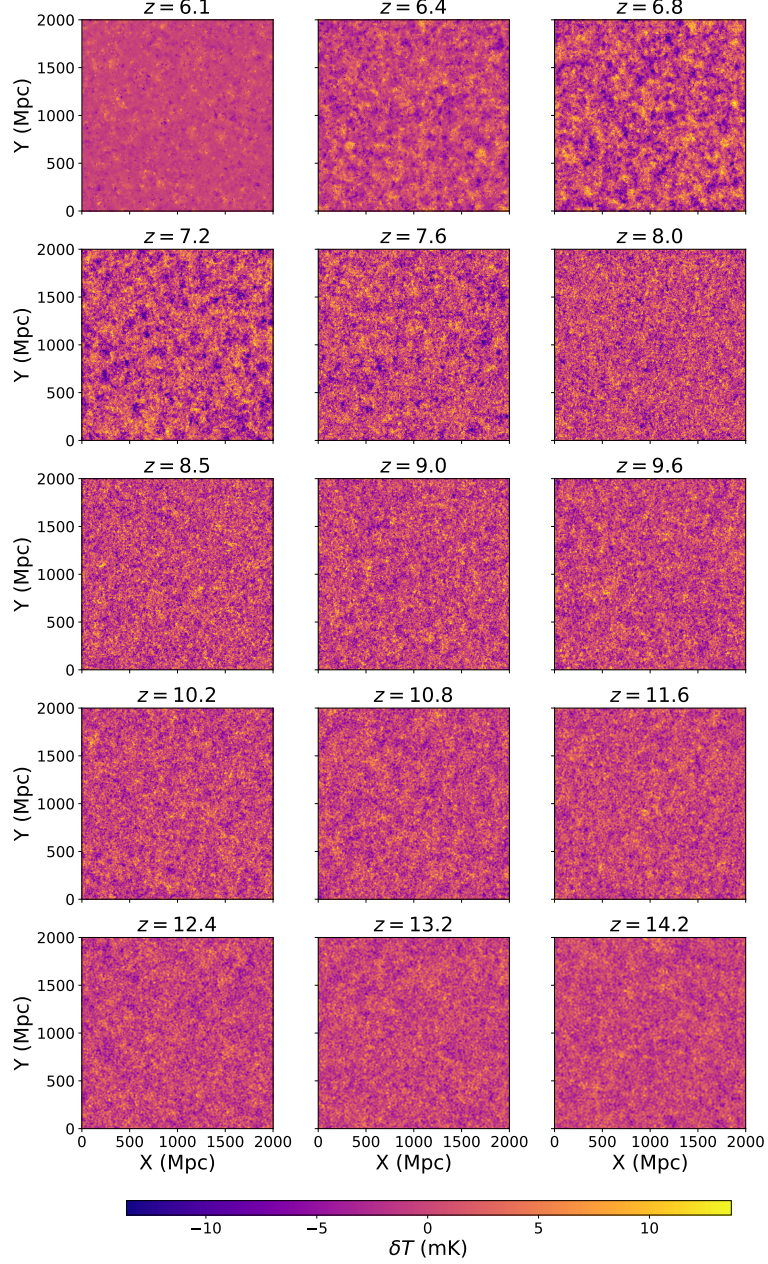


FIGURE 6.1: An example of a 21cm cube sliced over redshift from the semi-analytic simulation training dataset. The ‘foreground wedge’ has been removed to mimic the extreme foreground mitigation strategy for reducing contamination along with removal of $k_{\perp}$ modes beyond those achievable by HERA-350 ($k_{\perp} \geq 0.5 h^{-1}\text{Mpc}$). The midpoint of reionization in this scenario is $z_{50\%} = 6.83$ with a duration of $\Delta z = 0.8$.

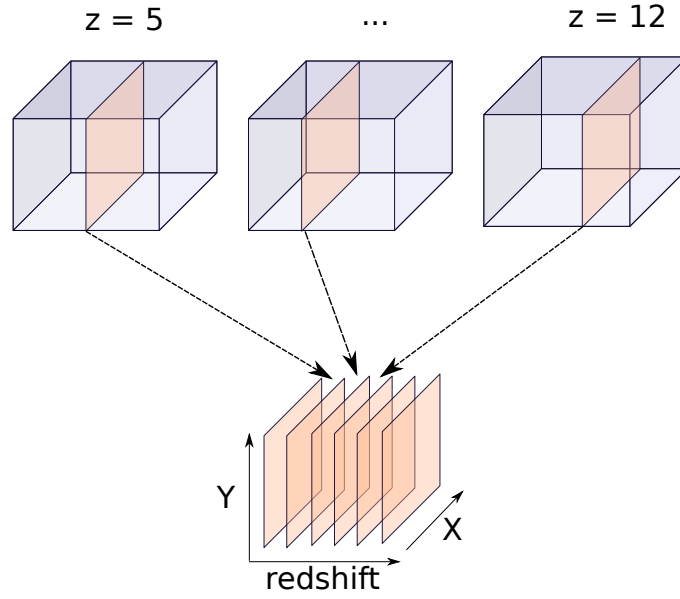


FIGURE 6.2: Cosmological volumes of size 2 Gpc on all sides are sliced into 2d images (X, Y physical space coordinates) at random orientations in each cube across redshifts z , at a redshift interval of $\delta z(z)$. The resulting cosmological slice per redshift cubes describe the progression of reionization, which contain the right amount of information for regressing on parameters such as the midpoint of reionization $z_{50\%}$, mean of reionization $\bar{z}$, and the duration of reionization Δz .

network.

6.2 Modified Convolutional Neural Network

I set apart the convolutional network used in this chapter from the one demonstrated as an effective method for mitigating RFI in Chapter 4 by noting the differences in architecture and predictive target space. In this model I forego any type of skip architecture or ‘ResNet’ style implementations. While I believe these are applicable in this instance, I use a simple architecture here for highlighting the uncertainty propagation aspect. The next biggest change to the architecture in this type of Fully Convolutional Neural Network (FCN) that I need to use for parameter regression is a Mean Squared Error (MSE) loss function which is simply

$$\text{MSE} = \frac{1}{n} \sum_i (y_t - y_p)^2 \quad (6.5)$$

where y_t are the target values and y_p the predicted values. In comparison, the output values for a neural network classification task are logits, which are a proxy to the probability per class estimates and require a loss function like the cross-entropy loss seen in Chapter 4. The regression task however requires the prediction to be a continuous variable, which is obtained by using the MSE as the loss function for the 3 parameters of interest $z_{50\%}$, Δz , and $\bar{z}$.

6.3 Parameter Recovery with an FCN

This work is based off of the Fully Connected Convolutional Neural Network application to parameter estimation on 21cm cosmological volumes presented in La Plante and Ntampaka (2018), where I replace the dense network layers with a Fully Convolutional Network. This improved type of network which I refer to as an FCN¹ is agnostic to the input cosmological cube shape because all layers involved are convolutional and do not have the rigid constraint

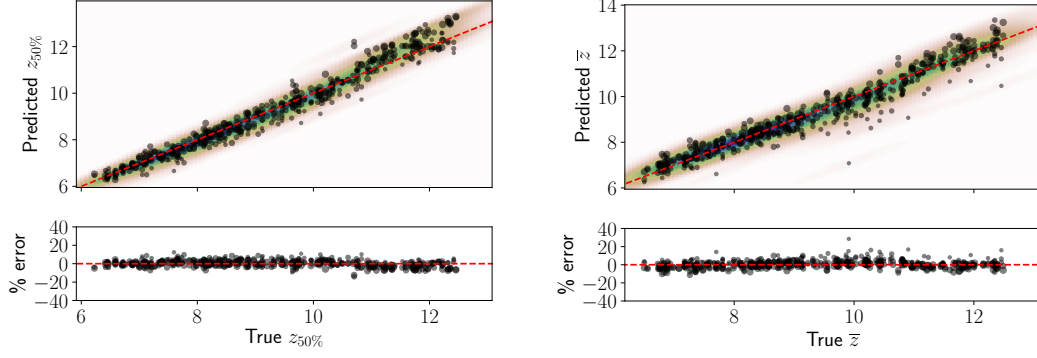
¹I do not label the type of network here as ‘deep’ because this network is subtly different in the sense that it does not have as large a training dataset to train on as in Chapter 4 and the depth of the filters and layers are sufficiently shallower by comparison. Admittedly the definition of a ‘deep’ network in the machine learning literature is a bit nebulous.

of node specified dense layers (every pixel is assigned a node on input). The only restriction in this case is maintaining the proper cosmological per-pixel scales ($\text{Mpc } h^{-1}/\text{px}$) used in the training data.

For comparison, I trained this network for 25 & 500 epochs, where a single epoch is a pass through of the entire training dataset, using an 80/20 training/test split from 1000 unique 21cm cosmological redshift slice cubes. I demonstrate in Figure 6.3 the successful recovery of the 3 parameters Δz , $z_{50\%}$, and $\bar{z}$ using only the moderately trained (25 epochs) FCN. I then show in Figure 6.4, that with substantially more training (500 epochs) $\bar{z}$ and $z_{50\%}$ can be recovered to within $\pm 4\%$ of the true parameters and Δz to within $\pm 10\%$ for long duration reionization scenarios. The model prediction bias seen at late reionization scenarios (low $z_{50\%}$, $\bar{z}$) requires additional investigation, but is potentially due to the missing cosmological k -modes removed in an attempt to imitate the angular resolution and foreground wedge. These preliminary results demonstrate the ability of the FCN to recover meaningful physical parameters from simulations. For this simple example there was only a 21cm cosmological signal, but more complex examples could be attempted which aim to recover these reionization parameters in the presence of foreground contamination or other instrument systematics.

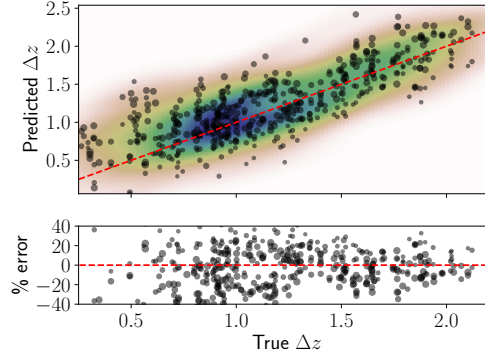
6.4 EKF Formalism

Many machine learning and more specifically neural network applications seen in astronomy and astrophysics today lack an estimate of the uncertainty associated with their prediction. Any real observational data that is input into a trained neural network model will have its own measurement noise, but there is also the noise contribution from the trained model's weights and biases that will accumulate in the prediction. I approach this concern with the aim of applying a trained FCN on simulated 21cm cosmological volumes in conjunction with the novel uncertainty propagation technique outlined in Titensky, Jananthan, and Kepner, 2018. This technique uses Extended Kalman Filtering (EKF) to determine the inherent uncertainty in a deep neural network (DNN) when applied to the MNIST digits classification problem. EKF is the special case application of the standard Kalman filter



(a) Midpoint of reionization

(b) Mean of reionization



(c) Duration of reionization

FIGURE 6.3: The recovered parameters of reionization as predicted by the trained Fully Convolutional Neural Network. Data point size corresponds to physical scales of the input cosmological cubes ($250\text{Mpc} \leq X, Y \leq 1000\text{Mpc}$) and the colored distribution is that of a kernel density estimate of the regressed parameters. I am able to regress on all three parameters $z_{50\%}$ (a), $\bar{z}$ (b), and Δz (c) simultaneously. When predicting $z_{50\%}$ and $\bar{z}$ I can achieve $\pm 10\%$ of the expected value. Predicting Δz however becomes more difficult as the cosmological slice sampling in z is sparse and at low Δz ($\Delta z < 0.7$) it begins to approach the redshift intervals $\delta(z)$ of the cubes resulting in poorer predictions.

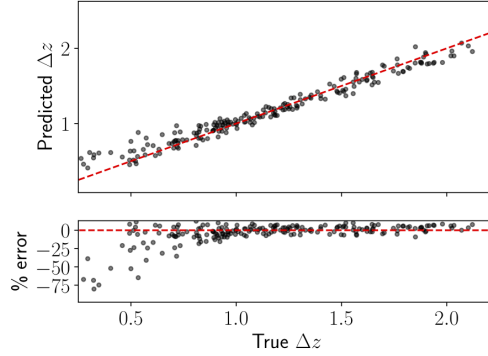
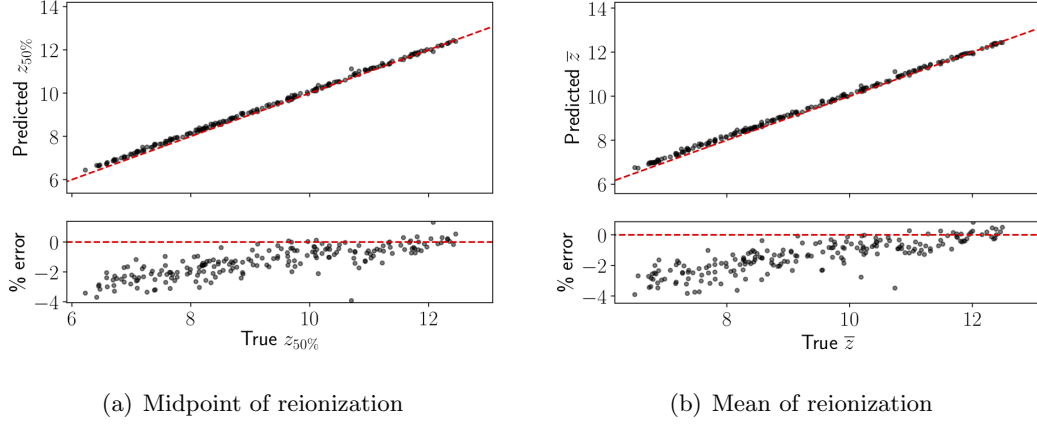


FIGURE 6.4: The same FCN as in Figure 6.3 with a model trained to 500 epochs. Even with the cosmological spatial modes removed due to the HERA-350 angular resolution and foreground wedge, the network is capable of recovering $z_{50\%}$ and $\bar{z}$ to within $\pm 4\%$ but shows a bias in the model for recovering late reionization scenarios. The recovery of duration Δz is also improved upon to well within $\pm 10\%$ at long durations ($\Delta z > 1$) but struggles with overfitting for short duration reionizations.

for state estimates of nonlinear systems. By treating every hidden layer in the DNN as a discrete time step it becomes an ideal case for neural networks as they are by design nonlinear through the choice of the layer activation functions.

To propagate the uncertainty of the FCN model to the predictions I begin with the EKF equations for dense neural networks outlined in Titensky, Jananthan, and Kepner, 2018, which are

$$\mathbf{x}_\ell = f(\mathbf{W}_\ell \cdot \mathbf{x}_{\ell-1} + \mathbf{b}_\ell), \quad (6.6)$$

and

$$\mathbf{\Sigma}_\ell = \mathbf{F}_\ell \mathbf{\Sigma}_{\ell-1} \mathbf{F}_\ell^T + \mathbf{Q}_\ell, \quad (6.7)$$

where $\mathbf{x}_\ell$ is the output vector at layer ℓ , $\mathbf{W}_\ell$ and $\mathbf{b}_\ell$ the weights and biases of layer ℓ , $\mathbf{F}_\ell = \nabla_{\mathbf{x}_{\ell-1}} \mathbf{x}_\ell = \partial \mathbf{x}_\ell(i) / \partial \mathbf{x}_{\ell-1}(j)$ is the partial derivative of between layer outputs (Jacobian) of node i and j , and $\mathbf{Q}_\ell$ is the sample covariance of a population of test data. The function $f(\cdot)$ represents the nonlinearity in each of the layers due to the activation function. In all instances here this will be a rectified linear unit (ReLU) which has the form

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (6.8)$$

This activation function is beneficial because of its ease of use and direct linearization about each nonlinear output layer allowing for an easier application of EKF.

When taken into context with a neural network, $\mathbf{x}_\ell$ is simply the state estimate of a single hidden layer and $\mathbf{\Sigma}_\ell$ is the covariance of each state. In this implementation of the EKF uncertainty propagation technique, the per layer state estimates are altered from $f(\mathbf{W}_\ell \cdot \mathbf{x}_{\ell-1} + \mathbf{b}_\ell) \rightarrow f(\mathbf{W}_\ell * \mathbf{x}_{\ell-1} + \mathbf{b}_\ell)$ because of the need for a convolutional network where $*$ represents a 2d spatial convolution. This means that I am now attempting to estimate the noise associated with the learned filters rather than an ensemble of weights, which differs only in concept.

The Jacobian of the between layer states is then

$$\begin{aligned}
\mathbf{F}_\ell &= \frac{\partial \mathbf{x}_\ell(i)}{\partial \mathbf{x}_{\ell-1}(j)} \\
&= \frac{\partial f(\sum_m W_\ell * x_{\ell-1} + b_{\ell-1})}{\partial \mathbf{x}_{\ell-1}(j)} \\
&= \frac{\partial f(\sum_m W_\ell * x_{\ell-1} + b_{\ell-1})}{\partial (\sum_m W_\ell * x_{\ell-1} + b_{\ell-1})} \cdot \frac{\partial \sum_m W_\ell * x_{\ell-1} + b_{\ell-1}}{\partial \mathbf{x}_{\ell-1}(j)},
\end{aligned} \tag{6.9}$$

where I can exchange $W_\ell * x_{\ell-1}$ with $x_{\ell-1} * W_\ell$ thanks to the commutative property of the discrete convolution and also work directly within the convolution using $(f * g)(i, j) = \sum_{m,n} f[i, j] \cdot g[i - m, j - n]$. I would like to remind the reader that x and w are 3 and 4 dimension tensors respectively in this case, where the first 2 dimensions of both tensors are where the spatial convolutions are being performed.² Because of the ReLU activation function $f(\cdot)$ the first term on the third line of Equation 6.9 cancels with the denominator and I arrive at

$$\begin{aligned}
\mathbf{F}_\ell(k_x, k_y, i, j) &= \frac{\partial (\sum_m W_\ell * x_{\ell-1} + b_{\ell-1})}{\partial \mathbf{x}_{\ell-1}(j)} \\
&= \frac{\partial (\sum_m [\sum_{n_x} \sum_{n_y} x_{\ell-1}(k_x, k_y, m) \cdot W_\ell(k_x - n_x, k_y - n_y, m, j)] + b_{\ell-1})}{\partial \mathbf{x}_{\ell-1}(k_x, k_y, j)} \\
&= \sum_{n_x} \sum_{n_y} W_\ell(k_x - n_x, k_y - n_y, i, j) \\
&= \begin{cases} W_\ell * J, & \text{if } \tilde{x}_\ell(i) \geq 1/2 \\ 0, & \text{otherwise,} \end{cases}
\end{aligned} \tag{6.10}$$

where J is a matrix-of-ones with the same dimensions as $x_{\ell-1}$. The conditional $\tilde{x}_\ell$ is the characterization of the ‘negativity’ in the spatial layer estimates x_ℓ and is defined as

$$\tilde{x}_\ell(i) = \begin{cases} \sum_{a,b} \frac{1}{N_{k_x} + N_{k_y}}, & \text{if } x_\ell(a, b, i) > 0 \\ 0, & \text{otherwise,} \end{cases} \tag{6.11}$$

where N_{k_x}, N_{k_y} are the spatial dimensions in pixels for the prediction from layer ℓ . Approximating the spatial response of the output x_ℓ by using the average number of positive activations is a substitution to performing the full tensor product, and care should be

²The padding necessary for the output convolution size to remain the same is assumed.

taken by inspecting the activation maps prior to using this approach. The $W_\ell * J$ convolution results in simply being a smoothing kernel that gets normalized according to the standard deviation, σ_{W_ℓ} , of the weights layer prior to convolving. For every Jacobian estimate $\mathbf{F}_\ell(k_x, k_y, i, j)$, the spatial weight dimensions k_x, k_y are averaged over for propagation through to each Σ_ℓ covariance estimate. The state estimate x_ℓ and covariance estimate Σ_ℓ can thus be performed successively for every layer up to the final network prediction.

6.5 EKF-FCN Future Work

In this chapter I introduced a novel method for propagating the uncertainty of a prediction from a convolutional neural network. I presented the mathematical basis for this approach which stems from the Extended Kalman Filtering formalism and the work presented in Titensky, Jananthan, and Kepner, 2018. Future work that should be attempted with the EKF-FCN technique is the empirical comparison between a distribution of predictions with a fully propagated EKF prediction. Ideally, the EKF variance prediction should align with the variance of the distribution. Going beyond this test, the scenarios for predicting within could be expanded to include instrument thermal noise and Galactic and extragalactic foregrounds. The impact of foregrounds on parameter estimation results using neural networks would be particularly interesting, especially in the context of determining optimal foreground mitigation required for an estimate. Beyond 21cm EoR meta-parameter estimation, this technique is extremely general and should be applicable in almost any regression or classification problem where convolutional neural networks are used. The 21cm EoR cosmological parameter estimation work from this chapter can be found at <https://github.com/jkerrigan/CosmoFCN>.

Acknowledgments

I would like to thank Paul La Plante for providing the 21cm reionization simulations used in this chapter, access to computing resources, and help with this research.

Chapter 7

Conclusion

7.1 The Future of 21cm EoR Cosmology

Generally the techniques and analysis strategies I developed in this thesis, to include several novel methods for an improved independent analysis pipeline, should be directly applicable to any drift scan wide-field interferometer, and with minor adaptations appropriate for steerable phased arrays (e.g. MWA) too. This analysis pipeline worked from raw interferometer visibilities and applied a pioneering fully convolutional neural network for identifying RFI in this data (Chapter 4). As arrays begin to grow in size, the traditional algorithms for RFI identification will become a growing bottleneck. RFI flagging speed and precision offered by neural networks will make processing of this data near real time, quite tractable. Following RFI flagging in pipelines such as HERA is traditionally redundant calibration. While redundant calibration has been shown to be quite successful it can still be prone to biases. On top of redundant calibration, there is also the problem of the following absolute calibration which typically uses only a small number of calibrators on the sky and no primary beam model. I introduced a sky calibration using the FHD simulation software for calibrating 18 nights of HERA-52 data (Chapter 5). In doing this, the long term stability of the HERA array could be understood so as to quantifiably determine its trustworthiness. Following calibration of the instrument the second largest concern is usually the bright galactic and extragalactic foregrounds. I approached this by investigating foreground subtraction inline

with foreground filtering in the context of the PAPER-64 analysis pipeline which further reduced contamination into the EoR window (Chapter 3). With an improved primary beam model for HERA the potential for reducing the effects of beam chromaticity into larger cosmological k -modes for HERA EoR studies is very possible. Finally I demonstrated what I believe will be a major aspect of future 21cm parameter estimation techniques by using a convolutional neural network for regressing on 21cm cosmological cubes (Chapter 6). More importantly I showed that I may be able to determine an uncertainty estimate on these parameters from the combined neural net model and input cube noise. Parameter estimation in this fashion has the potential for directly being fed from the output of this analysis pipeline accounting for interferometer systematics and potentially only mild foreground mitigation.

Appendix A

Foreground Avoidance Signal Loss

A.1 Wide-band Deconvolution Filter Signal Loss

The iterative wide-band deconvolution filter used in §3.1 has the unfortunate side effect of destroying information at k -modes within the “EoR Window” nearest the horizon limit. The original analysis done in Ali et al. (2015) approached signal loss using a Blackman-Harris window function and a residual threshold (ratio of final to initial delay-space component) of 10^{-9} . I extend the original analysis by quantifying the signal loss at all k -modes accessible by the PAPER-64 instrument for both the Blackman-Harris and Top-hat windows using several values of the residual tolerance. This is by no means an exhaustive analysis, but covers the range of filters used in this paper and other similar analyses from the PAPER collaboration. The amount of signal recovered is identified by injecting a known ‘EoR-like’ signal (V_{eor}) to a simulated PAPER-64 baseline visibility containing foregrounds (V_{fgs}). The visibility including the injected EoR signal $V_{\text{fgs+eor}}$ and V_{fgs} are then wide-band filtered. I then create the power spectra P_{fgs} and $P_{\text{fgs+eor}}$ and I compare the recovered signal in the form of

$$\bar{Y}(k)_{\text{Frac. of Signal Rec.}} = \left\langle \frac{\text{Filtered}(P_{\text{eor+data}}) - \text{Filtered}(P_{\text{data}})}{P_{\text{eor}}} \right\rangle, \quad (\text{A.1})$$

where $\langle \cdot \rangle$ represents the ensemble average and $\bar{Y}(k)_{\text{Frac. of Signal Rec.}}$ is the ensemble average over the fraction of signal that is recovered from the wide-band filtering process. The extent

of the typical (average) or extreme limit (3σ) for signal loss can be seen by looking at the ensemble average of recovered signal to injected signal over $N = 1000$ trials where I vary both injected signal realization and power ($10^{-10} \cdot P_{\text{data}} \leq P_{\text{eor}} \leq 10^{-4} \cdot P_{\text{data}}$). The fraction of signal recovered when using a Blackman-Harris window and Top-hat window is shown in Figure A.1(a) and Figure A.1(b). The Top-hat window outperforms the Blackman-Harris with respect to recovering signal in the mode nearest the horizon filter limit ($k \approx 0.079 \text{ hMpc}^{-1}$) because I sacrifice dynamic range for reducing correlation between adjacent k -modes. While both signal loss analyses in Figure A.1(a) and Figure A.1(b) demonstrate fairly similar worst case ($\bar{Y} + 3\sigma$) signal loss across the tolerances demonstrated, its important to note that the magnitude of foregrounds removed from within the filter limit and the number of iterations to convergence (time) varies between tolerance. This leaves the optimal choice of 10^{-9} for the iterative wide-band deconvolution filter.

I also investigate how the increasing width of the filter with increasing baseline length affects signal loss. The results of applying the same signal loss analysis but now to baselines of lengths 30 m, 120 m, and 300 m are shown in Figure A.1(c). I see that moving the filter limit gives a very similar result to the standard 30 m filter limit, simply pushed further in k : modes inside the horizon limit have very large values of signal loss, while modes outside the horizon are essentially unaffected. This analysis is useful in understanding the MWA 2D power spectra which contain different length baselines, seen in Figure 3.9 and Figure 3.10.

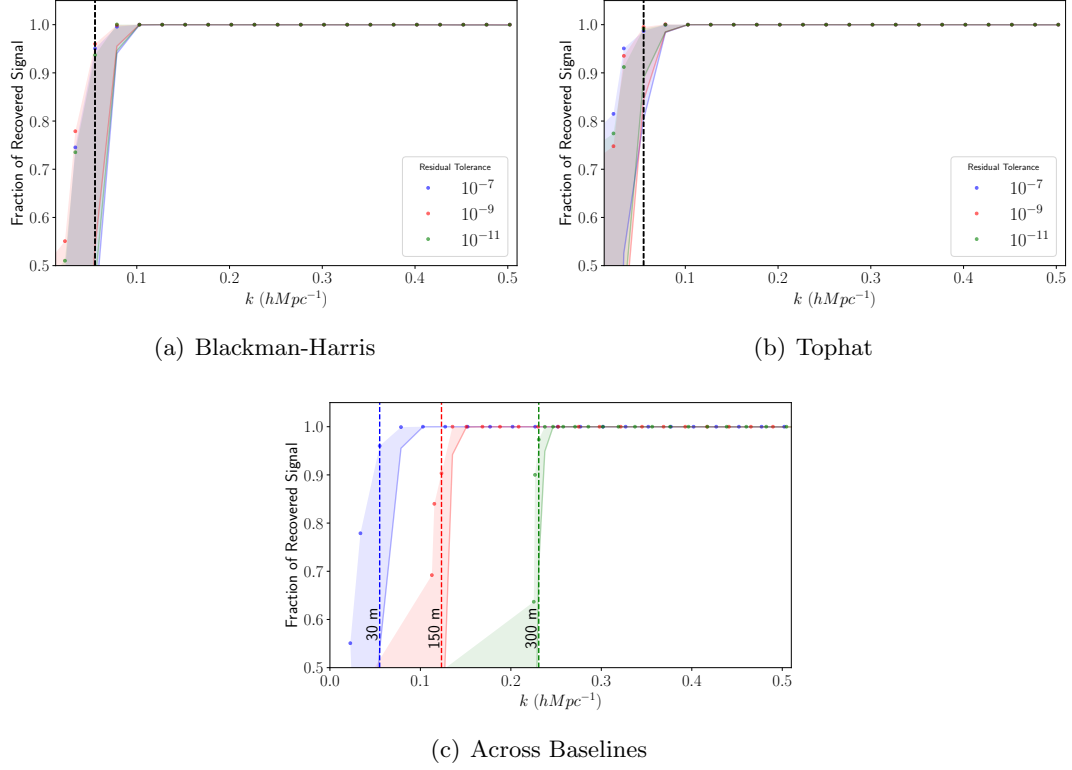


FIGURE A.1: (a) The narrow-band power spectrum signal recovered after performing the wide-band iterative deconvolution filter using a Blackman-Harris window function. The dots represent the average signal recovered at each k -mode and the shaded region is the $3\text{-}\sigma$ limit for the worst case signal recovery. Using the filtering residual tolerance of 10^{-9} I see that for the k -mode right outside the filter limit at $k = 0.079 \text{ hMpc}^{-1}$ I can expect an average signal loss of $8.2 \times 10^{-2}\%$ with a $3\text{-}\sigma$ upper-limit of 4.6% . (b) I approach the same analysis as (a) however I swap out the Blackman-Harris window with a Top-Hat window. The typical signal loss of the mode nearest the filter limit at $k = 0.079 \text{ hMpc}^{-1}$ is $8.4 \times 10^{-2}\%$ with a $3\text{-}\sigma$ upper-limit of 1.6% . (c) Although the PAPER-64 analysis consists of only one baseline type (30 m) I also want to demonstrate the fraction of signal recovered over different length baselines (wider horizon filter limits) which is important to an MWA foreground subtraction and filtering analysis. The residual tolerance is 10^{-9} with a Blackman-Harris window applied, and the filtering is performed over baselines of the length 30 m, 120 m, and 300 m. I show that the amount of signal recovered should essentially be a translation in k when compared to (a) & (b).

Appendix B

Remedying Model Visibility Issues with Fast Holographic Deconvolution

B.1 Creating Model Visibilities with Fast Holographic Deconvolution

In Chapter 3 & 5, I use FHD (Sullivan et al., 2012) to create model visibilities from a foreground source catalog and instrument model. These model visibilities are then directly subtracted from the visibility data to perform the foreground subtraction step. The FHD software package can perform many different operations, including calibration, image-based deconvolution, and visibility simulation; use of FHD as a visibility simulator has been presented in Barry et al., 2016 and Pober et al., 2016. Here, I describe the key steps in FHD visibility simulation to understand where the numerical artifacts described in Section 3.4.3 appear.

When using a point source catalog as input, FHD uses a Discrete Fourier Transform (DFT) to take a source at a floating point location in (RA, Dec) to a discretized uv plane. Both the size of pixels in this uv plane and the number of pixels (i.e. the maximum extent

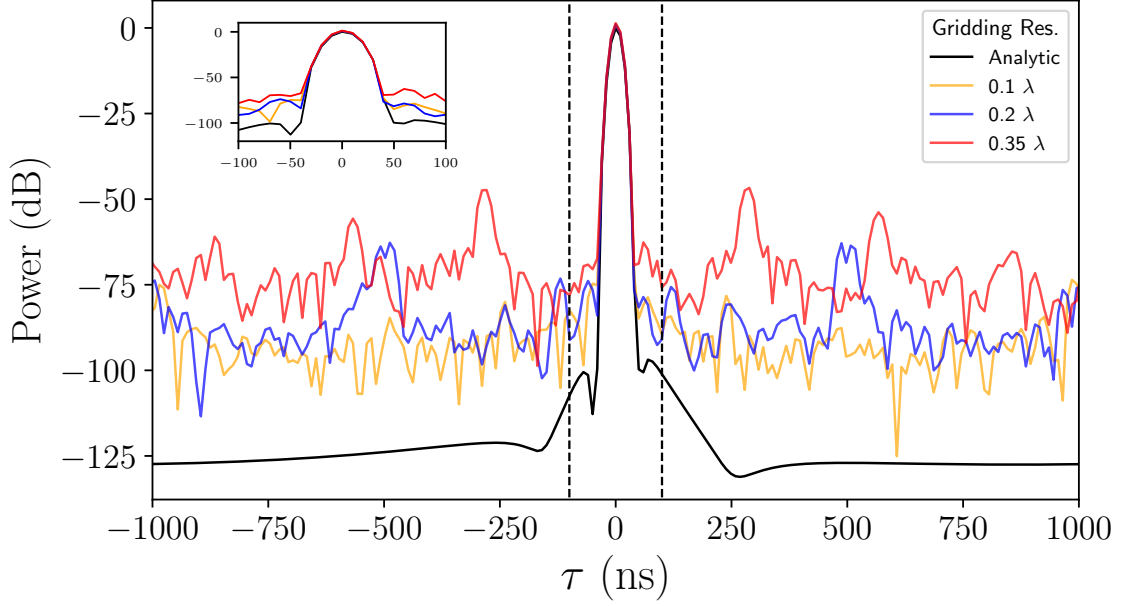


FIGURE B.1: Simulation of a single point source in a 30 m baseline visibility at zenith for different uv -plane resolutions in FHD: 0.1λ (orange), 0.2λ (blue), and 0.35λ (red). I also show the analytic response I expect for the same source (black). The inset figure shows the normalized peak spectra relative to the analytic source. At smaller uv -gridding resolutions, the power at high delays decreases while the peak source brightness converges to the analytic source. These model delay spectra examples also demonstrate a clear aliasing floor and thus lead to a motivation for filtering out high delays prior to subtraction.

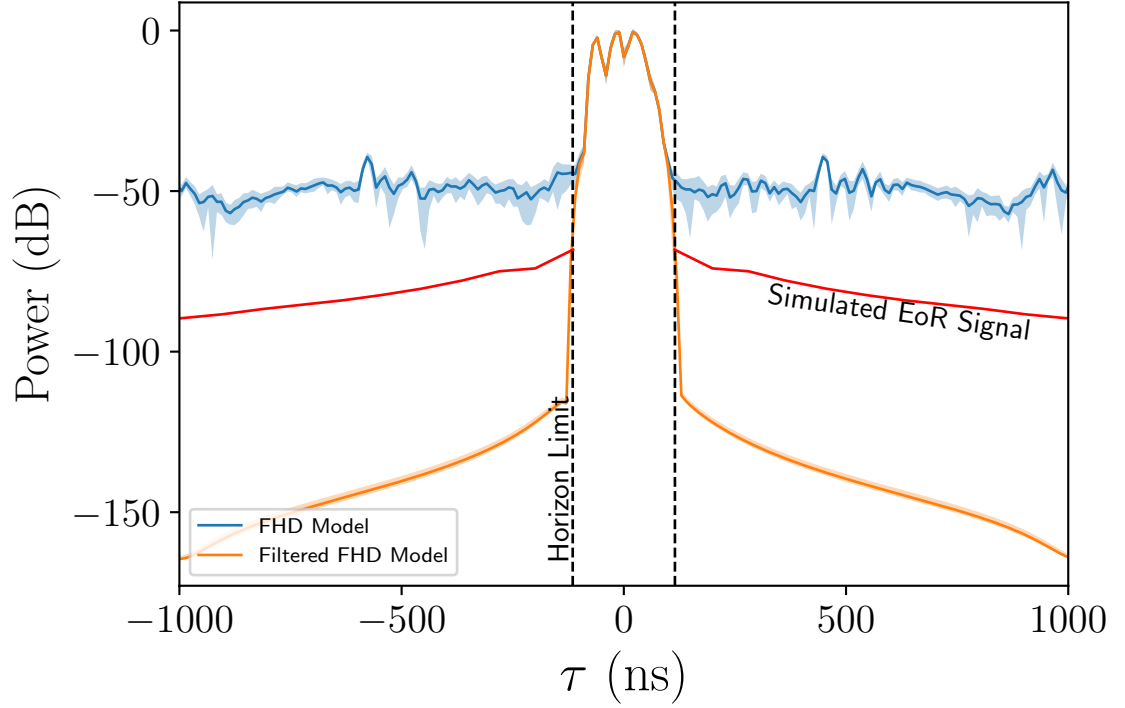


FIGURE B.2: A many sources example (compare to Figure B.1) of a delay transformed and peak power normalized FHD model visibility ($k_{res} = .35\lambda$) for a 30 m baseline averaged over a single observation (blue), and the same model after low pass filtering (orange) both with shaded regions demonstrating the variance over time. A Blackman-Harris window function has been applied in both cases. In the unaltered curve, large amplitude spectral structure can be seen outside of the delay horizon, > 115 ns. As described in Chapter 3, the origin of this high delay power is not physical. A fiducial 21cm Power Spectrum at 50% reionization (Lidz et al., 2008)(red) is shown as reference to demonstrate that the FHD model must be filtered prior to subtraction so that the EoR signal is not contaminated.

of the uv plane) are free parameters in FHD. All uv planes for each point source being simulated are summed to create a model uv sky-brightness distribution. Model visibilities for each baseline are then created by integrating the product of the uv response function (the Fourier transform of the primary beam from image domain to uv space; Morales and Matejek 2009) at the (u, v) coordinate of the baseline of interest and the model uv sky-brightness distribution. Because the uv plane has been discretized into pixels, however, this integral is approximated as a discrete sum and can introduce numerical errors, particularly as a function of frequency (since the (u, v) coordinates of a baseline are frequency dependent).

To illustrate these errors, I show the delay spectrum of a FHD-generated model visibility for a 30m East-West baseline and a single source in Figure B.1. Different colored curves correspond to different pixel areas in the uv plane in units of wavelength: 0.1λ (orange), 0.2λ (blue) and 0.35λ (red). Increasing the uv -plane resolution clearly mitigates the high delay power created by spurious spectral structure, but comes at the cost of increased computational resources and would not be feasible without specialized high RAM machines. The black curve represents the analytic value assuming a Blackman-Harris window was used in the Fourier transform to delay space. The gridding resolution 0.35λ is the resolution used for the analysis in Chapter 3, and was used strictly for the purpose of efficiently calibrating, modeling, and subtracting on a large dataset; running the full analysis at these higher resolutions is not computationally feasible, and even then, it does not bring the numerical error significantly below the level of the EoR signal (which is also approximately 80 dB below the foregrounds). Small flux density scaling errors between the different resolutions are a point of concern, but is largely mitigated in this analysis, as I effectively calibrate the flux scale of the FHD model to the absolutely calibrated PAPER data.

As noted in Section 3.4.3, high-delay power is the limiting factor in this analysis. If it is not removed, it sets a floor to best obtainable limits on the EoR signal with PAPER-64 data. To remedy this problem I use the iterative deconvolution filter to low pass filter the 0.35λ model visibilities with filter horizon of 90 ns for the 30 m baselines. Figure B.2 illustrates the effect of this filter. The delay spectrum of the FHD model prior to filtering is shown in blue and the delay spectrum after filtering is shown in orange. For comparison,

a model EoR signal is shown in red. I acknowledge that this filter will also remove any *real* high delay power (introduced by the instrument)¹ from the model visibilities. Because the high delay power in the model is a limiting systematic, however, I choose to forgo the possibility of using the model to subtract high delay power from the data. The fact that I still see improvements in the power spectrum (Section 3.6.1) shows that I am reducing the overall power within the horizon to the point where the filtering process is not dynamic range limited. Going forward, using either a modified version of FHD or another (likely more computationally intensive) code for simulating visibilities, I would not have to low pass filter the models and any improvement of the filtering+subtraction technique near the horizon should come from both effects: removing the chromatic footprint of the PAPER antenna response and reducing the dynamic range requirements on the delay filter.

¹Because I model the sources on the sky as flat spectrum, they will not be a source of high delay power.

Bibliography

- Abadi, Martín et al. (2016). “TensorFlow: A System for Large-scale Machine Learning”. In: *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*. OSDI’16. Savannah, GA, USA: USENIX Association, pp. 265–283. ISBN: 978-1-931971-33-1. URL: <http://dl.acm.org/citation.cfm?id=3026877.3026899>.
- Akeret, J. et al. (2017). “Radio frequency interference mitigation using deep convolutional neural networks”. In: *Astronomy and Computing* 18, pp. 35–39. DOI: 10.1016/j.ascom.2017.01.002.
- Ali, Zaki S. et al. (2015). “PAPER-64 Constraints on Reionization: The 21 cm Power Spectrum at $z = 8.4$ ”. In: 809.1, p. 61. URL: <http://stacks.iop.org/0004-637X/809/i=1/a=61>.
- Barkana, R. and A. Loeb (2001). “In the beginning: the first sources of light and the reionization of the universe”. In: 349.2, pp. 125–238. DOI: 10.1016/S0370-1573(01)00019-9. arXiv: astro-ph/0010468 [astro-ph].
- (2004). “Unusually Large Fluctuations in the Statistics of Galaxy Formation at High Redshift”. In: 609, pp. 474–481. DOI: 10.1086/421079. eprint: astro-ph/0310338.
- Barry, N. et al. (2016). “Calibration requirements for detecting the 21 cm epoch of reionization power spectrum and implications for the SKA”. In: 461, pp. 3135–3144. DOI: 10.1093/mnras/stw1380. arXiv: 1603.00607 [astro-ph.IM].
- Battaglia, N. et al. (2013). “Reionization on Large Scales. I. A Parametric Model Constructed from Radiation-hydrodynamic Simulations”. In: 776, 81, p. 81. DOI: 10.1088/0004-637X/776/2/81. arXiv: 1211.2821 [astro-ph.CO].

- Beardsley, A. P. et al. (2016). “First Season MWA EoR Power spectrum Results at Redshift 7”. In: 833, 102, p. 102. DOI: 10.3847/1538-4357/833/1/102. arXiv: 1608.06281 [astro-ph.IM].
- Becker, R. H. et al. (2001). “Evidence for Reionization at $z \sim 6$: Detection of a Gunn-Peterson Trough in a $z=6.28$ Quasar”. In: 122, pp. 2850–2857. DOI: 10.1086/324231. eprint: astro-ph/0108097.
- Bengio, Yoshua, Aaron Courville, and Pascal Vincent (2012). *Representation Learning: A Review and New Perspectives*.
- Bennett, C. L. et al. (2013). “Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results”. In: 208, 20, p. 20. DOI: 10.1088/0067-0049/208/2/20. arXiv: 1212.5225.
- Bernardi, G. et al. (2011). “Subtraction of point sources from interferometric radio images through an algebraic forward modelling scheme”. In: 413, pp. 411–422. DOI: 10.1111/j.1365-2966.2010.18145.x. arXiv: 1012.3719.
- Bhatnagar, S. et al. (2008). “Correcting direction-dependent gains in the deconvolution of radio interferometric images”. In: 487, pp. 419–429. DOI: 10.1051/0004-6361:20079284. arXiv: 0805.0834.
- Born, M. and E. Wolf (1999). *Principles of optics : electromagnetic theory of propagation, interference and diffraction of light*.
- Bosman, S. E. I. et al. (2018). “New constraints on Lyman- α opacity with a sample of 62 quasars at $z \sim 5.7$ ”. In: 479, pp. 1055–1076. DOI: 10.1093/mnras/sty1344. arXiv: 1802.08177.
- Bowman, J. D., A. E. E. Rogers, and J. N. Hewitt (2008). “Toward Empirical Constraints on the Global Redshifted 21 cm Brightness Temperature During the Epoch of Reionization”. In: 676, pp. 1–9. DOI: 10.1086/528675. arXiv: 0710.2541.
- Burd, P. R. et al. (2018). “Detecting Radio Frequency Interference in radio-antenna arrays with the Recurrent Neural Network algorithm”. In: *ArXiv e-prints*. arXiv: 1808.09739 [astro-ph.IM].

- Byrne, Ruby et al. (2019). “Fundamental Limitations on the Calibration of Redundant 21 cm Cosmology Instruments and Implications for HERA and the SKA”. In: 875.1, 70, p. 70. DOI: 10.3847/1538-4357/ab107d. arXiv: 1811.01378 [astro-ph.IM].
- Cardoso, J.-F. et al. (2008). “Component Separation With Flexible Models Application to Multichannel Astrophysical Observations”. In: *IEEE Journal of Selected Topics in Signal Processing* 2, pp. 735–746. DOI: 10.1109/JSTSP.2008.2005346.
- Carozzi, T. D. and G. Woan (2009). “A generalized measurement equation and van Cittert-Zernike theorem for wide-field radio astronomical interferometry”. In: 395, pp. 1558–1568. DOI: 10.1111/j.1365-2966.2009.14642.x. arXiv: 0812.0141.
- Chapman, E. et al. (2013). “The scale of the problem: recovering images of reionization with Generalized Morphological Component Analysis”. In: 429, pp. 165–176. DOI: 10.1093/mnras/sts333. arXiv: 1209.4769 [astro-ph.CO].
- Cheng, C. et al. (2018). “Characterizing Signal Loss in the 21 cm Reionization Power Spectrum: A Revised Study of PAPER-64”. In: 868, 26, p. 26. DOI: 10.3847/1538-4357/aae833. arXiv: 1810.05175 [astro-ph.IM].
- Clark, B. G. (1980). “An efficient implementation of the algorithm ‘CLEAN’”. In: 89, p. 377.
- Cohen, A. S. et al. (2007). “The VLA Low-Frequency Sky Survey”. In: 134, pp. 1245–1262. DOI: 10.1086/520719. arXiv: 0706.1191.
- Condon, J. and S. Ransom (2016). *Essential Radio Astronomy*. <https://www.cv.nrao.edu/course/ast534/Interferometers1.html>. (Accessed on 05/08/2019).
- Datta, Abhirup et al. (2014). “Effects Of The Ionosphere On Ground-Based Detection Of The Global 21 CM Signal From The Cosmic Dawn And The Dark Ages”. In: *arXiv e-prints*, arXiv:1409.0513, arXiv:1409.0513. arXiv: 1409.0513 [astro-ph.CO].
- Datta, Kanan K. et al. (2012). “Light-cone effect on the reionization 21-cm power spectrum”. In: 424.3, pp. 1877–1891. DOI: 10.1111/j.1365-2966.2012.21293.x. arXiv: 1109.1284 [astro-ph.CO].
- Davies, F. B., G. D. Becker, and S. R. Furlanetto (2018). “Determining the Nature of Late Gunn-Peterson Troughs with Galaxy Surveys”. In: 860, 155, p. 155. DOI: 10.3847/1538-4357/aac2d6. arXiv: 1708.08927.

- de Oliveira-Costa, A. et al. (2008). “A model of diffuse Galactic radio emission from 10 MHz to 100 GHz”. In: 388, pp. 247–260. DOI: 10.1111/j.1365-2966.2008.13376.x. arXiv: 0802.1525.
- DeBoer, David R. et al. (2017). “Hydrogen Epoch of Reionization Array (HERA)”. In: 129.974, p. 045001. DOI: 10.1088/1538-3873/129/974/045001. arXiv: 1606.07473 [astro-ph.IM].
- Delabrouille, J. et al. (2009). “A full sky, low foreground, high resolution CMB map from WMAP”. In: 493, pp. 835–857. DOI: 10.1051/0004-6361:200810514. arXiv: 0807.0773.
- Dillon, J. S., A. Liu, and M. Tegmark (2013). “A fast method for power spectrum and foreground analysis for 21 cm cosmology”. In: 87.4, 043005, p. 043005. DOI: 10.1103/PhysRevD.87.043005. arXiv: 1211.2232 [astro-ph.CO].
- Dillon, J. S. et al. (2015). “Empirical covariance modeling for 21 cm power spectrum estimation: A method demonstration and new limits from early Murchison Widefield Array 128-tile data”. In: 91.12, 123011, p. 123011. DOI: 10.1103/PhysRevD.91.123011. arXiv: 1506.01026.
- Dillon, J. S. et al. (2018). “Polarized redundant-baseline calibration for 21 cm cosmology without adding spectral structure”. In: 477, pp. 5670–5681. DOI: 10.1093/mnras/sty1060. arXiv: 1712.07212 [astro-ph.IM].
- Dodelson, S. and G. Efstathiou (2004). “Modern Cosmology”. In: *Physics Today* 57, pp. 60–61. DOI: 10.1063/1.1784308.
- Eriksen, H. K. et al. (2008). “Joint Bayesian Component Separation and CMB Power Spectrum Estimation”. In: 676, pp. 10–32. DOI: 10.1086/525277. arXiv: 0709.1058.
- Fan, X. et al. (2006). “A Survey of z5.7 Quasars in the Sloan Digital Sky Survey. IV. Discovery of Seven Additional Quasars”. In: 131, pp. 1203–1209. DOI: 10.1086/500296. eprint: astro-ph/0512080.
- Fernández-Cobos, R. et al. (2012). “Multiresolution internal template cleaning: an application to the Wilkinson Microwave Anisotropy Probe 7-yr polarization data”. In: 420,

- pp. 2162–2169. DOI: 10.1111/j.1365-2966.2011.20182.x. arXiv: 1106.2016 [astro-ph.CO].
- Field, G. B. (1959). “The Spin Temperature of Intergalactic Neutral Hydrogen.” In: 129, p. 536. DOI: 10.1086/146653.
- Furlanetto, Steven R., S. Peng Oh, and Frank H. Briggs (2006). “Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe”. In: 433.4-6, pp. 181–301. DOI: 10.1016/j.physrep.2006.08.002. arXiv: astro-ph/0608032 [astro-ph].
- Geil, P. M., B. M. Gaensler, and J. S. B. Wyithe (2011). “Polarized foreground removal at low radio frequencies using rotation measure synthesis: uncovering the signature of hydrogen reionization”. In: 418, pp. 516–535. DOI: 10.1111/j.1365-2966.2011.19509.x. arXiv: 1011.2321.
- Górski, K. M. et al. (2005). “HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere”. In: 622, pp. 759–771. DOI: 10.1086/427976. eprint: astro-ph/0409513.
- Griffiths, L. M., D. Barbosa, and A. R. Liddle (1999). “Cosmic microwave background constraints on the epoch of reionization”. In: 308, pp. 854–862. DOI: 10.1046/j.1365-8711.1999.02777.x. eprint: astro-ph/9812125.
- Gunn, J. E. and B. A. Peterson (1965). “On the Density of Neutral Hydrogen in Intergalactic Space.” In: 142, pp. 1633–1641. DOI: 10.1086/148444.
- He, Kaiming et al. (2016). “Deep Residual Learning for Image Recognition”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pp. 770–778. DOI: 10.1109/CVPR.2016.90. URL: <https://doi.org/10.1109/CVPR.2016.90>.
- Högbom, J. A. (1974). “Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines”. In: 15, p. 417.
- Hurley-Walker, N. et al. (2017a). “GaLactic and Extragalactic All-sky Murchison Widefield Array (GLEAM) survey - I. A low-frequency extragalactic catalogue”. In: 464, pp. 1146–1167. arXiv: 1610.08318.

- Hurley-Walker, N. et al. (2017b). “GaLactic and Extragalactic All-sky Murchison Widefield Array (GLEAM) survey - I. A low-frequency extragalactic catalogue”. In: 464, pp. 1146–1167. DOI: 10.1093/mnras/stw2337. arXiv: 1610.08318.
- Jacobs, D. C. et al. (2013). “A Flux Scale for Southern Hemisphere 21 cm Epoch of Reionization Experiments”. In: 776, 108, p. 108. DOI: 10.1088/0004-637X/776/2/108. arXiv: 1307.7716.
- Jacobs, D. C. et al. (2015). “Multiredshift Limits on the 21 cm Power Spectrum from PAPER”. In: 801, 51, p. 51. DOI: 10.1088/0004-637X/801/1/51. arXiv: 1408.3389.
- Jacobs, D. C. et al. (2016). “The Murchison Widefield Array 21 cm Power Spectrum Analysis Methodology”. In: 825, 114, p. 114. DOI: 10.3847/0004-637X/825/2/114. arXiv: 1605.06978 [astro-ph.IM].
- Kern N., Gianni B. and C. Carilli (2018). “ABSOLUTE CALIBRATION OF HERA-47 IDR1”. In: *Memo Series 42, HERA Collaboration*.
- Kerrigan, J. R. et al. (2018). “Improved 21 cm Epoch of Reionization Power Spectrum Measurements with a Hybrid Foreground Subtraction and Avoidance Technique”. In: 864, 131, p. 131. DOI: 10.3847/1538-4357/aad8bb. arXiv: 1801.00460 [astro-ph.IM].
- Kingma, D. P. and J. Ba (2014). “Adam: A Method for Stochastic Optimization”. In: *ArXiv e-prints*. arXiv: 1412.6980.
- Kohn, S. et al. (2018). “Polarized Foreground Power Spectra from the HERA-19 Commissioning Array”. In: *arXiv e-prints*. arXiv: 1802.04151 [astro-ph.IM].
- Kohn, S. A. (2016). “HERA-19 commissioning: radio frequency interference”. In: *Memo Series 19, HERA Collaboration*.
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton (2017). “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Commun. ACM* 60.6, pp. 84–90. ISSN: 0001-0782. DOI: 10.1145/3065386. URL: <http://doi.acm.org/10.1145/3065386>.
- La Plante, P. et al. (2014). “Reionization on Large Scales. IV. Predictions for the 21 cm Signal Incorporating the Light Cone Effect”. In: 789, 31, p. 31. DOI: 10.1088/0004-637X/789/1/31. arXiv: 1309.7056.

- La Plante, Paul and Michelle Ntampaka (2018). “Machine Learning Applied to the Reionization History of the Universe”. In: *arXiv e-prints*, arXiv:1810.08211, arXiv:1810.08211. arXiv: 1810.08211 [astro-ph.CO].
- Lane, W. M. et al. (2014). “VizieR Online Data Catalog: 74MHz VLA Low-frequency Sky Survey Redux (VLSSr) (Lane+, 2014)”. In: *VizieR Online Data Catalog* 8097.
- LeCun, Yann and Yoshua Bengio (1998). “The Handbook of Brain Theory and Neural Networks”. In: ed. by Michael A. Arbib. Cambridge, MA, USA: MIT Press. Chap. Convolutional Networks for Images, Speech, and Time Series, pp. 255–258. ISBN: 0-262-51102-9. URL: <http://dl.acm.org/citation.cfm?id=303568.303704>.
- Lecun, Yann et al. (1998). “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE*, pp. 2278–2324.
- Li, W. et al. (2018). “Comparing Redundant and Sky-model-based Interferometric Calibration: A First Look with Phase II of the MWA”. In: 863.2, 170, p. 170. DOI: 10.3847/1538-4357/aad3c3. arXiv: 1807.05312 [astro-ph.IM].
- Lidz, A. et al. (2008). “Detecting the Rise and Fall of 21 cm Fluctuations with the Murchison Widefield Array”. In: 680, 962-974, pp. 962–974. DOI: 10.1086/587618. arXiv: 0711.4373.
- Lin, M., Q. Chen, and S. Yan (2013). “Network In Network”. In: *ArXiv e-prints*. arXiv: 1312.4400.
- Liu, A., A. R. Parsons, and C. M. Trott (2014a). “Epoch of reionization window. I. Mathematical formalism”. In: 90.2, 023018, p. 023018. DOI: 10.1103/PhysRevD.90.023018. arXiv: 1404.2596.
- (2014b). “Epoch of reionization window. II. Statistical methods for foreground wedge reduction”. In: 90.2, 023019, p. 023019. DOI: 10.1103/PhysRevD.90.023019. arXiv: 1404.4372.
- (2014c). “Epoch of reionization window. II. Statistical methods for foreground wedge reduction”. In: 90.2, 023019, p. 023019. DOI: 10.1103/PhysRevD.90.023019. arXiv: 1404.4372.

- Liu, A. and M. Tegmark (2011). “A method for 21 cm power spectrum estimation in the presence of foregrounds”. In: 83.10, 103006, p. 103006. DOI: 10.1103/PhysRevD.83.103006. arXiv: 1103.0281 [astro-ph.CO].
- Liu, A. et al. (2013). “Global 21 cm signal experiments: A designer’s guide”. In: 87.4, 043002, p. 043002. DOI: 10.1103/PhysRevD.87.043002. arXiv: 1211.3743 [astro-ph.CO].
- Liu, A. et al. (2016). “Eliminating the optical depth nuisance from the CMB with 21 cm cosmology”. In: 93.4, 043013, p. 043013. DOI: 10.1103/PhysRevD.93.043013. arXiv: 1509.08463.
- Liu, Adrian et al. (2010). “Precision calibration of radio interferometers using redundant baselines”. In: 408.2, pp. 1029–1050. DOI: 10.1111/j.1365-2966.2010.17174.x. arXiv: 1001.5268 [astro-ph.IM].
- Long, J., E. Shelhamer, and T. Darrell (2014). “Fully Convolutional Networks for Semantic Segmentation”. In: *ArXiv e-prints*. arXiv: 1411.4038 [cs.CV].
- Maas, Andrew L., Awni Y. Hannun, and Andrew Y. Ng (2013). “Rectifier nonlinearities improve neural network acoustic models”. In: *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*.
- Majumdar, Suman et al. (2018). “Quantifying the non-Gaussianity in the EoR 21-cm signal through bispectrum”. In: 476.3, pp. 4007–4024. DOI: 10.1093/mnras/sty535. arXiv: 1708.08458 [astro-ph.CO].
- Mao, Y. et al. (2012). “Redshift-space distortion of the 21-cm background from the epoch of reionization - I. Methodology re-examined”. In: 422, pp. 926–954. DOI: 10.1111/j.1365-2966.2012.20471.x. arXiv: 1104.2094.
- Mesinger, Andrei, Steven Furlanetto, and Renyue Cen (2011). “21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal”. In: 411.2, pp. 955–972. DOI: 10.1111/j.1365-2966.2010.17731.x. arXiv: 1003.3878 [astro-ph.CO].
- Monsalve, R. A. et al. (2017). “Results from EDGES High-band. I. Constraints on Phenomenological Models for the Global 21 cm Signal”. In: 847, 64, p. 64. DOI: 10.3847/1538-4357/aa88d1. arXiv: 1708.05817.

- Morales, M. F. and M. Matejek (2009). “Software holography: interferometric data analysis for the challenges of next generation observatories”. In: 400, pp. 1814–1820. DOI: 10.1111/j.1365-2966.2009.15537.x. arXiv: 0810.5107.
- Morales, M. F. et al. (2012). “Four Fundamental Foreground Power Spectrum Shapes for 21 cm Cosmology Observations”. In: 752, 137, p. 137. DOI: 10.1088/0004-637X/752/2/137. arXiv: 1202.3830 [astro-ph.IM].
- Morales, Miguel F. and Jacqueline Hewitt (2004). “Toward Epoch of Reionization Measurements with Wide-Field Radio Observations”. In: *The Astrophysical Journal* 615.1, pp. 7–18. DOI: 10.1086/424437. URL: <https://doi.org/10.1086/424437>.
- Morales, Miguel F. et al. (2019). “Understanding the diversity of 21 cm cosmology analyses”. In: 483.2, pp. 2207–2216. DOI: 10.1093/mnras/sty2844. arXiv: 1810.08731 [astro-ph.CO].
- Offringa, A. R., F. Mertens, and L. V. E. Koopmans (2019). “The impact of interference excision on 21-cm epoch of reionization power spectrum analyses”. In: 484, pp. 2866–2875. DOI: 10.1093/mnras/stz175.
- Offringa, A. R., J. J. van de Gronde, and J. B. T. M. Roerdink (2012). “A morphological algorithm for improving radio-frequency interference detection”. In: 539, A95, A95. DOI: 10.1051/0004-6361/201118497. arXiv: 1201.3364 [astro-ph.IM].
- Offringa, A. R. et al. (2013). “The LOFAR radio environment”. In: 549, A11, A11. DOI: 10.1051/0004-6361/201220293. arXiv: 1210.0393 [astro-ph.IM].
- Offringa, A. R. et al. (2015). “The Low-Frequency Environment of the Murchison Widefield Array: Radio-Frequency Interference Analysis and Mitigation”. In: 32, e008, e008. DOI: 10.1017/pasa.2015.7. arXiv: 1501.03946 [astro-ph.IM].
- Paciga, G. et al. (2013). “A simulation-calibrated limit on the H I power spectrum from the GMRT Epoch of Reionization experiment”. In: 433, pp. 639–647. DOI: 10.1093/mnras/stt753. arXiv: 1301.5906.
- Parsons, A. (2015). “Power Spectrum Normalizations for HERA”. In: *Memo Series 27, HERA Collaboration*.

- Parsons, A. and A. Beardsley (2017). “Modeling Sky Temperature”. In: *Memo Series 34, HERA Collaboration*.
- Parsons, A. R. et al. (2012a). “A Per-baseline, Delay-spectrum Technique for Accessing the 21 cm Cosmic Reionization Signature”. In: 756, 165, p. 165. DOI: 10.1088/0004-637X/756/2/165. arXiv: 1204.4749 [astro-ph.IM].
- Parsons, A. R. et al. (2014). “New Limits on 21 cm Epoch of Reionization from PAPER-32 Consistent with an X-Ray Heated Intergalactic Medium at $z = 7.7$ ”. In: 788, 106, p. 106. DOI: 10.1088/0004-637X/788/2/106. arXiv: 1304.4991.
- Parsons, Aaron et al. (2012b). “A Sensitivity and Array-configuration Study for Measuring the Power Spectrum of 21 cm Emission from Reionization”. In: 753.1, 81, p. 81. DOI: 10.1088/0004-637X/753/1/81. arXiv: 1103.2135 [astro-ph.IM].
- Parsons, Aaron R. and Donald C. Backer (2009). “Calibration of Low-Frequency, Wide-Field Radio Interferometers Using Delay/Delay-Rate Filtering”. In: 138.1, p. 219. URL: <http://stacks.iop.org/1538-3881/138/i=1/a=219>.
- Parsons, Aaron R. et al. (2010). “The Precision Array for Probing the Epoch of Reionization: Eight Station Results”. In: 139.4, p. 1468. URL: <http://stacks.iop.org/1538-3881/139/i=4/a=1468>.
- Patil, A. H. et al. (2017). “Upper Limits on the 21 cm Epoch of Reionization Power Spectrum from One Night with LOFAR”. In: 838, 65, p. 65. DOI: 10.3847/1538-4357/aa63e7. arXiv: 1702.08679.
- Peacock, John A. (1999). *Cosmological Physics*.
- Planck Collaboration et al. (2016). “Planck 2015 results. I. Overview of products and scientific results”. In: 594, A1, A1. DOI: 10.1051/0004-6361/201527101. arXiv: 1502.01582.
- Planck Collaboration et al. (2018). “Planck 2018 results. I. Overview and the cosmological legacy of Planck”. In: *arXiv e-prints*, arXiv:1807.06205, arXiv:1807.06205. arXiv: 1807.06205 [astro-ph.CO].
- Pober, J. C. et al. (2012). “A Technique for Primary Beam Calibration of Drift-scanning, Wide-field Antenna Elements”. In: 143, 53, p. 53. DOI: 10.1088/0004-6256/143/2/53. arXiv: 1111.2882 [astro-ph.IM].

- Pober, J. C. et al. (2013). “Opening the 21 cm Epoch of Reionization Window: Measurements of Foreground Isolation with PAPER”. In: 768, L36, p. L36. DOI: 10.1088/2041-8205/768/2/L36. arXiv: 1301.7099.
- Pober, J. C. et al. (2016). “The Importance of Wide-field Foreground Removal for 21 cm Cosmology: A Demonstration with Early MWA Epoch of Reionization Observations”. In: 819, 8, p. 8. DOI: 10.3847/0004-637X/819/1/8. arXiv: 1601.06177 [astro-ph.IM].
- Pober, Jonathan C. et al. (2014). “What Next-generation 21 cm Power Spectrum Measurements can Teach us About the Epoch of Reionization”. In: 782.2, 66, p. 66. DOI: 10.1088/0004-637X/782/2/66. arXiv: 1310.7031 [astro-ph.CO].
- Pritchard, J. R. and S. R. Furlanetto (2007). “21-cm fluctuations from inhomogeneous X-ray heating before reionization”. In: 376, pp. 1680–1694. DOI: 10.1111/j.1365-2966.2007.11519.x. eprint: astro-ph/0607234.
- Pritchard, Jonathan R. and Abraham Loeb (2012). “21 cm cosmology in the 21st century”. In: *Reports on Progress in Physics* 75.8, 086901, p. 086901. DOI: 10.1088/0034-4885/75/8/086901. arXiv: 1109.6012 [astro-ph.CO].
- Rauch, Michael (1998). “THE LYMAN ALPHA FOREST IN THE SPECTRA OF QUASIS-TELLAR OBJECTS”. In: *Annual Review of Astronomy and Astrophysics* 36.1, pp. 267–316. DOI: 10.1146/annurev.astro.36.1.267. eprint: <https://doi.org/10.1146/annurev.astro.36.1.267>. URL: <https://doi.org/10.1146/annurev.astro.36.1.267>.
- Ronneberger, O., P. Fischer, and T. Brox (2015). “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *ArXiv e-prints*. arXiv: 1505.04597 [cs.CV].
- Salvini, S. and S. J. Wijnholds (2014). “Fast gain calibration in radio astronomy using alternating direction implicit methods: Analysis and applications”. In: 571, A97, A97. DOI: 10.1051/0004-6361/201424487. arXiv: 1410.2101 [astro-ph.IM].
- Smoot, G. F. et al. (1992). “Structure in the COBE differential microwave radiometer first-year maps”. In: 396, pp. L1–L5. DOI: 10.1086/186504.

- Spergel, D. N. et al. (2007). “Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology”. In: 170, pp. 377–408. DOI: 10.1086/513700. eprint: astro-ph/0603449.
- Sullivan, I. S. et al. (2012). “Fast Holographic Deconvolution: A New Technique for Precision Radio Interferometry”. In: 759, 17, p. 17. DOI: 10.1088/0004-637X/759/1/17. arXiv: 1209.1653 [astro-ph.IM].
- Sutinjo, A. T. et al. (2015). “Characterization of a Low-Frequency Radio Astronomy Prototype Array in Western Australia”. In: *IEEE Transactions on Antennas and Propagation* 63, pp. 5433–5442. DOI: 10.1109/TAP.2015.2487504. arXiv: 1510.01515 [astro-ph.IM].
- Taylor, G. B., C. L. Carilli, and R. A. Perley, eds. (1999). *Synthesis Imaging in Radio Astronomy II*. Vol. 180. Astronomical Society of the Pacific Conference Series.
- Tegmark, Max et al. (1997). “How Small Were the First Cosmological Objects?” In: 474, p. 1. DOI: 10.1086/303434. arXiv: astro-ph/9603007 [astro-ph].
- Thompson, A. R., J. M. Moran, and G. W. Swenson Jr. (2017). *Interferometry and Synthesis in Radio Astronomy, 3rd Edition*. DOI: 10.1007/978-3-319-44431-4.
- Thyagarajan, N. et al. (2015). “Foregrounds in Wide-field Redshifted 21 cm Power Spectra”. In: 804, 14, p. 14. DOI: 10.1088/0004-637X/804/1/14. arXiv: 1502.07596 [astro-ph.IM].
- Thyagarajan, N. et al. (2016a). “Effects of Antenna Beam Chromaticity on Redshifted 21 cm Power Spectrum and Implications for Hydrogen Epoch of Reionization Array”. In: 825, 9, p. 9. DOI: 10.3847/0004-637X/825/1/9. arXiv: 1603.08958.
- (2016b). “Effects of Antenna Beam Chromaticity on Redshifted 21 cm Power Spectrum and Implications for Hydrogen Epoch of Reionization Array”. In: 825, 9, p. 9. DOI: 10.3847/0004-637X/825/1/9. arXiv: 1603.08958.
- Tingay, S. J. et al. (2015). “A Search for Fast Radio Bursts at Low Frequencies with Murchison Widefield Array High Time Resolution Imaging”. In: 150, 199, p. 199. DOI: 10.1088/0004-6256/150/6/199. arXiv: 1511.02985.

- Titensky, Jessica S., Hayden Jananthan, and Jeremy Kepner (2018). “Uncertainty Propagation in Deep Neural Networks Using Extended Kalman Filtering”. In: *arxiv*. URL: <https://arxiv.org/abs/1809.06009>.
- Trott, C. M. et al. (2016). “CHIPS: The Cosmological H i Power Spectrum Estimator”. In: 818, 139, p. 139. DOI: 10.3847/0004-637X/818/2/139. arXiv: 1601.02073 [astro-ph.IM].
- Vedantham, H., N. Udaya Shankar, and R. Subrahmanyan (2012a). “Imaging the Epoch of Reionization: Limitations from Foreground Confusion and Imaging Algorithms”. In: 745, 176, p. 176. DOI: 10.1088/0004-637X/745/2/176. arXiv: 1106.1297 [astro-ph.IM].
- (2012b). “Imaging the Epoch of Reionization: Limitations from Foreground Confusion and Imaging Algorithms”. In: 745, 176, p. 176. DOI: 10.1088/0004-637X/745/2/176. arXiv: 1106.1297 [astro-ph.IM].
- Wang, J. et al. (2013). “Exploring the Cosmic Reionization Epoch in Frequency Space: An Improved Approach to Remove the Foreground in 21 cm Tomography”. In: 763, 90, p. 90. DOI: 10.1088/0004-637X/763/2/90. arXiv: 1211.6450.
- Wouthuysen, S. A. (1952). “On the excitation mechanism of the 21-cm (radio-frequency) interstellar hydrogen emission line.” In: 57, pp. 31–32. DOI: 10.1086/106661.
- Wrobel, J. M. and R. C. Walker (1999). “Sensitivity”. In: *Synthesis Imaging in Radio Astronomy II*. Ed. by G. B. Taylor, C. L. Carilli, and R. A. Perley. Vol. 180. Astronomical Society of the Pacific Conference Series, p. 171.
- Yosinski, Jason et al. (2014). “How Transferable Are Features in Deep Neural Networks?” In: *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*. NIPS’14. Montreal, Canada: MIT Press, pp. 3320–3328. URL: <http://dl.acm.org/citation.cfm?id=2969033.2969197>.
- Zaldarriaga, Matias, Steven R. Furlanetto, and Lars Hernquist (2004). “21 Centimeter Fluctuations from Cosmic Gas at High Redshifts”. In: *The Astrophysical Journal* 608.2, pp. 622–635. DOI: 10.1086/386327. URL: <https://doi.org/10.1086%2F386327>.

- Zaroubi, S. (2013). “The Epoch of Reionization”. In: *The First Galaxies*. Ed. by T. Wiklind, B. Mobasher, and V. Bromm. Vol. 396. Astrophysics and Space Science Library, p. 45. DOI: 10.1007/978-3-642-32362-1_2. arXiv: 1206.0267.
- Zhang, Y. G. et al. (2018). “Fast Radio Burst 121102 Pulse Detection and Periodicity: A Machine Learning Approach”. In: *ArXiv e-prints*. arXiv: 1809.03043 [astro-ph.HE].
- Zheng, H. et al. (2014). “MITEoR: a scalable interferometer for precision 21 cm cosmology”. In: 445, pp. 1084–1103. DOI: 10.1093/mnras/stu1773. arXiv: 1405.5527 [astro-ph.IM].
- Zheng, H. et al. (2017). “An improved model of diffuse galactic radio emission from 10 MHz to 5 THz”. In: 464.3, pp. 3486–3497. DOI: 10.1093/mnras/stw2525. arXiv: 1605.04920 [astro-ph.CO].